



HAL
open science

Essays on matching with limited information and bounded rationality

Laure Goursat

► **To cite this version:**

Laure Goursat. Essays on matching with limited information and bounded rationality. Economics and Finance. École des Ponts ParisTech, 2024. English. NNT : 2024ENPC0017 . tel-04757027

HAL Id: tel-04757027

<https://pastel.hal.science/tel-04757027v1>

Submitted on 28 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



École des Ponts
ParisTech



THÈSE DE DOCTORAT
de l'École des Ponts ParisTech

**Essays on Matching
with Limited Information and Bounded Rationality**

École doctorale N°465, Économie Panthéon-Sorbonne (EPS)

Théorie, Micro-économie

Thèse préparée à PSE-École d'Économie de Paris

Laure GOURSAT

Thèse soutenue le 29 janvier 2024

Composition du jury :

Olivier COMPTE
École d'Économie de Paris
École des Ponts ParisTech

Président

Yair ANTLER
Tel Aviv University

Rapporteur

Hector CHADE
Arizona State University

Rapporteur

George J. MAILATH
University of Pennsylvania
Australian National University

Examineur

Leeat YARIV
Princeton University

Examinatrice

Philippe JEHIEL
École d'Économie de Paris
École des Ponts ParisTech
University College London

Directeur de thèse

Francis BLOCH
École d'Économie de Paris
Université Paris 1 Panthéon-Sorbonne

Co-directeur de thèse

Acknowledgments / Remerciements

Je dédie cette thèse à tous ceux qui l'ont rendue possible, par leur soutien intellectuel, financier, moral et affectif.

A mes directeurs de thèse, Philippe JEHIEL et Francis BLOCH pour leur double encadrement, dans deux styles bien différents. J'ai été marquée par leurs travaux inspirants, leurs visions de l'économie et de la recherche. Je leur suis très reconnaissante pour la liberté qu'ils m'ont laissée dans le choix de mes sujets, leurs précieux conseils sur mes travaux, leur disponibilité, leur écoute, leur patience et leur enthousiasme. Je les remercie en particulier pour leur soutien au moment du covid, de ma visite à l'Université de Pennsylvanie et de ma recherche de contrat post doctoral. J'espère que notre filiation de recherche perdurera bien au-delà de cette thèse.

Aux membres de mon comité de thèse, Olivier COMPTE et Olivier TERCIEUX, pour leur intérêt pour mes travaux, leurs conseils avisés et leur bienveillance.

To the members of my Ph.D. thesis jury, Yair ANTLER, Hector CHADE, Olivier COMPTE, George MAILATH and Leeat YARIV. I am grateful for their interest in my work and for accepting my invitation to join the jury. I thank them in advance for their comments during the Ph.D. defense. Special thanks to the referees Yair ANTLER and Hector CHADE for reading the thesis in detail and for making many recommendations during the pre-defense on ways to improve this research in the short, medium, and long run.

To the community of researchers at PSE-Ecole d'Economie de Paris and to the theory group in particular. I feel grateful for the intellectually stimulating environment, the inspiring discussions, and the generous advice. Special thanks to the organizers or the various reading and working groups for Ph.D. students Francis BLOCH, Evan FRIEDMAN, Jeanne HAGENBACH, Frederic KOESSLER, Eduardo PEREZ-RICHET and Nikhil VELLODI.

To my (hopefully) future coauthors Francis BLOCH and Giacomo WEBER for the likable experience of collaborating on a research project.

To George MAILATH for his warm welcome, kind interest, and outstanding advice during my visit at the University of Pennsylvania, Philadelphia, US (from January to April 2023). I am extremely grateful for this privileged experience. I have found George's precision and rigor in his approach to economic theory inspiring. I miss our weekly meetings. I am also grateful to the rest of the theory group at Penn for interesting discussions, in particular to Kevin HE, Andrew POSTLEWAITE, Rakesh VOHRA and beyond to Alex REES-JONES (Wharton) and Sampath KANNAN (engineering). I am also grateful to Larry SAMUELSON: although he could not host me at Yale, I have had the chance to meet him at several conferences, and I am grateful for his constant interest and kind advice.

Au personnel administratif, d'encadrement et de support des doctorants à PSE et à la direction de la recherche de l'Ecole Nationale des Ponts et Chaussées. Je suis reconnaissante pour les bonnes conditions de travail et de recherche à PSE. Je remercie particulièrement à PSE Lucia Roxana BAN, Véronique GUILLOTIN, Catherine BOBTCHIEFF et Sylvie LAMBERT pour leur professionnalisme, leur disponibilité et leur aide. Je remercie également Stéphane BRICE et son équipe pour leur assistance et leur aide en informatique. Je suis particulièrement reconnaissante à Dalida CHARTREL pour sa défense des intérêts des doctorants au sein de la direction de la recherche de l'ENPC et son soutien précieux sur la question du financement. Je lui dois en grande partie d'avoir pu finir ma thèse dans de bonnes conditions matérielles.

A la direction de la recherche de l'ENPC (contrat doctoral n°20/092) et au Conseil Européen de la Recherche (bourse ERC LTCSEI attribuée à Philippe JEHIEL) pour avoir financé cette thèse. Je remercie aussi PSE et la région Ile-de-France pour leur soutien financier à ma visite à Penn. Je remercie PSE et l'Association pour le Développement de la Recherche en Economie et en Statistique pour le financement

de conférences à Paris, Toulouse, Oxford, Tel Aviv, Tokyo. Ces conférences m'ont permis de présenter mes travaux et de m'informer sur ceux des autres chercheurs. Elles m'ont aussi motivée à achever cette thèse pour intégrer la communauté des chercheurs en économie théorique. Elles m'ont donné l'occasion de rencontrer les chercheurs externes à PSE que j'admire pour leur travail et leur engagement pour la recherche et la science.

A Eduardo PEREZ-RICHET, pour le contrat de post-doctorat à Sciences Po (financé par sa bourse ERC IMEDMC) qui débutera en février 2024. La perspective d'intégrer Sciences Po pour y développer ma recherche et de nouvelles collaborations me réjouit.

To my various teachers at school and university for educating me and making school the place of my fulfilment. I have had a long academic track record with obvious detours. But I have been privileged to attend amazing institutions: Ecole de la Marche, Collège Yves du Manoir, Lycée Hoche, HEC Paris, Université Paris-Sud, Ecole des Ponts, Ecole Polytechnique, Ecole d'Economie de Paris. I am glad I have finally found research (and economics). Special thanks to my teachers in Game Theory (so far): Olivier COMPTE, Philippe JEHIEL, Frederic KOESSLER, Yukio KORIYAMA, Antonin MACÉ, Jörgen WEIBULL, Bruno ZILLOTTO.

To my students at PSE, ENPC, and Paris 1, for the challenge of presenting economic models in the clearest possible way. Through their curiosity as beginners in economics, they have helped me reflect on the way we usually model our world in economics. I also thank Francis BLOCH, Bernard CAILLAUD, Nicolas JACQUEMET, and Bruno ZILLOTTO for trusting me and giving me these opportunities to teach. Special thanks to Pierre BERTRAND for entrusting me with the main lectures of Game Theory in second year at ENPC.

To my fellow Ph.D. students at PSE, for the friendly moments there and for their support in periods of doubt or stress. Special thanks to my friends Giacomo WEBER and Pierre BARDIER. Their presence at PSE and our friendship have been very precious to me along this thesis. Special thanks also to Benjamin BLUMENTHAL, Béla ELMHAUSER, Alexis GHERSENGORIN, Monica GOMEZ-OSPINA, Euiyoung JUNG, Xavier KOCH, Sofiiia MUN, and Eric VANSTEENBERGHE.

To Ph.D. students and visiting students at Penn Economic Department during the Spring term 2023, for making my visit there a surprising friendship experience, in particular to Wei XIONG, Anaïs FABRE, Javiera GARCIA, Kathleen HUI, Natacha SPITZER. I already miss our talks in Penn and our trips in Philly and beyond, and I miss them all. Special thanks to Natacha SPITZER and Kathleen HUI for being so welcoming in the theory group / office.

A mes amis externes à PSE et Penn, en particulier à Xavier GONTIER et Aurora SIYI DING, pour leur intérêt sincère, leur soutien et leur amitié.

A ma famille, que j'aime si fort. Je veux remercier mes parents et mes frères et sœurs au sens large pour leur soutien affectif pendant toute la durée de ma thèse. Le fait qu'ils n'en comprennent pas toujours bien le sens ou l'intérêt ne rend pas leur soutien moins important à mes yeux, bien au contraire. Je remercie en particulier mes parents pour l'éducation qu'ils m'ont donnée, et la compréhension et l'amour qu'ils me donnent encore. Je veux les remercier d'avoir fondé cette merveilleuse famille recomposée. Je suis infiniment reconnaissante à mon Papa "Sounet" de nous avoir adoptés, Astrid, Dauphine, Sébastien et moi. Je dédie cette thèse tout particulièrement à ma Maman qui, je l'espère, et avec de bonnes raisons, est en train de gagner son combat contre la maladie. Je remercie mes frères et sœurs Astrid, Dauphine et Sébastien d'être ces personnalités diverses, brillantes, et fascinantes avec qui j'ai eu la chance de grandir, et d'être un repère et un pilier dans ma vie. Je remercie tout particulièrement Astrid pour ses soins et son travail sur ma jument Koo Star, notamment pendant ma visite aux Etats-Unis et le rendu de ma thèse.

A Lucas, que j'aime tant, et qui n'a eu de cesse de m'encourager dans cette thèse. Il a aussi contribué à cette thèse en étant mon modèle d'éthique de travail, de curiosité et d'honnêteté intellectuelle. Je lui serai toujours reconnaissante pour sa présence, son soutien et son amour ces 4 / 6 / 10 dernières années.

Contents

Preamble	3
Introduction	4
Summary (in English)	9
Résumé (en français)	12
1 Information and Discrimination in Matching with Priority Scores	17
1 Introduction	18
2 Literature review	20
3 Model	22
3.1 Frictional market and Application Game	22
3.2 Benchmark Market with perfect information	23
3.3 Leading example: Social Housing in Europe	23
3.4 Model justifications	24
4 Equilibria	25
4.1 Preliminaries	25
4.2 Example	26
4.3 General results	28
5 Welfare	33
5.1 Ex ante	33
5.2 Interim	33
6 Conclusion	37
Appendix A Proofs for results in the main	39
Appendix B Supplements	50
B.1 Pure (asymmetric) Bayes-Nash equilibrium - Partial characterization	50
B.2 Symmetric (interior) equilibrium - Comparative statics	51
B.3 Extension: Endogenous cost and the dynamic	53
B.4 Extension: Larger or no truncation	55
B.5 Extension: Imperfectly correlated preferences	55
B.6 Proofs for supplements	57
2 A Heuristic Approach to Matching and Stability	61
1 Introduction	62
2 Literature review	64
3 Model: Valuation and v-stability (pure)	65
3.1 The marriage market	65
3.2 The information	65
3.3 The valuation	66
3.4 Connection with literature	67
3.5 V-stability	68
4 Results: V-stable matchings (pure)	69
4.1 Characterization: Happiness sorting	69
4.2 Structures on preferences and existence	71
4.3 Existence of v-stable matchings	73

4.4	Blocking pair dynamic and convergence	74
5	Model: Valuation and v-stability (mixed)	77
5.1	Mixed matchings	77
5.2	The valuation	79
5.3	V-stability	80
6	Results: v-stable matchings (mixed)	82
6.1	General result: Existence	82
6.2	Misaligned preferences: Characterization	84
7	Discussions and extensions	84
7.1	Unmatched agents	84
7.2	Valuation with bundles of goods	86
8	Conclusion	86
Appendix A	Proofs	92
Appendix B	Additional content	101
B.1	Valuation with memory	101
B.2	V-stable mechanism	102
3	Robust Incomplete Information Stability	105
1	Introduction	106
2	Literature review	108
3	Model: Matching with incomplete information and no transfers	110
3.1	The market	110
3.2	The information	110
3.3	An example	111
3.4	Complete information stability	112
3.5	Robust incomplete information stability	113
4	Results: Robust stable matchings	114
4.1	Common preferences	114
4.2	Comparison with LMPS	119
4.3	Other preference structures	121
5	Concluding comments	123
Appendix A	Proofs	125
4	Pre-Matching Information Acquisition	129
1	Introduction	130
2	Literature review	131
3	Model	134
3.1	The college admission problem	134
3.2	Information and learning	134
3.3	Matching	135
3.4	Bayesian Game	135
4	Results	135
4.1	Preliminaries	135
4.2	The case $K = 1$	136
5	Work in Progress	140
5.1	Matching and welfare in the baseline model	140
5.2	Larger capacities at colleges	141
5.3	Larger learning capacities	141
5.4	Other mechanisms	142
6	Provisional conclusion	143
Appendix A	Proofs	145

PREAMBLE

Introduction and summary / résumé

Introduction

As an introduction to my Ph.D. thesis, I would like to explain in simple words what matching, economics, and theory are (or rather what I have understood of these so far) and how research in these fields can (or should) connect with real life.

Matching

Partnerships

Most of our human production and consumption cannot be generated by one single person. In many cases, it can only result from some kind of partnership between several people. Only a pair of two people, one male and one female, can jointly produce offspring. Only a firm owning capital and workers owning skills can jointly produce goods and services. This is due to some underlying physical or biological constraint in our environment that we may summarize under the word complementarities. The fact that human beings are social animals directly implies that we have based our societies on these partnerships.

Matching problems

Matching problems study the formation of these partnerships or “matches” between people or “participants”. A matching market is the place where matches between participants are formed and broken. There are many examples of real-life matching markets. I have mentioned the dating market and the job market. The problems of allocating housing to households, students to schools, and organs to patients can also be viewed as matching problems.

A matching market is typically two-sided (participants are grouped into two sides, where matches can only form between agents from different sides) and one-to-one (each participant can match with at most one agent). A matching market can be with monetary transfers. In the private job market (private housing market), firms and workers (landlords and tenants) can freely negotiate on the wage (rent). But many matching markets forbid or regulate monetary transfers for ethical or fairness reasons. In the public job market for civil servants (allocation of doctors to hospitals, of teachers to schools), wages are fixed and cannot be negotiated. In social housing, rents are set in advance. Trading organs is forbidden in most countries. A two-sided, one-to-one, no transfer matching market is usually called, by analogy, a “marriage market”.

Important primitives of matching problems are the preferences (which specify how participants value or rank each other) and the allocation mechanism (which specify how participants can express their preferences and how they are matched).

For example, preferences are common if all agents from a given side of the market value the agents on the other side in the same way. Preferences are aligned if whenever an agent A likes a potential partner B a lot, then B also likes A a lot. Preferences are idiosyncratic when they vary within and across sides. Deferred Acceptance (Gale and Shapley (1962) [2]) is the most famous allocation mechanism for marriage markets. In this mechanism, agents from both sides of the market are asked to rank potential partners. In the male-proposing Deferred Acceptance, each male then asks his most preferred female whether she wants to mate with him. Each female chooses her most preferred male among the applicants, if any, and tentative matches form. In the next round, previously rejected males apply to their second choice. Females receiving new applications are allowed to substitute a tentative partner from a previous round for a preferred applicant. The process continues until all males are matched. Deferred Acceptance is widely used worldwide, especially for matching students to universities.

Matching in the eye of the analyst

Economics

Economics is the social science studying scarcity and its implication on welfare. In particular, microeconomics studies individual behaviors at the level of one market. Economic theory uses models –

formal simplified representations of reality to gain insight into economic problems. In these models, an agent's satisfaction from an economic outcome is modeled through a "utility function", summarizing all the benefits and costs associated with the outcome, and each agent seeks to maximize her utility. In matching specifically, the utility functions enclose the preferences over the various possible matches.

Decision and Game Theories

In some economic situations, the maximization problems that agents solve remain independent. Each agent's decision or "action" impacts only her own utility. The agent's maximization problem, therefore, remains reasonably simple to solve. It may become more complex when information about the situation the agent is facing is imperfect. For example, if there is some uncertainty on the utilities associated with the various outcomes. In this case, before making a decision, the agent needs to form beliefs about the world she is in. Decision theory is the field of mathematics formally modeling and analyzing decision problems.

In most economic situations, though, utilities are interdependent. There are some "externalities", meaning that the action of one agent affects the utilities of other agents. This creates a strategic interaction or game, where agents try to respond to the externalities by adjusting their behaviors or "strategies" to other agents' strategies. Game Theory is the field of mathematics formally modeling and analyzing strategic interactions.

Decision and Game Theory propose solution concepts, which describe a generic way in which economic agents behave in a decision problem or game. Solution concepts usually have an interpretation in terms of rationality. The most famous solution concept for games with complete information is Nash equilibrium (Nash (1950) [6]). It demands that no agent wants to unilaterally deviate from her action after observing other agents' actions. To play Nash equilibrium, players must hold rational expectations about the behavior of other players. A stronger concept is equilibrium in dominant strategies: all agents play a strategy that is better than any other strategy, whatever the other agents are doing. In decision problems with incomplete information, the main solution concept is Bayesian rationality. It demands that a player's beliefs are equal to the true probabilities of each state of the world, conditional on the player's information. For example, suppose butterflies go out in the mornings on only half (respectively a fourth) of the sunny (resp. rainy) days, and that sunny and rainy days are equally likely. A Bayesian agent seeing a butterfly in the morning would think that the day will be sunny with a probability of two-thirds. In games with incomplete information, the concept of Bayes-Nash equilibrium (Harsanyi [3]) mixes the two approaches.

Micro-economic theory uses decision and game theory as a tool. Most of the job of micro theorists consists of modeling an interesting economic environment as a game and applying a solution concept from Game Theory to this game. This results in a solution of the game, predicting how agents will behave in the economic situation considered. Economists then study the welfare properties of the solution. The main social objective is efficiency, with various criteria. For example, a solution is said to be Pareto-optimal if it is not possible to improve one agent's outcome without deteriorating the outcome of another agent. A solution is utilitarian optimal if it maximizes the sum of all agents' utilities. Fairness can also be viewed as a social objective with, again, multiple criteria.

Matching as an economic theory problem

The outcome of matching markets (job, marriage) governs much of our lives and happiness. Ideally, we would like agents with high complementarities to be effectively matched. Said differently, welfare stakes in matching are very high. Thus, matching appears as a natural and priority field of research in economics. Indeed, in the last sixty years, matching problems have received much interest from economists.

Matching markets involve some obvious strategic interactions. In general, the fact that participant *A* applies to *B* decreases the chances of another participant *C* to match with *B*. So, a matching market induces a game between participants, and Game Theory is needed for the analysis of this market. The convention there is to call a strategic participant an "agent" and a non-strategic participant (behaving according to some exogenous fixed rule an "object").

Matching Theory

In parallel, matching theory has developed its own solution concepts to describe the expected outcome on the matching market directly in terms of matching rather than through behaviors. A pairwise stable matching is a matching where no currently matched agent would prefer to be single, and there is no pair of two participants that would both prefer matching with each other rather than staying with their current partners. This latter condition is called the “no blocking pair condition” and can be interpreted as a fairness criterion: it demands that there is “no justified envy” on the market. Pairwise stable matchings appear as a credible prediction of the long-run outcomes in a market where agents can unilaterally break a match and freely form a new match if both new partners agree. Fortunately, at least when all agents have perfect knowledge and understanding of the market, a pairwise stable matching always exists (Gale and Shapley (1962) [2]).

Market Design

Economists ask two main questions about markets. The first one is: If we let agents behave independently and maximize their own utility functions, what outcome should we expect? This outcome is called the “decentralized equilibrium” of the market. The answer to this question depends both on the market considered and on the solution concept used. The second one is: If a social planner could force agents to behave in a given and coordinated way to maximize a social objective, what could be the outcome? This outcome is called the “centralized social optima”. The answer to this question depends both on the considered market and on the social objective targeted.

Ideally, we would like the decentralized equilibrium and the centralized social optima to coincide. If they do not, it becomes interesting to engage in market design. The design question is the following: can we organize the market environment (the mechanism, the information) in a way that ensures that the decentralized equilibrium achieves the social optima? Instead of considering that the environment is fixed and mechanically applying a solution concept to derive a solution, we fix what is a desirable solution and characterize the environment that enables us to reach it. The economist becomes an engineer. If such a design exists and we are able to find it, we should recommend policymakers and market operators to adopt it.

In standard markets for homogeneous goods using prices as a mechanism, and with perfect competition, the decentralized equilibrium (defined as demand equals supply) coincides to a large extent with the efficient outcomes (in the Pareto sense). We usually refer to these results as the “Welfare Theorems”.

Unfortunately, we do not have equivalent results for matching markets. In matching markets, in general, stability and efficiency are not compatible. One example can be given with Deferred Acceptance. Consider a marriage market where matching occurs with the (male-proposing) Deferred Acceptance mechanism. At Nash equilibrium of the induced game, all agents truthfully report their preferences (this is even a dominant strategy, which we summarize by saying that Deferred Acceptance is a “strategy-proof mechanism”), and whatever the preferences, this results in a pairwise stable matching. Yet, in terms of efficiency, this matching is Pareto-optimal only for the male side. For the female side, it is the worst matching among the set of all stable matchings.

The consequence is that there is much need and room for design in matching markets. This is probably why matching is one of the rare fields where economic research has managed to connect with political decisions. The most prominent example is college admissions in the US, where numerous student assignment systems have been redesigned in the last decades using input from economists.

Matching in the eye of the participant

Matching complexity

Nobel prize Alvin Roth defines a matching market as “a market in which prices do not do all the work” (Roth (2017) [7]), implicitly comparing standard markets for homogeneous goods and matching markets.

A standard market connects sellers willing to sell goods and buyers willing to buy and consume goods. All relevant information is summarized in the price, which acts as a signaling device, and agents just need to solve a (simple) decision problem: whether they want to buy / sell (and how many units) at a given price.

On matching markets with monetary transfers, prices still do “part of the job”; on matching markets with no transfers, prices are absent. An important corollary is that a matching market participant cannot just choose an object as long as she can afford it; she has to be chosen as well. There is an unfortunate applied consequence to this. To be able to play optimally on matching markets, an agent needs to know the market she is in finely.

The issue is that matching markets are utterly complex environments where perfect information and rationality are unlikely to happen. In general, participants only have limited information about their environment and bounded rationality about how they could try to understand it better.

Limited information

Playing optimally in a matching game demands information about the preferences and the mechanism. Unfortunately, this information may not be readily available on the market.

One’s own preferences describe how happy one would be in all possible matches. But intuition suggests it may be hard to assess future happiness prior to experiencing the match, so before effectively being married or hired. This is especially true when preferences include some idiosyncrasies.

This is mostly because the utility from a match or “match utility” depends not only on one’s own productive characteristics but also on the partner’s characteristics. However, collecting information on a potential partner before approaching him and starting to produce with him can prove challenging. When accepting a job position, some aspects of the job (quality of daily interactions with colleagues, effective workload) are not written in the contract and can only be discovered through experience in the job.

In addition, the kind of partnerships we have in mind are rare events, and the same partnerships are maintained over long time horizons. At the extreme, some people get married, go to college, receive an organ only once, and occupy only a few different houses or job positions in their whole lives. These people lack experience and are unable to collect data signaling their potential partner’s characteristics and their future happiness in different matches.

Finally, even in contexts where participants can acquire information by approaching other participants during an exploration stage before matching, information acquisition is either constrained or costly. A high school student who wants to learn about colleges is time-constrained and will not be able to learn about all the colleges she could rank in the mechanism. Still in the school choice example, the best way to acquire information is likely to visit the colleges’ campuses, hence moving effort and travel costs. In this case, even a rational agent, could decide to remain (partially) uninformed if the cost of acquiring information outweighs the expected benefit of being informed.

As soon as the allocation mechanism is not strategy-proof, agents also want to know about their priorities. The priorities describe how the (non-strategic) objects value the (strategic) agents. They determine the chances of an agent being accepted at an object conditional on applying. An example of a non-strategy proof mechanism that is widely used in practice (especially for college admission) is truncated Deferred Acceptance, where agents cannot list all available options in their rank-ordered lists. In this mechanism, at Nash equilibrium, a student wants to list safe schools (where her probability of acceptance is high) and does not always list (first) her most preferred schools.

The issue is that priorities are often uncertain due to underlying uncertainty about the competition that one faces. In social housing, a household usually observes that the market operator has assigned a priority score to her, reflecting the level of emergency of her housing need. However, she is often uncertain about the number of other applicants and their scores.

The empirical matching literature supports these intuitions, providing evidence from various real-life markets that participants are, in general, poorly informed about their preferences and priorities.¹

Bounded rationality

In the matching markets where information is indeed imperfect, participants need to form beliefs. The issue is that the standard Bayesian paradigm appears far too demanding in complex environments like matching markets. It demands that agents store and process a large quantity of data in a highly sophisticated way. In a marriage market, for instance, there are many individual characteristics that are relevant for the outcome of the marriage (health, fertility, cooking abilities...). A Bayesian agent should, at any point in time, hold a joint probability distribution over possible sets of vectors of characteristic of all agents on the market. Whenever a new match would be formed, the Bayesian agent would interpret the matching and the resulting utilities as weak signals over the characteristics and would update her beliefs accordingly through Bayes' rule.

There is overwhelming evidence coming from the field or the lab that agents are unable to do so.¹ They rather show systematic biases in the way they analyze the data, which are consistent with some simple heuristics of belief formation. The inference is then coarser but simpler.

¹We review these references in the introduction of each chapter.

Summary (in English)

This Ph.D. thesis models and analyses matching problems with limited information and bounded rationality. It is made of four chapters - 2 main articles (1) and (2), and 2 shorter and preliminary notes (3) and (4):

- (1) Information and Discrimination in Matching with Priority Scores
- (2) A Heuristic Approach to Matching and Stability
- (3) Robust Incomplete Information Stability
- (4) Pre-Matching Information Acquisition

Chapters (1), (2) and (3) are single-authored. Chapter (4) is co-authored with Francis BLOCH.

Motivations

This thesis is mainly motivated by applied concerns about how real-life matching markets work. It also seeks to address a gap in the economic matching literature.

Theoretical matching papers most often do not investigate the effect of limited information and bounded rationality or of any kind of friction that would burden real-life operations on matching markets.² With respect to information, most existing papers assume perfect information on one's own preferences and set strategy-proof mechanisms (hence, no incentive to know about others). This is mainly for tractability reasons but at the expense of realism. With respect to rationality, the majority of existing papers that look at matching in incomplete information model perfectly Bayesian agents. The literature on behavioral matching remains quite scarce but growing. Behavioral matching papers typically plug a known behavioral bias into the standard matching framework.

The perfect information literature is still very useful as a benchmark for the pure allocation problem. It often defines the maximum welfare that can be achieved on the market. But these models should not be taken too seriously to the field: they are bound to produce incorrect predictions in terms of matching and welfare, and biased design recommendations.

Research question

The ambition of this research has been to model and analyze matching markets featuring participants with limited information and bounded rationality. The analysis has aimed at both a descriptive value (improving our understanding of matching markets and of the world) and a normative value (making policy design recommendations to achieve a number of social objectives such as efficiency or fairness). Common sense suggests both the predictions and the recommendations should be revised once one introduces the informational and cognitive frictions.

Methodology

The exact way I model limited information and bounded rationality depends on the motivations and the real-life applications I had in mind. I have tried to design it as simple and natural as possible. The limited information always applies to the preferences and priorities. The information structure is often (a) private information (an agent only observes her own preferences and priorities and is ignorant of others' preferences and priorities), (b) present information (an agent only observes the current outcome on the markets and the resulting signals on preferences and priorities), (c) asymmetric information (agents from one side of the market have full information whereas agents from the other side of the market do not observe productive characteristics of potential partners), (d) constrained information (agents can acquire information but with is a limit: a capacity constraint or a cost), or rather a combination of these.

In terms of bounded rationality, this thesis makes a central distinction between instrumentation (the ability, given beliefs on the environments, to optimize) and cognition (the ability to form correct beliefs on the environment). It does not question perfect instrumentation in matching environments and

²With some interesting exceptions, which I discuss in the literature reviews of each chapter.

rather argues that cognition is the challenging part. In chapter (2), I model agents forming beliefs through a simple (non-Bayesian) heuristic. In the paper, I argue that this heuristic summarizes a very natural way to exploit the very limited information available on preferences. I also show that the heuristic captures a famous behavioral bias from the psychological literature, the “projection bias” (Dawes (1989) [1], Krueger and Clement (1994) [4]). In chapter (3), I model incomplete information agents as holding a set of reasonable beliefs on the characteristics of potential partners (rather than a probability distribution over these possible characteristics as in the Bayesian framework). Agents can only dismiss some beliefs in this set, using their observation that the current matching is stable.

The analysis uses game theory as a tool to analyze the strategic interactions at play – both non-cooperative game theory (computing Bayes-Nash equilibria in chapters (1) and (4)) and cooperative game theory (characterizing pairwise stability in chapters (2) and (3)). In this sense, this thesis is mostly applied theory.

It also proposes methodological contributions by identifying novel strategic interactions and equilibrium structures (in chapter (1), the interactions within the agent side of an agent-object market, and the block structure arising at symmetric Bayes-Nash equilibrium) and by defining new solution concepts (in chapter (2), a learning heuristic and a notion of (stable) mixed matching).

Summary of each chapter

Chapter (1) models a matching market where agents have private information on their priority scores and must choose an object to which they apply. The analysis derives the Bayes-Nash equilibria, computes welfare ex-ante and interim, and discusses implications for market design. Three main findings emerge. One, there is no symmetric equilibrium in pure strategies. Second, the symmetric equilibrium exhibits a block structure: agents sort into a finite number of classes of neighboring scores where they use the same application strategy. Third, the inefficiencies proceeding from the frictional market design prove interim asymmetric: low-score agents are better off under private information than under public information. In total, private information mitigates the discriminatory power of the priority system.

Chapter (2) models a marriage market with unknown preferences (agents only observe the current matching and realized match utilities) and defines a novel and natural heuristic of belief formation (valuation), which incorporates a well-known and documented cognitive bias (the projection bias). Under this heuristic, an agent estimates a counterfactual match utility by extrapolating from realized match utilities: his own utility and the weighted average utility of all current partners of the targeted partner’s type. I study how this reshuffles the market outcome, as given by pairwise stable matchings when agents have valuation beliefs (v -stability). When restricting attention to pure matchings, I find that v -stability is equivalent to any two partners holding the same rank according to current utilities (happiness sorting). The predictions under specific preference structures are then straightforward. The alignment of interests across the market governs the size of the v -stable set from empty to maximal. The correlation of preferences by agent or target stabilizes the positive assortative matching. For a generic market, though, we get neither the existence of a pure v -stable matching nor the convergence of a dynamic blocking pair process (predicting persistent moves on the market). The most general version of the model defines a notion of mixed matching, characterizing the proportions of each productive type matched with each partner type. The main result is a general existence theorem for v -stable matchings in the mixed extension.

In the short chapter (3), I consider a matching market with no transfers and incomplete asymmetric information: on one side, agents do not observe types of potential partners; they just observe the type of their current partner. The model can represent civil servants’ job markets where wages are regulated and where employers have trouble learning about workers’ productivity prior to hiring. I apply the definition of incomplete-information stable matchings by Liu, Mailath, Postlewaite, and Samuelson (2014) [5]: a pair is blocking if both partners strictly want to block under any reasonable beliefs they may have using their private information and common knowledge of stability. Even under monotonic payoffs, the incomplete-information stable set may be large – it depends finely on the market structure and the prior belief support. If the unknown workers’ type function is a bijection, the stable sets with

complete and incomplete information perfectly coincide (to include only positive assortative matchings). I show, using examples, that the robust approach can reach precise predictions even beyond the monotonic case.

In the short chapter (4), we study a college admission problem gathering heterogeneous students and colleges where students can endogenously acquire information on their own preferences. Students' preferences over colleges include a common component, which is common knowledge, and a private component, which is unknown ex-ante. Students can learn about the private components, before matching occurs through a standard Deferred Acceptance mechanism with common priorities. The question is: What information do students acquire, as a function of their priority rank? With unit constraint on learning and unit capacities at colleges, we find that the best student learns about one of the best colleges. Students with lower-priority learn about the best college among the ones where they are admitted for sure. The proof uncovers a novel additive property of the values of information. We discuss matching and welfare implications and ongoing generalizations.

Epilog

I view the stakes of this strand of research as high and still growing. That's why I would like to continue developing this research in the future.

Decentralized (matching) markets, where strategic agents sequentially meet and trade (match), are as old as humankind. But centralized markets, where agents' meetings are coordinated in space and time, have gradually developed with our ability to organize in societies with rules. Centralized matching markets specifically have appeared relatively recently, and the current trend is much towards more centralization. The public markets for civil servants' jobs, slots at universities, social housing, and organs in developed economies have been centralized at the national scale in the last decades. This gives economists avenues for market design and data collection.

Résumé (en français)

Cette thèse de doctorat modélise et analyse des marchés d'appariement avec information et rationalité limitées. Elle est divisée en quatre chapitres - 2 articles principaux (1) et (2), et deux notes plus courtes et préliminaires (3) et (4) :

- (1) Information et discrimination sur des marchés d'appariement avec scores de priorité
- (2) Une approche heuristique du problème de stabilité sur un marché d'appariement
- (3) Stabilité robuste en information incomplète
- (4) Acquisition d'information avant appariement

Je suis la seule auteure des chapitres (1), (2) et (3). Le chapitre (4) est coauteuré avec Francis BLOCH.

Motivations

Cette thèse est principalement motivée par des considérations d'économie appliquée sur le fonctionnement des marchés d'appariement réels. Elle cherche également à répondre à une carence dans la littérature économique actuelle sur les marchés d'appariement.

La plupart des articles théoriques sur les marchés d'appariement n'étudient pas l'effet de l'information ou de la rationalité limitée, ni en fait de n'importe quelle friction qui empêcherait ces marchés de fonctionner de façon optimale.³ En ce qui concerne l'information, une majorité d'articles fait l'hypothèse que l'information sur ses propres préférences est parfaite, et modélise des mécanismes d'allocation où aucun agent n'a intérêt à manipuler ses préférences (donc où il n'y a aucune incitation à apprendre sur les autres agents). Cette hypothèse est faite principalement pour que le modèle théorique reste relativement simple à analyser, mais le rend peu réaliste. Du côté de la rationalité, la majorité des articles qui s'intéressent aux marchés d'appariement en information incomplète modélise des agents bayésiens. La littérature mêlant économie comportementale et marchés d'appariement reste réduite, mais en croissance. La plupart des articles dans ce domaine introduisent un biais comportemental connu dans le problème d'appariement classique.

La littérature avec information parfaite est très utile en tant que point de référence pour le problème d'allocation pur. Généralement, elle définit le niveau de bien-être collectif maximal qui peut être atteint sur le marché. Mais ces modèles ne sont que de peu d'utilité sur le terrain : ils produisent nécessairement des prédictions incorrectes sur l'appariement final et le bien-être en résultant, d'où des recommandations politiques biaisées.

Question de recherche

L'ambition de cette thèse a donc été de modéliser et d'analyser des marchés où les participants n'ont qu'une information et une rationalité limitées. L'analyse se veut à la fois descriptive (pour améliorer notre compréhension des marchés d'appariement et du monde dans lequel nous vivons), mais aussi normative (faisant des recommandations politiques sur la meilleure façon d'atteindre des objectifs sociaux comme l'efficacité ou l'équité). On s'attend à ce que les prédictions et les recommandations issues du modèle d'appariement soient modifiées dès lors que les frictions informationnelles et cognitives y sont introduites.

Méthodologie

La façon exacte donc je modélise l'information et la rationalité limitée dépend des motivations de chaque chapitre et des applications concrètes associées. Je me suis efforcée de la penser comme la plus simple et naturelle possible.

Dans tous les chapitres de la thèse, l'information limitée porte sur les préférences et les priorités. Selon le contexte, c'est (a) de l'information privée (l'agent observe seulement ses propres préférences et priorités et ignore celles des autres), (b) de l'information courante (l'agent observe seulement

³Il existe cependant certaines exceptions intéressantes, que je présente dans les revues de littérature de chacun des chapitres.

l'issue courante sur le marché et les signaux sur les préférences et priorités qui en résultent), (c) de l'information asymétrique (les agents appartenant à un côté du marché n'observent pas les caractéristiques de leurs partenaires potentiels), (d) de l'information contrainte (il est possible d'acquérir de l'information mais avec une limite de capacité ou un coût), ou plutôt une combinaison de plusieurs de ces aspects.

Pour ce qui est de la rationalité, cette thèse fait une distinction importante entre l'instrumentation (la capacité, étant donné certaines croyances sur son environnement, de choisir les stratégies optimales) et la cognition (la capacité de former des croyances correctes sur son environnement). Elle ne remet pas en cause l'instrumentation parfaite sur les marchés d'appariement, et considère la cognition comme la tâche la plus difficile. Dans le chapitre (2), je modélise des agents qui forment des croyances au travers d'une heuristique (non bayésienne). Je démontre que cette heuristique représente une forme très naturelle d'apprentissage exploitant l'information disponible. Je soutiens également que cette heuristique incarne un biais comportemental clairement identifié dans la littérature en psychologie, le "biais de projection" (Dawes (1989) [1], Krueger et Clement (1994) [4]). Dans le chapitre (3), je modélise les agents en information incomplète comme ayant un ensemble de croyances raisonnables sur les caractéristiques de leurs partenaires potentiels (plutôt qu'une distribution de probabilités sur les caractéristiques possibles comme dans le cadre bayésien). Les agents peuvent seulement éliminer certaines croyances de cet ensemble, en remarquant que l'appariement courant est stable.

Ces travaux utilisent la théorie des jeux comme un outil pour analyser les interactions stratégiques – à la fois la théorie des jeux non coopérative (en calculant les équilibres de Bayes-Nash dans les chapitres (1) et (4)) et la théorie des jeux coopérative (en caractérisant les appariements stables par paire dans les chapitres (2) et (3)). En ce sens, cette thèse est principalement une thèse de théorie appliquée. Je propose aussi une contribution méthodologique en identifiant des interactions stratégiques et des structures d'équilibre nouvelles (dans le chapitre (1), les interactions entre agents sur un marché agent-objet et la structure en blocs de l'équilibre de Bayes-Nash symétrique), et en définissant de nouveaux concepts de solution (dans le chapitre (2), une heuristique d'apprentissage et une notion d'appariement mixte (stable)).

Résumé de chaque chapitre

Le chapitre (1) modélise un marché d'appariement où l'information sur les scores de priorités est privée et où chaque agent doit choisir un objet auquel il candidate. L'analyse caractérise les équilibres de Bayes-Nash, calcule le bien-être social du point de vue ex ante et intérim, et en tire des conclusions sur la meilleure façon d'organiser les marchés. Trois résultats principaux apparaissent. Tout d'abord, il n'existe pas d'équilibre symétrique en stratégies pures. Ensuite, l'équilibre symétrique présente une structure en blocs où les agents se rangent dans un nombre fini de catégories de scores pour lesquels ils utilisent exactement la même stratégie mixte. Enfin, les inefficacités liées aux frictions sur le marché sont asymétriques à l'étape intérim : les agents à faible score préfèrent l'information privée à l'information publique. Au total, l'information privée atténue la discrimination induite par le système de priorité.

Le chapitre (2) modélise un marché où les préférences ne sont pas connues (les agents observent seulement l'appariement courant et les utilités réalisées) et propose une nouvelle heuristique naturelle de formation des croyances (la valuation), qui intègre un biais cognitif connu (le biais de projection). Sous cette heuristique, un agent estime une utilité contrefactuelle en extrapolant à partir des utilités réalisées : sa propre utilité et la moyenne de l'utilité de tous les agents actuellement appariés à un partenaire du type visé. J'étudie la façon dont ce mode de formation des croyances perturbe les issues possibles sur le marché, données par l'ensemble des appariements stables par paires quand tous les agents ont des croyances de type valuation (*v*-stabilité). En me concentrant sur les appariements purs exclusivement, je trouve une condition nécessaire et suffisante pour la *v*-stabilité : les agents appariés doivent avoir le même rang quand on ordonne les utilités courantes au sein de chaque côté du marché (ordre sur le bonheur). Les prédictions sous des préférences standards s'en déduisent rapidement. L'alignement des intérêts contrôle la taille de l'ensemble des appariements *v*-stables, de l'ensemble vide à la taille maximale. La corrélation des préférences, au niveau de l'agent ou de la cible, stabilise

l'appariement positivement assortatif. Pour un marché quelconque, cependant, un appariement v -stable n'existe pas forcément, et un processus dynamique de paire bloquante ne converge pas nécessairement. La version la plus générale du modèle définit une notion d'appariement mixte, qui fixe les proportions de chaque type productif apparié avec chaque type de partenaire possible. Le résultat principal est un théorème d'existence des appariements v -stables dans cette extension mixte.

Dans le court chapitre (3), je m'intéresse à un marché d'appariement sans transferts et avec une information imparfaite asymétrique, telle que d'un côté du marché, les agents n'observent pas les types de leurs partenaires potentiels ; ils observent uniquement le type de leur partenaire actuel. Ce modèle peut représenter le marché du travail des agents de la fonction publique, où les salaires sont encadrés et où les employeurs publics ont du mal à évaluer la productivité des travailleurs avant de les engager. J'applique la définition des appariements stables en information incomplète de Liu, Mailath, Postlewaite et Samuelson (2014) [5] – une paire est bloquante si les deux agents souhaitent former cette paire quelles que soit leurs croyances raisonnables étant données leur information privée et leur connaissance commune de la stabilité. Même quand les paiements sont monotones, l'ensemble des appariements stables en information incomplète peut être grand – il dépend finement de la structure du marché et du support des croyances possibles ex ante. Si la fonction qui définit les types des travailleurs est une bijection, les ensembles stables en information complète et incomplète coïncident (ils incluent seulement les appariements positivement assortatifs). Je montre, sur la base d'exemples, que l'approche robuste peut parvenir à des prédictions relativement précises même sans l'hypothèse de monotonie des paiements.

Dans le court chapitre (4), nous étudions un problème d'allocation d'étudiants hétérogènes dans des universités hétérogènes, quand les étudiants ont la possibilité d'acquérir de l'information sur leurs propres préférences. Les préférences des étudiants sur les universités sont composées de paramètres communs, qui sont connaissance commune, et de paramètres privés, qui sont inconnus ex ante. Les étudiants peuvent apprendre sur les paramètres privés, avant que l'allocation ne soit réalisée par un mécanisme classique d'Acceptation Différée avec priorités communes. Nous posons la question : Qu'apprennent les étudiants en fonction de leur rang de priorité ? Quand l'apprentissage est limité à une seule université et que chaque université n'a qu'une seule place disponible, nous prouvons que le meilleur étudiant apprend sur une des deux meilleures universités. Les étudiants moins bien classés apprennent sur la meilleure des universités à laquelle ils sont sûrs de pouvoir être admis. La preuve de l'équilibre met en lumière une propriété nouvelle sur l'additivité des valeurs de l'information. Nous évoquons les conséquences en termes d'appariement et de bien-être, et les généralisations du modèle en cours.

Epilogue

De mon point de vue l'intérêt et la portée de ces thématiques de recherche sont importants et croissants. C'est pourquoi j'aimerais continuer à développer cette recherche à l'avenir. Les marchés (d'appariements) décentralisés, où des agents stratégiques échangent entre eux (forment des partenariats) au cours du temps, sont aussi anciens que l'espèce humaine. Mais les marchés centralisés, où les rencontres entre agents sont coordonnées dans l'espace et le temps se sont développés de façon graduelle en parallèle de notre capacité à nous organiser en sociétés régies par des règles. Les marchés d'appariements centralisés sont apparus particulièrement tard et la tendance actuelle va clairement vers plus de centralisation. Dans les économies développées, les marchés publics pour les postes de fonctionnaires, les places à l'université, les logements sociaux, les organes se sont concentrés à l'échelle nationale ces dernières décennies. Cette tendance fournit aux économistes des perspectives intéressantes d'ingénierie des marchés et de collecte de données.

References

- [1] R. M. Dawes. Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, 25(1):1–17, 1989.
- [2] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- [3] J. C. Harsanyi. Games with incomplete information played by “bayesian” players, i–iii part i. the basic model. *Management science*, 14(3):159–182, 1967.
- [4] J. Krueger and R. W. Clement. The truly false consensus effect: an ineradicable and egocentric bias in social perception. *Journal of personality and social psychology*, 67(4):596, 1994.
- [5] Q. Liu, G. J. Mailath, A. Postlewaite, and L. Samuelson. Stable matching with incomplete information. *Econometrica*, 82(2):541–587, 2014.
- [6] J. F. Nash et al. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.
- [7] A. E. Roth. Al roth on matching markets. *Social Science Bits Podcast*, 2017.

Chapter 1

WHETHER AND WHERE TO APPLY

Information and Discrimination in Matching with Priority Scores

Laure GOURSAT* – December 2023

We thank Yair ANTLER, Nick ARNOSTI, Francis BLOCH, Catherine BOBTCHEFF, Hector CHADE, Yeon-Koo CHE, Olivier COMPTE, Yannai GONCZAROWSKI, Michael GREINECKER, Guillaume HAERINGER, Kevin HE, Philippe JEHIEL, Sampath KANNAN, Frederic KOESSLER, Antonin MACÉ, Eduardo PEREZ-RICHET, George J. MAILATH, Alex REES-JONES, Larry SAMUELSON, Ludvig SINANDER, Marciano SINISCALCHI, Colin STEWART, Olivier TERCIEUX, Nikhil VELLODI, Rakesh VOHRA, Leeat YARIV, seminar participants at PSE, Sciences Po, ENS Paris Saclay, and conference participants at the Annual Conference of the Society for the Advancement of Economic Theory (Paris, France - July 2024), Summer School in Economic Theory (Tokyo, Japan - August 2024), Transatlantic Theory Workshop (Oxford, UK - September 2024), for helpful comments and conversations.

We thank the European Research Council (Philippe JEHIEL's advanced grant 2018-2022 LTCSEI) and Ecole Nationale des Ponts et Chaussées (doctoral funding contract n°20/092) for financial support.

Abstract: This paper considers a matching market where agents have private information on their priority scores and must choose an object to which they apply. The analysis derives the Bayes-Nash equilibria, computes welfare ex ante and interim, and discusses implications for market design. Three main findings emerge. One, there is no symmetric equilibrium in pure strategies. Second, the symmetric equilibrium exhibits a block structure: agents sort into a finite number of classes of neighboring scores where they use the same application strategy. Third, the inefficiencies proceeding from the frictional market design prove interim asymmetric: low-score agents are better off under private information than under public information. In total, private information mitigates the discriminatory power of the priority system.

Keywords: Matching markets, priorities, private information, discrimination.

JEL codes: D82, C78.

*Ph.D. candidate at Paris School of Economics (PSE) - 48 boulevard Jourdan, 75014 Paris, France (Office R6-41) - laure.goursat@psemail.eu

1 Introduction

Matching

Matching problems study the formation of productive partnerships, with numerous applications to marriage, labor, housing, college admissions, organ donation, and many more. In centralized matching markets, a mechanism summarizes the rules used to match participants. Important primitives of any matching problem and inputs into the mechanism are the preferences and priorities, which specify how participants value each other. "Preferences" ("priorities") summarize the tastes of strategic participants or "agents" (non-strategic participants or "objects").

Strong evidence from the field shows that most mechanisms used in practice use priority scores and are truncated.

Priority scores

A participant's priority score summarizes the participant's characteristics in a single real number. A higher score is associated with a higher chance of obtaining a desirable partner in the mechanism. Then, priorities are homogeneous (the score controls the priority of an agent to any object) and cardinal (rather than ordinal).

Examples of real-life markets with scores include college admissions, where a student is assessed based on her score in a standardized test; civil servant job markets such as teacher or doctor allocation, where civil servants accumulate points for experience or performance; social housing, where households are assigned a score reflecting the emergency of the housing need; or credit, where the credit score measure an individual's creditworthiness. This suggests that scores help to discriminate¹ between agents on the market.

Truncated mechanisms

In a truncated mechanism, agents are allowed to apply to only a limited number of objects. On the matching field, truncation is the rule more than the exception. Students, civil servants, households rank only a subset of the whole set of national Universities, job positions, social housing units. The motivations of the market operators for imposing such truncation are still debated. In his survey on school choice, Pathak (2016) argues that there are hidden operational costs associated with reviewing applications.

One important feature of truncated mechanism is that they are manipulable, even when the untruncated version is strategy-proof. The profitable preference manipulation consists in ranking safe objects high. Hence, agents would like to know about their order in priorities, so about other agents' priority scores.

Private information

The issue is that the default information structure on priority scores, absent any intervention by the market operator, is private information. An agent observes her own priority score but is ignorant of other agents' scores. This is because, in general, these scores are computed based on criteria of private circumstances.

In social housing, households are awarded priority points if they are currently homeless, have large families, or have health issues. A household knows about her status with respect to every criterion and is able to compute her aggregate priority score. However, she has no idea about other households' status with respect to the same criteria and cannot compute competitors' priority scores.

The empirical matching literature supports this intuition providing evidence that participants are poorly informed about their order in priority. Kapor, Neilson, and Zimmerman (2020) [12] using data from admissions of students to colleges in New Haven, US, show that beliefs about admissions chances differ from rational expectations values. They predict choice behaviors and quantify the welfare costs

¹Throughout the article, we use the word in a positive sense: discriminating means allocating a good (with a high probability) to those who need it the most.

of belief errors. Fabre et al. (2021) [8], using data from Chile, show that on-the-fly information about college programs' cutoff scores has a causal effect on reducing students' biases and application mistakes and improving students' outcomes.

Research question

Yet, theoretical matching papers do not investigate the effect of privately known priority scores.² Most papers model ordinal priorities and ordinal mechanisms (where agents are asked to rank available options) rather than cardinal ones (where agents reveal how much they like each available option). With respect to information, they mostly either assume perfect information (which is unrealistic) or set strategy-proof mechanisms (hence, no incentive to know about others).

In this paper, we explicitly model private information on priority scores, jointly with the other realistic and standard market frictions (truncation on the mechanism, participation cost). We characterize application behaviors, the allocation, and the welfare. The stakes are both descriptive (to improve our understanding of real-life markets) and normative (to engineer matching markets efficiently).

Overview of model

Our model formalizes the dilemma of "Whether and Where to Apply?" and gives insights about any market with uncertainty on priorities.

We define a stylized frictional matching market: two-sided, one-to-one, agent-object, with non-transferable utilities. Preferences and priorities are homogeneous, meaning that each object is characterized by a unique objective value, and conversely, each agent is endowed with a single (privately known) priority score. The allocation occurs through a truncated Deferred Acceptance (equivalently, truncated Serial Dictatorship) mechanism, with truncation one. Agents independently and simultaneously decide to apply to one or no object, where an application is costly. Then, each object goes to the highest priority agent in the pool of applicants.

We model strategic interactions on the market as a Bayesian Game of incomplete information termed "Application Game". On the market defined, any participant suffers uncertainty on who else applies and wonders "Whether and Where to Apply?".³ Her answer to that question should depend on the other participants' strategies since any higher priority agent applying to the same object eliminates her chances of getting the object. Thus, in building one's application strategy, the agent must consider the trade-off between being "ambitious", accepting the prospect of competition (targeting high-value objects), or being "practical", seeking coordination (targeting under-demanded objects).

Preview of results

The analysis elicits the equilibrium application strategies as defined by the Bayes-Nash equilibria of the Application Game. It finds that in any equilibrium, high-score agents are ambitious, and low-score agents are practical. We fully characterize the symmetric equilibrium, and uncover two salient features of the symmetric equilibrium structure. First, it is necessarily interior. Second, it has a "block structure": agents with scores on a continuous support sort into discrete classes (defined as groups of close priorities) where they adopt exactly the same strategy.

We compute the welfare associated with equilibrium outcomes using the two criteria of ex-ante and interim expected payoffs, and compare it with the level of welfare achieved without the frictions. We derive implications for market design. Although the frictional market design is sub-optimal for the criterion of ex-ante welfare, it maximizes participation, and the inefficiencies associated with the described market design are interim asymmetric. In many instances of the Application Game, we even find that low-score agents are better off in the sub-optimal (private information) design than in the optimal (public information) design. The conclusion is that the frictional design is less efficient but

²With some interesting exceptions, which we discuss in section §2.

³Because the model is static, agents do not wonder *when* to apply.

more egalitarian than the frictionless design. Private information mitigates the discriminatory power of the priority score system. This calls for a joint design of the priority score computation rule and the information structure.

Outline of paper

The rest of the paper is structured as follows. Section §2 reviews the related literature. Section §3 models the market and the associated game. Section §4 characterizes the Bayes-Nash equilibria. Section §5 investigates the welfare. Section §6 concludes.

All proofs are available in appendix §A. Appendix §B presents additional results. In particular, it proposes and processes three natural model extensions. It shows that the lessons learned from the baseline model extend to more sophisticated and realistic markets, where the cost is endogenous to the score, the truncation on the mechanism is larger than one, or the preferences are correlated.

2 Literature review

Matching with uncertain priorities

There is a recent but active literature on matching with private information or uncertainty. In the majority of existing matching papers, the uncertainty applies to one's own ex-post payoff in the match (preferences). We discuss three notable exceptions, where the uncertainty applies to the probability of acceptance (priorities), just as in the present paper.

Chade, Lewis, and Smith (2014) [5] consider a decentralized Bayesian game of admissions gathering two colleges and many heterogeneous students. Colleges have the same value to all students, and each student is characterized by a unique score. There is a cost of application for students, and in addition, colleges' evaluations of students' applications are uncertain.⁴ A student designs her application strategy⁵ (no application, application to one college, to both colleges) to her expected payoff, and a college designs her admission standard to maximize the total score of its student body under capacity constraints. Their model differs from ours in two ways. In their model, strategic interactions happen within the college side and between students and colleges, whereas in our model, strategic interactions happen only within the agent side. The uncertainty also differs. They set common knowledge of students' priority types but exogenous noise on the allocation, whereas uncertainty on priority order in our model endogenously arises from private information on priority scores. Their analysis finds that at Bayes-Nash equilibrium, student-college sorting may fail in two ways: first, weaker students sometimes apply more aggressively; second, weaker colleges might impose higher standards. Our analysis (see section §4.4) finds the opposite and more standard pattern: at the Bayes-Nash equilibria of the Application Game, higher-value objects are played more often and at higher scores.

Ali and Shorrer (2021) [1] define the general decision problem of students who are uncertain about their (correlated) priorities (whereas our paper studies a game between participants). Their focus is on the correlation between admission chances and the subsequent signaling effects. Because priorities, and admissions decisions, are correlated, the optimal portfolio involves applying to a combination of "reach," "match," and "safety" schools. In our model, as in many applications beyond school choice, priorities are homogeneous (perfectly correlated). We could generate results similar to Ali and Shorrer (2021) only after enlarging the truncation on the mechanism.

Avery and Levin (2010) [2] model students who are differentiated in their academic ability and in their fit for different schools. Each student knows her ability only imperfectly and, thus, is uncertain about the priority order. The focus of the paper is on early admissions (dynamic concern, contrasting with our static model). They show that early admissions have a sorting effect (early applications convey a signal of good fit from students to schools) and a competitive effect (lower-ranked colleges attract cautious high-ranked students).

⁴The key assumption about this uncertainty is a monotone likelihood property for the distribution of signals on students' scores to colleges. Therefore, a higher-score student always sends a higher proportion of good signals (vs. bad signals) on her score than a lower-score student, so that colleges use cutoff strategies at equilibrium.

⁵The paper-specific terminology says that students make "portfolio choices".

Matching with uncertain competitors' preferences

In our paper, the uncertainty on priorities endogenously emerges from a lack of information about competitors. It can be, therefore, connected with the wider literature on agent-agent matching with uncertainty on other agents' preferences.

The main technical difference is that in agent-agent matching with uncertainty on others' preferences, an agent's payoff depends on the types of other players only indirectly through the actions of the players. In agent-object matching with uncertainty on priorities, the payoff more generally depends both on the actions and on the types of the other players. In addition, there are no reporting issues: by definition, the market designer designs the priority system and perfectly observes the priority scores.

Roth (1989) [18] models agent-agent markets with private information on preferences on both sides: an agent only observes her own utility function and holds a prior distribution over the possible vectors of other agents' utilities. He shows that results on dominant and dominated strategies are similar to the standard results from the complete information benchmark. However, the results on Bayes-Nash equilibria are negative: for any mechanism, there exist some prior distributions for which at least some Bayes-Nash equilibria of the resulting game produce unstable matchings.

Kloosterman and Troyan (2020) [13] show that when preferences are uncertain but correlated, DA is no longer strategy-proof or stable, and less informed students are worse off due to a curse of acceptance (being accepted at a school signals that the school's quality is low). They show that priority design (so that any student is guaranteed a safe school) mitigates these issues.

Gleyze and Pernoud [9] model the information acquisition problem where agents on a matching market can learn on their own and other agents' preferences at a cost. Their main result is that a mechanism does not incentivize information acquisition on other agents if and only if it is *de facto* dictatorial.

Frictional matching

This paper more generally relates to the literature on frictional matching. The "Whether and Where to Apply" dilemma stems from the fact that truncating the Deferred Acceptance mechanism sacrifices the strategy-proofness (Haeringer and Klijn (2009) [10]).

The novelty in our approach comes from the fact that frictions add up and interplay to create novel strategic interactions arising within the agent side of the market. It combines some aspects of centralized matching (coordination in the timing of application) with some decentralized aspects (private information).

Multi-item auction

Because it features homogeneous preferences and priorities, the coming model also bears some similarities with a multi-item auction. In Demange, Gale, and Sotomayor (1986) [7], a collection of items is to be distributed among several bidders. All bidders rank items in the same way, and each bidder is to receive at most one item. The truncated Deferred Acceptance mechanism in our matching model and the generalized first-price or second-price mechanisms in the auction model have in common that they fail to be strategy-proof and that they result in the same allocation (when bids in the auction are consistent with priority scores in matching).

3 Model

3.1 Frictional market and Application Game

We consider a market with n agents and m objects ($m \leq n$). Agents (resp. objects) are numbered by $i \in \{1, \dots, n\}$ ($j \in \{1, \dots, m\}$).⁶ Agents have common cardinal preferences over objects: all agents assign the same objective value to each object, denoted a^j for object $j \in \{1, \dots, m\}$. The convention is set that object 1 (resp. m) is the highest (lowest) value object: $a^m < \dots < a^1$. Objects' preferences over agents (equivalently priorities) are also cardinally common: each agent i is characterized by a unique priority score, denoted ω_i . Priority scores are independently and identically distributed according to some (cumulative) distribution F on the unit interval: $\omega_i \sim F([0, 1])$, $i \in \{1, \dots, n\}$.

Information about priority is private: an agent i only knows her own priority score ω_i , but is ignorant of the priority scores ω_j of other agents $j \neq i$.

The allocation occurs through a truncated Deferred Acceptance mechanism, with truncation one. In this mechanism, agents are asked to independently and simultaneously choose whether to apply to an object and, if yes, to which object. Because priorities are common, the mechanism is equivalent to a serial dictatorship where the serial order is given by the score.

Application is costly, it costs $c < a^m$. If a given object receives no application, then it is wasted. If an object receives exactly one application, it goes to the single applicant. If an object receives at least two applications (crowding), the mechanism selects the agent with the highest priority score among the pool of applicants and endows this agent with the object.⁷ In this latter crowding case, we say that the agent who gets the object succeeds, while the other applicants fail.

A successful agent receives the value of the object she is assigned minus the application cost. An agent who has failed just pays the application cost, hence a negative utility. An agent who has chosen not to apply secures a reservation utility of zero.

This model poses a symmetric Bayesian Game of incomplete information that we call ‘‘Application Game’’ (AG). This game comprises n players, with action space $A_i = \{A^1, \dots, A^m, N\}$ - where A^j denotes the action of applying to object j and N ⁸ stands for the action of not applying -, privately known independent types, prior F over $[0, 1]$, and payoffs:

$$u_i(X_i, X_{-i}) = \begin{cases} a^j - c & \text{if } X_i = A^j \text{ and } \{l \in \{1, \dots, n\} \setminus \{i\} | X_l = A^j, \omega_l > \omega_i\} = \emptyset \\ -c & \text{if } X_i = A^j \text{ and } \{l \in \{1, \dots, n\} \setminus \{i\} | X_l = A^j, \omega_l > \omega_i\} \neq \emptyset \\ 0 & \text{if } X_i = N \end{cases}$$

where $u_i(X_i, X_{-i})$ denotes the payoff of player i when she plays action X_i and the rest of agents play according to action profile X_{-i} .

A pure strategy $s : [0, 1] \rightarrow \{A^1, \dots, A^m, N\}$ in the AG is a mapping from the interval of scores to the set of available actions (distributions over actions). A mixed or behavioral strategy $p : [0, 1] \rightarrow \Delta\{A^1, \dots, A^m, N\}$ is a mapping from the support of scores into the simplex of the action set: $p : [0, 1] \rightarrow \Delta\{A^1, \dots, A^m, N\}$. The probability $p_i^j(\omega)$, $j \in \{1, \dots, m+1\}$ stands for the odds that agent i chooses action A^j when her score is ω . A strategy is interior whenever the probability distribution is non-degenerate on more than a finite number of points, that is when $\exists \omega' < \omega \in [0, 1]$, $j \in \{1, \dots, m+1\}$ s.t. $\forall \omega \in [\omega', \omega] : 0 < p_i^j(\omega) < 1$.

A strategy can easily be represented graphically, displaying areas of scores where the agent chooses each action in one dimension ($[0, 1]$ line) for pure strategies and two dimensions ($[0, 1]$ square) for behavioral strategies.

⁶Notation: In all the following, numerals for agents (resp. objects) are written in indices (exponents).

⁷The mechanism does not specify how to break ties in case crowding happens between several agents with the same priority scores. But since F is a continuous probability distribution, ties occur with probability 0, hence no consequence on payoffs and equilibrium behaviors.

⁸Occasionally, to ease the notations in the rest of the analysis, we denote the action N as an additional application action, $A^{m+1} := N$, and define $a^{m+1} := c$.

3.2 Benchmark Market with perfect information

Throughout the analysis, we continuously refer to an alternative market design with public information. More precisely, priority scores are common knowledge.⁹ We call this design the “friction-less” or “benchmark” market.¹⁰ It defines a game with perfect information called a “Sorting Game.” The following table summarizes the difference between the two markets.

Design	Game	Information on priorities
Frictional market	Application Game	Private (observation of own score)
Benchmark market	Sorting Game	Public (common knowledge of scores)

FIGURE 1.1: Summary of two market designs

The comparison between the two designs, both in terms of predicted behaviors and subsequent welfare, will shed light on the most interesting features of the frictional market. By definition, the benchmark frictionless market is efficient, whereas the frictional market is inefficient. In section §5, we precisely characterize the inefficiencies associated with the frictions. The general idea is that providing public information on priorities, removing the truncation, or making the mechanism sequential would enable to capture the maximum welfare.

3.3 Leading example: Social Housing in Europe

A prominent example is the assignment of social housing units in Paris. Since 2016, the municipality has allocated around 4,500 housing units a year through an online scheme called “*LOC’annonces*”.¹¹ The allocation occurs in three steps. In the first step, households register as social housing seekers. The market operator performs eligibility checks and places households in rent and bedroom categories depending on their earnings and sizes. Most importantly, households are assigned priority scores based on their circumstances. The computation mode awards points for homelessness or unsuitable current housing (overcrowding), ill-health status, and more criteria. In the second step, households apply for vacant housing units. More specifically, vacancies are advertised on a dedicated website¹² from each Tuesday morning until the following Wednesday midnight. Households apply to one housing unit per round or choose not to apply. Very importantly, no precise feedback information is provided on the identities of other applicants to the same housing units. Following the application closing, applicants who have applied and who are on a shortlist of the highest priorities can view the accommodation and decide to maintain or withdraw their bids. In the third and final step, each vacant property goes to the applicant with the highest priority score among those who have applied for it. The whole allocation process, from the application closing to the final allocation, can take a maximum of three months.¹³ During the three-month period, households may miss attractive opportunities from the private sector.

The London social housing allocation scheme (“Choice Based Lettings Scheme”) much resembles the Paris scheme, except for the information. On the dedicated website,¹⁴ the application period runs from each Thursday morning until the following Monday midnight. During this period, when a household applies, she observes her position in a priority ranking of all current applications on her targeted housing unit. Through of trials and withdrawals, it is then possible to recover common knowledge on (the order of) priority scores.

⁹Common knowledge of the order of score is the most parsimonious information structure leading to the same equilibria.

¹⁰An alternative but equivalent design would consist in making the mechanism sequential rather than static. The market operator would organize the timing of applications by decreasing order of scores so that lower-ranked agents have the opportunity to observe the highest-priority agent assignments before submitting their own applications.

¹¹To be translated as “rental advertisements”.

¹²<https://teleservices.paris.fr/locannonces>.

¹³The example of social housing is, in fact, dynamic, with successive rounds on application. An agent who does not apply or fails in a given round is offered the opportunity to apply again in the next round. With a static model, we certainly miss this aspect. We address this issue in appendix §B.3 by introducing in the static model an endogenous cost of participation, which captures one of the main effects of the dynamic. The cost decreases with the priority, reflecting that high-priority agents keep their high score in successive rounds.

¹⁴www.homeconnections.org.uk.

3.4 Model justifications

Modeling a sub-optimal mechanism follows from the observation that market operators most often stick to the frictional design. Paris social housing is just one example. More precisely, the model features three frictions: the private information on priority scores, the truncation on the application list, and the application cost.

Private information stands as the default information structure on priorities, absent any intervention by the market operator. This is because, in general, the priority scores are computed based on criteria of private circumstances. The table below shows the scoring rule used in London social housing. FIGURE 2 below illustrates this point.

City of London Allocations Scheme		Secondary Points																
Primary Group	Primary Points	Overcrowding			Wellbeing				Unsuitable Housing Conditions				Housing Management					
		Per room lacking	Mixed sharing		Medical		Welfare		Sharing		Lack of tenancy	Bedroom Cap	Long TA stay	Advice & Engagement	Intentionality	Decant Urgency		
				S	M	L	S	M	L	F	1-4	5+						
Management Transfer	800																	
Under-occupation	400												50					100 / 200
Severe Medical / Welfare	275	25	10	50	25	10	50	25	10	5	10	15	5			15	minus 50	
Severe Overcrowding	250	25	10		25	10		25	10	5	10	15	5			15	minus 50	
Studio Upgrade	250	25	10		25	10		25	10									
Decants	225	25	10	50	25	10	50	25	10									100 / 200
Moderate Medical / Welfare	225	25	10		25	10		25	10	5	10	15	5				minus 50	
Moderate Overcrowding	200		10			10			10	5	10	15	5				minus 50	
Homeless	140	25	10	50	25	10	50	25	10						150		minus 50	
Lower Income City Connection	100					10			10	5	10	15	5					
Sons and Daughters	50					10			10									
Low Priority	1					10			10	5	10	15	5					

FIGURE 1.2: London social housing - Priority score computation rule

Households are awarded priority points if they are currently homeless, have large families, or have health issues. Household *A* knows about her status with respect to every criterion and, when given this table, can compute her aggregate priority score. However, household *A* could not make the same computation for another household, say *B*, simply because she had no idea about the inputs.

About the truncation, Pathak (2016) [17] reports that in school choice, the truncation is more often the rule than the exception. We propose two families of explanations: constraints and hidden objectives.

On the constraint side, Pathak (2016) suspects that truncation is used because it saves on operational costs that are usually unmodelled in theoretical matching papers. In school choice, reviewing students' records takes time. In social housing, organizing viewings of the accommodations also takes a lot of time. Universities and social landlords could try to limit the number of applications they receive to save on this time. With respect to private information, Roth and Sotomayor (1990) [19] note that many two-sided matching markets - in particular, entry-level labor markets - use decentralized application procedures, where agents from the same side of the market are isolated from each other, and information on preferences or priorities is subsequently private. It could also be that public information cannot be achieved due to privacy concerns. Finally, sequentiality is not a solution if there is a high opportunity cost of time (in social housing, this corresponds to the cost of vacancy).

With respect to social objectives, we argue that market operators may have different or additional objectives than just maximizing the ex-ante aggregate welfare. With respect to efficiency, they could be interested in using more sophisticated criteria such as the Pareto order (considering that a market design is superior if it leads to a higher payoff for all levels of scores). In addition to efficiency concerns, they could be interested in participation. Indeed, in many social landlords reports, we observe that statistics about high or increasing number of applications received are proudly announced.

In the baseline model, we impose an extreme truncation on the mechanism: agents can apply to at most one object. This is mostly to keep the model simple and tractable. In appendix §B.24, we show that our results are robust to a larger truncation.

The cost of application models the opportunity cost of time and effort dedicated to the application (reviewing available objects, sending application), sometimes adding up to an objective application or participation fee.

Finally, the homogeneous preferences are an extreme version of correlated preferences,¹⁵ and

¹⁵Homogeneous or correlated preferences are problematic empirically, hence interesting theoretically, because they intro-

mostly a first approach. In appendix §B.5, we show that our results are robust to imperfect correlation in the preferences.

Homogeneous priorities (as materialized with a unique priority score for each agent) fit the many contexts where the needs of the object side of the market are similar and are very standard in many public economic applications. For college admissions, the SAT score, as a weighted average of student performance in a number of maths, reading, and writing exercises, serves as a measure of a high school student's readiness for any college. In social housing, the priority score, reflecting the emergency of the household's housing need, controls the priority to any vacant housing unit. In the case of teacher allocation to schools, the same priority score applies to any school where a teacher can apply.

4 Equilibria

To study behaviors in the AG, we use Bayes-Nash equilibrium (BNE) as equilibrium concept.

4.1 Preliminaries

We make a few preliminary qualitative remarks on the structure of the problem and introduce the formalism.

Existence

We first state the existence of an equilibrium in the AG.

Lemma 1. *[Existence]*

There exists a Bayes-Nash equilibrium of the Application Game.

The proof is by the Bayes-Nash existence theorem for games with finite action space and independent types (potentially infinite type space) - Milgrom and Weber (1985) [16].

Interim expected payoffs

The interim expected payoff of player i under strategy profile p , when her priority score is ω , is denoted $\mathbb{E}[u_i(p)|\omega]$. Due to the dictatorship, in the AG, interim payoffs depend on the strategy of other agents through the behaviors of higher score agents, yet are independent of the behaviors of lower score agents. Trivially, the interim payoffs also depend only on the agent's strategy through the agent's behavior at the set score and not at higher or lower scores. Thus in the formalism for interim payoffs, it is enough to specify for p just $p_i(\omega)$ for the agent, and $p_{-i}([\omega, 1])$ for the other agents: $\mathbb{E}[u_i(p)|\omega] = \mathbb{E}[u_i(p_i(\omega), p_{-i}([\omega, 1]))|\omega]$.

At BNE p^* , we have:

$$\forall i \in \{1, \dots, n\}, \forall \omega \in [0, 1] : p_i^*(\omega) \in \arg \max_{p_i(\omega)} \mathbb{E}[u_i(p_i(\omega), p_{-i}([\omega, 1]))|\omega]$$

In the case where the agent applies with full probability to one object at score ω , the interim payoff is a weighted sum of two ex-post payoffs: the (positive) object value and the (negative) application cost, where the value is weighted by the conditional probability of success, denoted $\mathbb{P}(S|p, \omega)$:

$$\mathbb{E}[u_i(A_i^j, p_{-i}([\omega, 1]))|\omega] = \mathbb{P}(S|p, \omega)a^j - c$$

The following lemma characterizes this interim expected payoff:

Lemma 2. *[Interim payoff after pure action - Characterization]*

- (i) Continuity: Interim payoff conditional on any action is continuous in the score.
- (ii) Monotonicity:

duce competition within the sides of the matching market.

- *Interim payoff conditional on not applying is constant equal to 0.*
- *Interim payoff conditional on any application action is increasing (constant) on any interval of scores where at least one other agent (no other agent) applies to the same object with a positive probability.*

(iii) *Value at highest bound: Interim payoff at score 1 conditional on applying to any object is equal to the value of the object minus the cost.*

Statement (i) states that there is no jump in the interim payoff: one's chance to get an object, hence one's payoff cannot dramatically change from one score to a neighbor score. Statement (ii) says that one's payoff increases when one's score rises if and only if the agent was facing crowding on the targeted object by agents with scores slightly above. Statement (iii) formalizes that the highest score agent is successful for sure.

The three statements are direct consequences of the serial dictatorship mechanism and the continuity of the priority score support.

Gross incentive analysis

It is well known that once truncated, the Deferred Acceptance mechanism is no longer strategy-proof (Haeringer and Klijn (2009) [10]). The typical preference manipulation consists of applying to "safe" objects (objects that accept the agent with high probabilities). It materializes in different ways at different levels of scores.

A player with a score close to 1 should feel confident that when applying to any object and in case of crowding, she will succeed. She should target high-value objects, fully accepting the prospect of competition.

Conversely, a player with a score approaching 0 should expect that when applying to any object and in case of crowding, she will fail. This low-score agent seeks to avoid competition and to coordinate with peers so as to target different objects: she should target under-demanded objects. Even more than that, she may be tempted not to apply to guarantee a utility of zero. Any agent with an intermediate score faces a trade-off between applying to high-value objects and risking failure or to under-demanded objects (or even giving up) and settling for a low (zero) satisfaction.

We illustrate this discussion in the graph below.

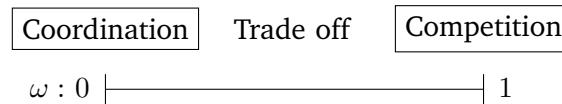


FIGURE 1.3: Coordination and competition behaviors in the AG as a function of priority scores

4.2 Example

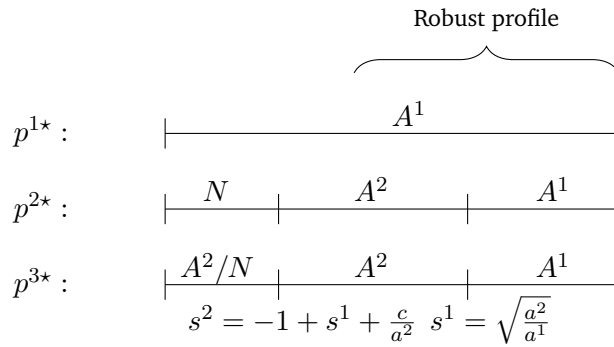
As a prelude for the general results, we display the equilibria in an example where dimensions are small ($n = 3 > m = 2$), and the distribution is uniform ($F \sim \mathcal{U}([0, 1])$). We provide a graphical representation, a description, and the intuition.

Pure (asymmetric) equilibrium

On the graph below,¹⁶ each line going from 0 to 1 stands for the score support $[0, 1]$, one line for each strategy of the three players, and the letters above stand for the action played at the corresponding scores.

The bracket "robust profile" specifies the part of the profile that is realized at the pure strategy equilibrium of any AG.

¹⁶We deal with the asymmetry by numbering players and assigning each of them to a specific role. Yet any permutation of strategies between players is again an equilibrium.

FIGURE 1.4: Pure (asymmetric) BNE - $n = 3$, $m = 2$, $F \sim \mathcal{U}$

The graph shows three intervals, where the intervals' bounds s^2 , s^1 are indifference points.

On an interval of high scores $[s^1, 1]$, all agents apply to object 1. This is because an agent with the highest possible score 1 always gets the object she has applied to. So, she applies to the highest value object and secures the highest possible payoff in the game $a^1 - c$.

At score s^1 , confidence in success in case of crowding when applying to object 1 becomes quite low. By contrast, the probability of success conditional on applying to object 2 is constant equal to 1. Algebraically, the interim payoff of applying to object 1 hits the value of object 2 minus cost $a^2 - c$. At s^1 , agents become indifferent between applying to objects 1 or 2.

On an interval of intermediate scores $[s^2, s^1]$, two agents (agents 2 and 3) apply to object 2. Because they compete, their interim payoff steadily decreases from s^1 leftward. The other agent (agent 1) keeps on applying to object 1. From her point of view, there is no competition anymore on object 1; her interim payoff is constant on the whole interval $[s^2, s^1]$.

The fact that agents share roles (with a majority of applicants to object 2 and a minority of applicants to object 1) breaks the possibility of a symmetric equilibrium in pure strategies. To get an intuition on the necessity of asymmetry, we can consider (by contradiction) a symmetric strategy profile where all players would shift to apply to object 2 below s^1 . Then, the interim payoff conditional on playing object 1 would be constant (no competition) as the score decreases below s^1 , whereas the interim payoff conditional on playing object 2 would increase (2 competitors). Consequently, the former would be higher than the latter at any score below s^1 , and any player would face a profitable deviation from object 1 to object 2.

At score s^2 , the interim payoff of agents 2 and 3 hit the zero bound. It becomes profitable for one of them (say agent 2) to deviate to N , to secure a payoff of zero. Conditional on that, agent 3 is indifferent between maintaining action A^2 with no competition or playing action N (both deliver the same zero payoff). Agent 1 keeps on playing object 1 on the whole interval $[s^2, s^1]$. She still faces no competition on object 1, hence a constant interim payoff.

The bracket “robust profile” shows that only the right side of the graph is realized at the pure strategy equilibrium of any AG. For some sets of parameters, we may observe only the top part of the profile, and not the bottom part. Agents with a score of 0 may all apply (potentially only to the two best objects).

Symmetric (interior) equilibrium

In the following graph, the horizontal line in the square still represents the score support $[0, 1]$. The vertical line represents the probabilities of each action in the behavioral strategy.

For example, $p_{[1,2]}^2$ denotes the probability with which an intermediate score agent applies to object 2.

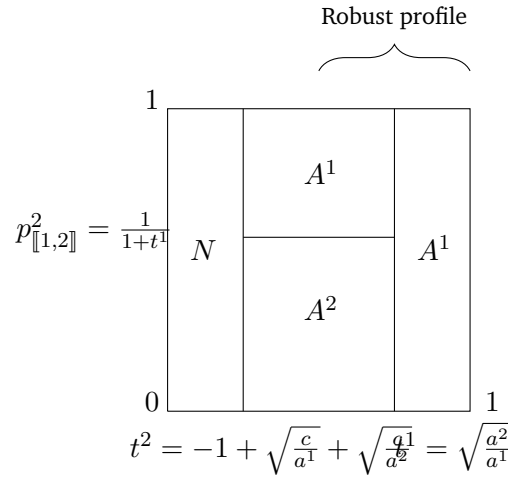


FIGURE 1.5: Symmetric (interior) BNE - $n = 3$, $m = 2$, $F \sim \mathcal{U}$

As in the pure strategy equilibrium, all agents apply to object 1 at scores belonging to an interval $[t^1, 1]$.

The equilibrium strategy becomes truly interior on an interval $[t^2, t^1]$, where an agent applies with positive probabilities to objects 1 or 2. What is striking here is that those probabilities are constant. This is what we call the "block structure of the equilibrium." In section §4.3.2, we discuss this central result.

Notably, intermediate score agents apply more often to object 2 than to object 1 ($p_{[1,2]}^2 > \frac{1}{2}$).

At some low score t^2 , due to competition by higher-ranked agents, the interim payoff of playing A^1 and A^2 hits the zero bound. Agents choose not to apply at any score below. Symmetry does not allow one player to keep on applying, unlike what happens in the pure case.

The bracket "robust profile" shows that for some parameters of the AG, all agents apply with probability one (potentially only randomizing on the two best objects).

Summary results from example

In both the pure and interior equilibria, confidence in success makes high-priority agents ambitious. At intermediate scores, it becomes rewarding to be less ambitious (say "practical") and to try to coordinate to avoid competition.

In the pure equilibrium, this happens through a sharing of roles between applicants to different objects; in the interior equilibrium, by positive probabilities to apply to both objects. Many intermediate score agents settle for the secure option. A remaining smaller group of intermediate score agents take advantage of alleviated competition to maintain high ambitions. At the lowest possible scores, agents may need to shift to the no application action so as to secure a positive payoff.

Although the structure of the equilibrium is very robust, whether the possibility to abstain or apply to low-value objects is used at equilibrium depends finely on the parameters of the AG.

4.3 General results

The analysis from the example generalizes to any number of agents n , objects m , and any priority score distribution F .

Nash equilibrium of the Sorting Game

Proposition 0. [(Unique) NE]

In the Sorting Game, there is a unique Nash equilibrium σ^* , where:

- (i) The agent ranked i^{th} , $2 \leq i \leq m$ in priority applies to and is allocated object ranked i^{th} in value.
- (ii) Agents ranked i^{th} , $m + 1 \leq i \leq n$ in priority do not participate.

The proof is done by induction, following the priority order for agents, which is also the serial order used in the dictatorship mechanism.

On the benchmark market, agents are able to perfectly tailor their ambitions to their ranks in the priority order and endogenously sort. The highest priority agent (second-highest priority agent) applies to the highest value object (second-highest value object), and so on. If there are strictly more agents than objects, the lowest score agents do not apply to avoid certain failure. In total, each available object receives exactly one application, which is accepted. No object is wasted, and no agent fails.

Pure (asymmetric) Bayes-Nash equilibrium of the Application Game

Proposition 1. *[Asymmetry of pure BNE]*

A pure strategy Bayes-Nash equilibrium of the Application Game is necessarily asymmetric.

The proof is done by contradiction, just as in the example. The short intuition is that competition by high-score agents on the highest-value objects smooths interim payoffs conditional on different applications. Consequently, at any intermediate or low score, several objects of different values are equally attractive. To guarantee the absence of profitable deviation, they all need to be targeted by at least one agent with positive probability. A pure symmetric profile would not allow this.

In the rest of the paper, we discard the BNE in pure strategies for two reasons. First, we remain skeptical about the capacity of (ex-ante symmetric) agents to coordinate and share the different roles in an asymmetric profile without any communication. Second, the pure strategy BNE structure is little robust. In the general model with any number of agents, objects, and any distribution, it depends very finely on the set of parameters of the game.¹⁷

Symmetric (interior) Bayes-Nash equilibrium of the Application Game

We find that the symmetric has a very specific structure, which we illustrate on the next figure:

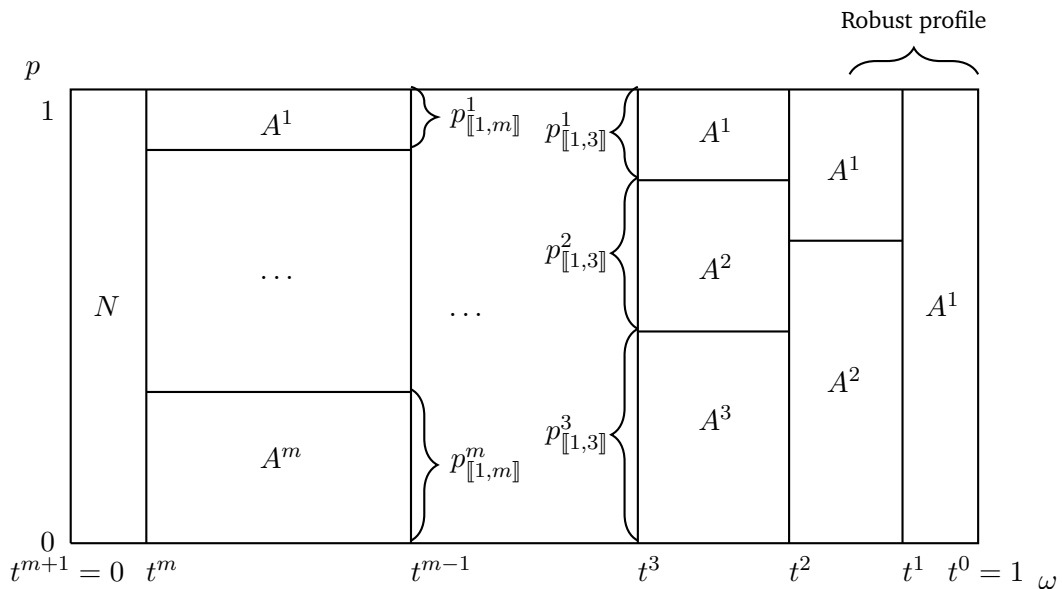


FIGURE 1.6: Symmetric (interior) BNE of the AG – General form

Theorem 1. states the existence and uniqueness of a symmetric BNE and describes the "block structure."

¹⁷We illustrate this lack of robustness in appendix §B.1.

Theorem 1. [Symmetric (interior) BNE]

A symmetric (interior) Bayes-Nash equilibrium p^* of the Application Game:

- (1) Exists and is unique.
- (2) Exhibits a "block structure," meaning that there is a finite number of intervals of scores, called "classes," where the interim strategy profile at any score is the same:
 - (i) There are between 2 and $m + 1$ classes: $k_0(p^*) \in \{2, \dots, m + 1\}$.
More precisely, there are exactly:

$$\begin{aligned}
 & - k, k \in \{2, \dots, m\} \text{ classes iff: } 1 + \sum_{l=1}^k \left(\frac{a^{k+1}}{a^l}\right)^{\frac{1}{n-1}} \leq k < 2 + \sum_{l=1}^{k-1} \left(\frac{a^k}{a^l}\right)^{\frac{1}{n-1}} \\
 & - m + 1 \text{ classes iff: } m < 1 + \sum_{l=1}^m \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}}
 \end{aligned}$$

- (ii) Conditional on existence, classes write $[t^k, t^{k-1}]$, with:

$$\begin{aligned}
 \forall k \in \{1, \dots, k_0(p^*) - 1\} : t^k &= F^{-1}\left(1 - k + \sum_{l=1}^k \left(\frac{a^{k+1}}{a^l}\right)^{\frac{1}{n-1}}\right) \\
 t^{k_0(p^*)} &= 0
 \end{aligned}$$

- (iii) Conditional on existence:

- In class $k \in \{1, \dots, m\}$, agents apply to object $j \in \{1, \dots, k\}$, with probability:

$$p_{[[1,k]]}^j := \left(\sum_{l=1}^k \left(\frac{a^j}{a^l}\right)^{\frac{1}{n-1}}\right)^{-1}$$

- In class $m + 1$, agents do not apply.

From now on, the bounds of the classes t^k are "thresholds," and the constant probabilities $p_{[[1,k]]}^j$ are "levels".¹⁸

The proof of Theorem 1. is done in three steps.

In the first step, we prove that we can divide the score support $[0, 1]$ in a finite number of intervals where agents with scores in one interval apply with positive probabilities to only object 1, then both object 1 and 2 until an interval where they potentially apply to all objects and a bottom interval with no application. This step relies heavily on Lemma 2., jointly with the intermediate value theorem applied recursively m times. The proof that t^1 necessarily exists is made by contradiction, exactly as in the sketched proof of Proposition 1..

The next step is to characterize the probability functions. The proof that they are piecewise constant is done by induction. At inductive step $k \in \{2, \dots, m - 1\}$, the strong indifference principle applied at a score $\omega^* \in (t^k, t^{k-1})$ delivers a system of $k - 1$ differential equations with $k - 1$ unknowns ($p^j(\omega^*), j \in \{2, \dots, k\}$). Substituting within the equations, we find a relation between the primitives of $f p^j$ and f , hence constant probabilities $\forall \omega \in [t^k, t^{k-1}]$, $p^j(\omega) := p_{[[1,k]]}^j$. Meanwhile, we use the differential equations again to get a recursive relation between all $p_{[[1,k]]}^j$, $j \in \{1, \dots, k\}$, and use the fact that they sum up to one to get the explicit formula.

In the third and final step, we use the thresholds' definitions (t^k , $k \in \{1, \dots, m\}$ is the highest score where there is indifference between all actions A^1, \dots, A^{k+1}) to find their expressions.

The symmetric equilibrium in small (§4.2) and general dimensions displays the same pattern, and the overall interpretation (intermediary between cooperation game at low scores and conflict game at high scores) is similar. However, the general version sheds light on two interesting equilibrium features deserving a dedicated discussion, in the coming paragraphs.

¹⁸Appendix §B.2 provides comparative statics showing how the thresholds and levels vary with the parameters. In particular, it shows that when the distribution of scores changes, the levels remain unchanged, and only the width of the classes adapts so that the expected mass of agents belonging to each class remains constant.

Block structure and sorting

This block structure of the symmetric equilibrium may appear surprising at first glance. From a theoretical perspective, it says that within a very large strategy space made of potentially uneven probability functions, agents effectively use a small number of application mixtures. Starting from a continuous type support, we end up with a discrete number of equilibrium behaviors. From an applied perspective, it means that agents self-sort according to their priority scores into a finite and quite small number of classes. Two households with close priority scores (in the same class) use exactly the same (potentially highly sophisticated) application mixture. We could have rather expected that higher priority agents would be strictly more ambitious than lower priority agents.

The algebraic necessity of the block structure is clear enough. It comes from the following observation (made on class $[t^2, t^1]$, to fix ideas). The indifference principle, when applied at a score slightly below threshold t^1 (say $t^1 - \epsilon$) when indifference at t^1 is already established, delivers exactly the same constraint on the probability function as when applied slightly leftward at a score $t^1 - 2\epsilon$, when indifference at $t^1 - \epsilon$ is already established. This constraint writes as the ratio of probabilities of success conditional on applying to different objects is equal to a constant. In both cases, it is thus expected to deliver the same level $p_{[1,2]}^2$. A more direct intuition combines the notions of ambition and risk. The class defines the strategy, hence a constant level of ambition in each class. By contrast, the risk (as measured by the probability of failure) varies within a given class: it is high (low) just above (below) the thresholds. Therefore, for agents at the bottom of a class, the equilibrium strategy features a given level of ambition and is risky; hence a low payoff. For agents at the top of the same class, the equilibrium strategy features the same level of ambition but a low risk, hence a high payoff. The risk and payoff variations are smooth in between. We illustrate this line of intuition below with a figure eliciting the variations of risk and ambition within and between classes.

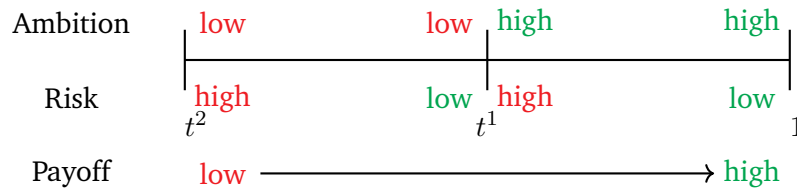


FIGURE 1.7: Ambition, risk and interim payoffs at symmetric BNE of the AG

In total, the block structure is reminiscent of the "class segregation result" in the dynamic search problems. The dynamic search literature (McNamara and Collins (1990) [14], Burdett and Coles (1997) [4], Bloch and Ryder (2000) [3], Jacquet and Tan (2007) [11]) studies two-sided agent-agent markets where each agent is characterized by a value distributed on a continuous support. At each time period, agents are tentatively matched, they observe each other's values and decide to accept or reject the proposed match. At equilibrium, agents sort into a finite number of classes (value intervals) where all agents use exactly the same acceptance cutoff and match within classes. In their case, the equilibrium with class segregation is in pure strategies. In our case, the block structure is even more surprising as agents use sophisticated behavioral strategies.

A related interesting question is whether, at equilibrium, agents are able to self-sort, conforming their ambitions to their scores. It matters in relation to the frictionless benchmark, where the equilibrium outcome is the perfectly positive assortative matching. In our environment with private information, truncation, and application cost, we would like to know whether some assortativity remains. The following corollary answers in a positive way. There is a kind of sorting, although, by the block structure, it proceeds with discrete jumps.

Corollary 1. [Sorting at symmetric BNE]

At the unique symmetric (interior) Bayes-Nash equilibrium of the Application Game:

- (i) For a given object $j \in \{1, \dots, m\}$ and two scores $0 \leq \omega < \omega' \leq 1$ where the object receives applications, it receives more applications at the higher score ω' .

- (ii) At any two scores $0 \leq \omega < \omega' \leq 1$, the probability level vector at the higher score $p(\omega')$ first-order stochastically dominates the same vector at the lower score $p(\omega)$.
- (iii) For a given score $0 \leq \omega \leq 1$ and two objects $1 \leq j < j' \leq m$ that receive applications at this score, the lower value object j' receives strictly more applications.
- (iv) In total, conditional on existence of class m , the ex ante probability to apply to object k is $1 - \left(\frac{c}{a^k}\right)^{\frac{1}{n-1}}$.

Statement (i) states that as the score increases, agents apply more and more frequently to high-value objects.

Statement (iii) uses the standard criterion of First Order Stochastic Dominance (FOSD).¹⁹ What it means is that the probability of playing the highest value object is increasing in the score, the cumulative probability of playing one of the two highest value objects is also increasing in the score, and so forth. Interestingly, statement (iii) also claims that within class $[t^2, t^1]$ (for example), agents apply more often to object 2 than to object 1 (as displayed on FIGURE 6). Hence, the equilibrium appears closer to a pure strategy equilibrium with perfect vertical sorting than to a single-class block equilibrium with no sorting.

Statement (iv) computes the ex-ante probability with which an agent applies to a given object, giving a sense of how bankable an object is. It finds that it is increasing with the value of the object a^k (higher-value objects are played more often), decreasing with the application cost c , independent of other objects' values a^j , $j \neq k$ and of the score distribution.

Robust profile and participation

Theorem 1. (2)(i) ensures that the two highest classes $[t^1, 1]$ and $[t^2, t^1]$ are realized at symmetric equilibrium of any AG. Yet, it does not guarantee that any lower class is reached. For some sets of parameters (low cost, low and balanced number of objects and agents, heterogeneous values), classes 3 to $m + 1$ (interval $[0, t^2]$) may not be observed ($t^2 = 0$), and low score agents may apply to objects 1 and 2 only. This is the meaning of the vocabulary “robust profile” for $[t^2, 1]$.

In particular, in the extreme case with only 2 classes, this suggests that agents with very different scores (distant from about a half) use exactly the same randomization over actions.

The robust profile bracket also characterizes participation in the mechanism. We say that an agent (fully) participates if she applies with positive (one) probability. The next corollary discusses participation in the AG.

Corollary 2.

At the unique symmetric (interior) Bayes-Nash equilibrium of the Application Game:

- (i) If $m \geq 1 + \sum_{l=1}^m \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}}$, all agents on the market participate.
- (ii) If $m \geq \frac{n}{n-1} \sum_{l=1}^m \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}}$, expected participation is higher than in the Nash equilibrium of the Sorting Game.

Statement (i) states that whenever class $[0, t^m]$ is not reached, all agents on the market - including the ones with the lowest possible scores-, fully participate. They do so in spite of the congestion (the fact that there are fewer objects than agents on the market). This is a major contrast with the benchmark design, where the agents with the lowest possible scores do not apply.

The more general statement (ii) provides a sufficient condition for higher participation in the frictional market. The condition is easily satisfied, as soon as n is not too small or values are quite heterogeneous, or the cost is low.

The lesson is that frictions enhance participation. This has major welfare consequences, which we explore in the next section §5.

¹⁹First-order stochastic dominance translated to our (discrete) case is given by: $p(\omega) \succeq_{FOSD} p(\omega')$ if $\forall j \in \{1, \dots, m-1\}$, $\sum_{l=1}^j p^l(\omega) \geq \sum_{l=1}^j p^l(\omega')$.

5 Welfare

In this section, we compute the welfare at the symmetric equilibrium of the AG (§4.3.2). We compare this welfare from the fictional market to the welfare on the benchmark frictionless market.

5.1 Ex ante

We first examine the ex-ante payoff, that is, for an anonymous agent, before her priority score is realized. It is a measure of aggregate welfare.

Inefficiencies

The ex-ante payoff on the benchmark market is given by:

$$W^B := \mathbb{E}[u(\sigma^*)] = \left(\sum_{k=1}^m a^k \right) - cm$$

There is no waste (agents collectively capture the whole sum of objects' values) and no failure (the number of agents paying the cost equals the number of applicants equals the number of objects (m)).

The ex-ante expected payoff on the frictional market²⁰ is given by:

$$W^F := \mathbb{E}[u(p^*)] = \frac{1}{n} \left(\sum_{k=1}^m a^k \right) - \frac{c}{n} \left(m - (n-1) \sum_{j=1}^m \left(\frac{c}{a^j} \right)^{\frac{1}{n-1}} \right)$$

Therefore, the difference in welfare is:

$$W^B - W^F := c(n-1) \left(m - \sum_{j=1}^m \left(\frac{c}{a^j} \right)^{\frac{1}{n-1}} \right)$$

Proposition 2. [ex ante welfare]

(B) The benchmark market design is efficient.

(F) The frictional market is inefficient: $W^B - W^F > 0$.

The size of the inefficiency increases with the values of all objects and with the cost of application.

(B) recalls that, by definition, the frictionless design achieves a higher total welfare than the frictional design. The comparative statics in (F) is straightforward. When an object's value increases, waste on this object is more detrimental to welfare. When the application cost increases, failure is also more consequential.

5.2 Interim

We push the characterization of the inefficiencies identified in section §5.1. We are not only interested in the size, but also in the shape of the inefficiencies.

Interim expected payoffs

The interim expected equilibrium payoff in the benchmark market is given by:

$$W^B(\omega) := \mathbb{E}[u(\sigma^*)|\omega] = \sum_{j=1}^m \binom{n-1}{j-1} (1 - F(\omega))^{j-1} F(\omega)^{n-j} (a^j - c)$$

²⁰The formula is written for the case $t^m \geq 0$. For case $t^m < 0$, the formula is more sophisticated but Proposition 2. remains valid.

It is just the sum of probabilities that the agent with score ω is ranked i^{th} , $i \geq m$ in priority multiplied by the ex-post payoff in this case (object value minus application cost). If the agent is ranked lower (i^{th} , $i > m$), the ex-post payoff is zero.

The interim expected equilibrium payoff in the frictional market is given by:

$$W^F(\omega) := \mathbb{E}[u(p^*)|\omega] = \begin{cases} \sum_{j=1}^k p_{[[1,k]]}^j \cdot \left(\left(\frac{a^k}{a^j} \right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j \right)^{n-1} a^j - c \\ , \omega \in [t^k, t^{k-1}], k \in \{1, \dots, m\} \\ 0, \omega \in [0, t^m] \end{cases}$$

At any score where the agent applies, it is an expected sum. $p_{[[1,k]]}^j$ is the probability that the agent applies to object j , it multiplies the probability of success conditional on applying to object j . To succeed, one needs that each of all other agents ($n-1$) does not apply to the same object or has a lower score, which happens with probability $\left(\frac{a^k}{a^j} \right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j$.²¹

How the interim expected payoff varies with the score indicates the effect of priority on individual welfare. If the priority score system fulfills the role of discriminating between agents with different levels of priority, the interim expected payoff should increase with the score.

Lemma 3. [Equilibrium interim payoffs]

(B) On the benchmark market, the interim equilibrium payoff is continuous and strictly increasing with priority score ω .

(F) On the frictional market, the interim equilibrium payoff is continuous and:

- Constant on $[0, t^m]$ ($t^m > 0$).
- Strictly increasing on $[t^m, 1]$

The proof of (B) is from the formula above, plus the continuity and monotonicity of the cdf F . The proof for (F) just stems from the initial characterization of interim payoffs in Lemma 2. and the symmetric equilibrium structure from Theorem 1..

Overall, in both kinds of markets, higher score agents are always better off. This means the block structure (featuring constant ambition on each class of scores) still allows for continuously and strictly increasing payoffs. For illustration, we display below two graphs for interim expected payoffs on the frictional market, in the context of the example from section §4.2: one with $k_0 = 2$ classes, one with $k_0 = 3$ classes.

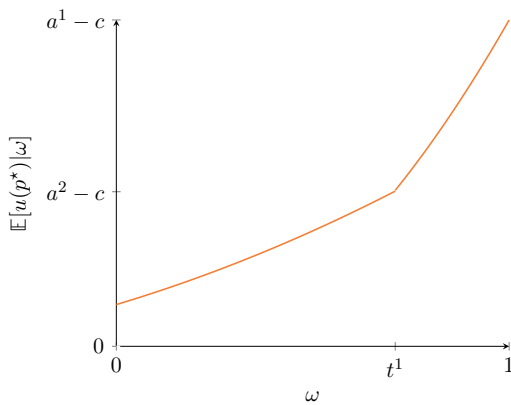


FIGURE 1.8: Interim expected payoff at the symmetric (interior) BNE - $m = 2$, $n = 3$, $a^2 = 2$, $a^1 = 4$, $c = 0.2$, $F \sim \mathcal{U}$

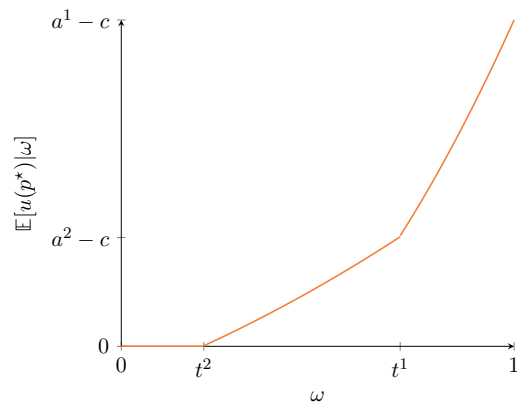


FIGURE 1.9: Interim expected payoff at the symmetric (interior) BNE - $m = 2$, $n = 3$, $a^2 = 2$, $a^1 = 4$, $c = 1$, $F \sim \mathcal{U}$

²¹The fraction $\left(\frac{a^k}{a^j} \right)^{\frac{1}{n-1}}$ simplifies the whole probability of an agent having a score in a higher class and applying to the same object j .

Asymmetries

FIGURE 8 specifically shows that on the frictional market, when $k_0 < m$, agents with low scores apply and expect a positive payoff. In the benchmark market (Proposition 0.), by contrast, low-score agents decide not to apply to secure a zero payoff.

This is a clue that there is a score-related asymmetry in the way the inefficiencies associated with private information affect market participants. The next proposition shows that this asymmetry can make low-score agents prefer the situation when everyone has less information.

Proposition 3. [Low score agents' preference for private information]

If $m > 1 + \sum_{l=1}^k \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}}$, then $\exists \omega' \in (0, 1)$ s.t.: $\forall \omega \in [0, \omega')$, $W^B(\omega) < W^F(\omega)$.

Thus, in some instances of the AG, low-score agents are better off with less (private) information than with full (public) information. The condition on parameters looks quite general, easily satisfied if the number of objects is large enough and the application cost is reasonably low.²² One important implication is that we cannot rank the two benchmark and frictional designs through a Pareto order.

The proof uses the definition of t^m as the indifference point between application actions and the no application action, hence an expected payoff of zero, jointly with continuity and strict monotonicity of interim payoff above t^m (Lemma 2.). When $t^m < 0$, the monotonicity mechanically induces a positive interim payoff at score zero, to compare with an interim payoff in the benchmark design of 0. The continuity of payoffs extends the comparison to a non-degenerate interval of low scores $[0, \omega')$.

A graphical illustration is given below. FIGURE 10 (left) shows the interim expected payoffs from the frictional market in orange, and the interim expected payoffs from the benchmark market in green. On FIGURE 10 (right) the filled orange area displays the difference in interim payoffs $W^B(\omega) - W^F(\omega)$.

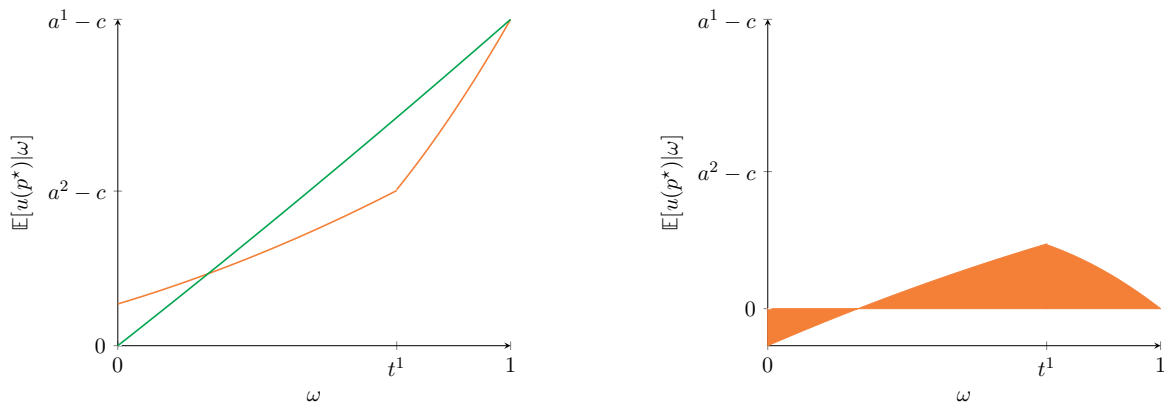


FIGURE 1.10: Frictional vs. benchmark markets: Interim expected payoffs at symmetric (interior) BNE
 $m = 2$, $n = 3$, $a^2 = 2$, $a^1 = 4$, $c = 0.2$, $F \sim \mathcal{U}$

A more general statement from this graph is that intermediate score agents bear most of the burden associated with the inefficiency of private information.

Signal quality, competitive advantage, and competition easing

We formulate the intuition in successive effects of private information on cardinal scores: discrepant signal quality, competitive advantage, and competition easing.

In the application game, although all agents receive signals of the same nature (perfect observation of their own priority score), the informative value of the signal depends on the score level. A low-score agent, by observing a signal at the bottom of the prior distribution support, realizes that she is almost surely the lowest type, ranked at the bottom of the priority order. By contrast, for an intermediate

²²Notably, it seems to match our empirical driving empirical motivation of social housing.

score agent, observing her score is a poor signal of her ranking. But ranking or ordinality of scores (as opposed to cardinality) is what decides on the allocation in case of crowding. When it comes to information, the low-score agent benefits from a relative competitive advantage. This advantage more than offsets the fact that, in absolute terms, the private information they get is slightly less revealing. It translates into actions in the following way. The intermediate score agent, misguided by imprecise posterior on ranking, is likely to make mistakes (compared to what she would do with perfect information): being too or too little ambitious, miscoordinating with intermediate-to-high score agents. In total, and due to the recursive structure of the AG, these mistakes tend to alleviate the fierceness of competition by intermediate score agents.

Thanks to the precision of her posterior, the low-score agent avoids any mistake due to unrealistic ambitions. More importantly, she is fully aware of the ease of competition. Therefore, she understands that this leaves some room for her to apply. She applies and captures a positive expected payoff.²³

The result that frictions favor participation and equity is very general and can be reconstructed in many different economic settings. For example, Mekonnen (2019) [15] compares random and directed searches on an agent-object market with also common preferences but homogeneous agents. This is equivalent to comparing a no information design to a full information design. At equilibrium, an agent is better off under the random search because she benefits more from the ease of congestion on high-value objects than she suffers from not being able to target objects accurately. Che and Tercieux (2021) [6] study the optimal design of a queueing system when agents' arrival and servicing are governed by a general Markov process. They show the optimal information is no information (beyond recommendation to join, stay in, or leave the queue). The intuition is that no information pools the various incentive constraints, ensuring more participation, which, in the queue environment, increases efficiency.

Implications for decentralized matching markets

We have found that in a very stylized frictional market, private information on priorities tends to favor low-score agents.

The question therefore arises: is it an issue for the proper functioning of markets?

The answer to that question rests on the role of priority score systems. In social housing, for instance, the priority system recognizes differential rights to housing based on different levels of emergency. During the allocation stage, it maps those differential rights into proportionally different probabilities of satisfaction. In particular, point rules and allocation mechanisms are often jointly thought of as giving twice more chances to an agent with a score 2ω of getting an object or the best object than to an agent with a score ω .

Yet, in the considered design, private information artificially distorts in an increasing fashion the probability for low-score agents to get a valuable object with respect to intermediate and high-score agents. In other words, private information mitigates the ability of the priority system to discriminate between agents.

From a design perspective, we have found that three elements matter for the final allocation and for welfare: the allocation mechanism, the inputs in the mechanism (the exact rule used to compute priority scores), and the information structure. Our analysis suggests that they should be designed jointly in order to keep control over the amount of discrimination on the market.

²³ A wonder may be: how does private information affect the competition between high-score agents and intermediate-score agents? A high-score agent observes a score close to the higher bound of the support, which is a clear signal that she is likely the highest type, so this agent also benefits from a relative informational competitive advantage. But because she always plays the same action (applying to the best object) in the two designs, she cannot exploit this advantage, and there is no competition easing effect from which the intermediate score agent could benefit.

6 Conclusion

This paper models a stylized market where agents with homogeneous preferences and privately known priority scores can (costly) apply to at most one object, and each object is assigned to its highest priority applicant. In this market, the frictions (private information, truncation of the mechanism, and application cost) ask for a trade-off between competition and coordination, with participants wondering "Whether and Where to Apply?". They consider the trade-off between being "ambitious", accepting the prospect of competition (targeting high-value objects), or being "practical", seeking coordination (targeting under-demanded objects).

We find that in all equilibria, high-score agents are ambitious, and low-score agents are practical. The analysis also uncovers three salient and surprising features of the symmetric equilibrium. One, the symmetric equilibrium necessarily involves agents randomizing between applications. Second, in this equilibrium, agents with scores on a continuous support sort into discrete classes, defined as intervals of priority scores, where they adopt exactly the same strategy. Third, the frictional market design is less efficient but more egalitarian than the benchmark design. Indeed, low-score agents may be better off with private information than with public information because they benefit from a relative informational competitive advantage.

The value of this work is two-fold. It illustrates the role of information on priorities matching markets, showing how uncertainty interplays with other standard market frictions to distort the allocation. This results in a straightforward design recommendation that should be applied broadly: the mechanism design, information design, and priority design should be performed jointly in order to achieve the desired pattern of discrimination. It also makes a methodological contribution by displaying a novel and rich mode of strategic interactions arising within the agent side of a matching market (the application game), resulting in an equally novel equilibrium structure (the block structure).

We believe the lessons learned from this work may generalize to more sophisticated (idiosyncratic, hybrid) preferences and mechanism (dynamic, with deterministic and stochastic stages). Ultimately, our framework should be able to accommodate more numerous and interesting empirical applications: beyond social housing, any market where priority is defined by a cardinal point system (teacher allocation, college admissions in many countries).

Another interesting challenge in this research is to provide micro-foundations for the use of a sub-optimal mechanism. Ideally, we would explicitly model the general design problem and show that once we consider a two-fold social objective (we care not only about efficiency but also about participation), the private information mechanism becomes optimal. Another foundation could come from political economy. In our applied example, the market frictions give a chance to middle-class households to be allocated social housing. Using a frictional design could be a way for a greedy politician to earn these households' votes for reelection.

References

- [1] S. N. Ali and R. I. Shorrer. The college portfolio problem. Technical report, Working Paper, 2021.
- [2] C. Avery and J. Levin. Early admissions at selective colleges. *American Economic Review*, 100(5):2125–2156, 2010.
- [3] F. Bloch and H. Ryder. Two-sided search, marriages, and matchmakers. *International Economic Review*, 41(1):93–116, 2000.
- [4] K. Burdett and M. G. Coles. Marriage and class. *The Quarterly Journal of Economics*, 112(1):141–168, 1997.
- [5] H. Chade, G. Lewis, and L. Smith. Student portfolios and the college admissions problem. *Review of Economic Studies*, 81(3):971–1002, 2014.
- [6] Y.-K. Che and O. Tercieux. Optimal queue design. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 312–313, 2021.
- [7] G. Demange, D. Gale, and M. Sotomayor. Multi-item auctions. *Journal of political economy*, 94(4):863–872, 1986.
- [8] A. Fabre, T. Larroucau, M. Martinez, C. Neilson, and I. Rios. Application mistakes and information frictions in college admissions. Technical report, Working Paper, 2021.
- [9] S. Gleyze and A. Pernoud. Informationally simple incentives. Available at SSRN 3517922, 2020.
- [10] G. Haeringer and F. Klijn. Constrained school choice. *Journal of Economic theory*, 144(5):1921–1947, 2009.
- [11] N. L. Jacquet and S. Tan. On the segmentation of markets. *Journal of Political Economy*, 115(4):639–664, 2007.
- [12] A. J. Kapor, C. A. Neilson, and S. D. Zimmerman. Heterogeneous beliefs and school choice mechanisms. *American Economic Review*, 110(5):1274–1315, 2020.
- [13] A. Kloosterman and P. Troyan. School choice with asymmetric information: Priority design and the curse of acceptance. *Theoretical Economics*, 15(3):1095–1133, 2020.
- [14] J. M. McNamara and E. Collins. The job search problem as an employer-candidate game. *Journal of Applied Probability*, pages 815–827, 1990.
- [15] T. Mekonnen. Random versus directed search for scarce resources. Available at SSRN 3275771, 2019.
- [16] P. R. Milgrom and R. J. Weber. Distributional strategies for games with incomplete information. *Mathematics of operations research*, 10(4):619–632, 1985.
- [17] P. A. Pathak. What really matters in designing school choice mechanisms. *Advances in Economics and Econometrics*, 1:176–214, 2017.
- [18] A. E. Roth. Two-sided matching with incomplete information about others’ preferences. *Games and Economic Behavior*, 1(2):191–209, 1989.
- [19] A. E. Roth and M. A. O. Sotomayor. Two-sided matching, volume 18 of econometric society monographs, 1990.

Appendix A Proofs for results in the main

Proof of Lemma 1.

- Action spaces $\forall i \in \{1, \dots, n\}$, $A_i = \{A^1, \dots, A^m, N\}$ are finite. So by Proposition 1 in Milgrom and Weber (1985) [16], payoffs are equicontinuous (R1).
- Types ω_i , $i \in \{1, \dots, n\}$ are independent. So by Proposition 3 in Milgrom and Weber (1985) [16], information is absolutely continuous (R2).

Finally, by Theorem 1 in Milgrom and Weber (1985) [16] applied to the AG satisfying R1 and R2, there exists a BNE in the AG.

Proof of Lemma 2.

- By definition, $\forall \omega \in [0, 1]$, $\mathbb{E}[u_i(N_i)|\omega] = 0$. This proves (i) and (ii) for $X_i = N_i$.
- For $X_i = A_i^j$, $j \in \{1, \dots, m\}$, we have:

$$\mathbb{E}[u_i(A_i^j, p_{-i}([\omega, 1]))|\omega] = \underbrace{\prod_{i' \in \{1, \dots, n\} \setminus \{i\}} \left(\omega + \int_{\omega}^1 (1 - p_{i'}^j(x)) dx \right)}_{\mathbb{P}(S|A_i^j, p_{-i}([\omega, 1]), \omega)} a^k - c$$

$\forall i' \in \{1, \dots, n\} \setminus \{i\}$, $j \in \{1, \dots, m\}$, $\omega \mapsto \int_{\omega}^1 (1 - p_{i'}^j(x)) dx$ is continuous.

So the probability of success $\omega \mapsto \mathbb{P}(S|A_i^j, p_{-i}([\omega, 1]), \omega)$ and the whole expectation $\mathbb{E}[u_i(A_i^j, p_{-i}([\omega, 1]))|\omega]$ are also continuous. This proves (i).

Set $\omega_- < \omega_+$.

$$\begin{aligned} \omega_- + \int_{\omega_-}^1 (1 - p_{i'}^j(x)) dx &= (\omega_+ + \int_{\omega_+}^1 (1 - p_{i'}^j(x)) dx) + \omega_- - \omega_+ + \int_{\omega_-}^{\omega_+} (1 - p_{i'}^j(x)) dx \\ &= (\omega_- + \int_{\omega_-}^1 (1 - p_{i'}^j(x)) dx) + \int_{\omega_-}^{\omega_+} (-p_{i'}^j(x)) dx \\ &< \omega_+ + \int_{\omega_+}^1 (1 - p_{i'}^j(x)) dx \\ &\implies \mathbb{P}(S|A_i^j, p_{-i}([\omega, 1]), \omega_-) < \mathbb{P}(S|A_i^j, p_{-i}([\omega, 1]), \omega_+) \\ &\implies \mathbb{E}[u_i(A_i^j, p_{-i}([\omega, 1]))|\omega_-] < \mathbb{E}[u_i(A_i^j, p_{-i}([\omega, 1]))|\omega_+] \end{aligned}$$

This proves (ii).

In addition:

$$\lim_{\omega \rightarrow 1} \mathbb{P}(S|A_i^j, p_{-i}([\omega, 1]), \omega) = 1 \implies \lim_{\omega \rightarrow 1} \mathbb{E}[u_i(A_i^j, p_{-i}([\omega, 1]))|\omega] = a^j \quad (iii)$$

Proof of Proposition 0.

For agent ranked i^{th} in priority, $i \in \{1, \dots, n\}$, denote $\mu(i) \in \{1, \dots, m, \emptyset\}$ her final allocation on the benchmark market.

(i) We prove an induction statement: $H(i) : \mu_i = i$, $i \in \{1, \dots, m\}$

– Initial step

$$\begin{aligned} \forall i \in \{1, \dots, n\} : \omega_1 > \omega_i &\implies u_1(A_1^i, \sigma_{-1}) = a^i - c \\ u^1(N_1, \sigma_{-1}) &= 0 \\ \implies \forall \sigma_{-1}, BR^1(\sigma_{-1}) &= \{A_1^1\} \\ \implies \sigma_1 = A_1^1, \mu_1 = 1 &H(1) \end{aligned}$$

– Inductive step

$$\begin{aligned} \text{Set } i \in \{1, \dots, m-1\} \text{ s.t. } H(1), \dots, H(i) &\text{ true.} \\ \forall j \in \{1, \dots, i\} : \omega_j > \omega_{i+1} &\implies u_{i+1}(A_i^j, \sigma_{-(i+1)}) = -c \\ \forall j \in \{i+1, \dots, n\} : \omega_{i+1} > \omega_j &\implies u_{i+1}(A_i^j, \sigma_{-(i+1)}) = a^j - c \\ u_i(N_i, \sigma_{-(i+1)}) &= 0 \\ \implies BR_{i+1}(\sigma_{-(i+1)}) &= \{A_{i+1}^{i+1}\} \\ \implies \sigma_{i+1} = A_{i+1}^{i+1}, \mu_{i+1} = i+1 &H(i+1) \end{aligned}$$

(ii) We set $m + 1 \leq i \leq n$.

$$\begin{aligned} \text{By (i): } & \forall j \in \{1, \dots, m\}, u_i(A_i^j, \sigma^{-i}) = -c \\ & u_i(N_i, \sigma^{-i}) = 0 \\ \implies & BR_i(\sigma_{-i}) = \{N_i\} \\ \implies & \sigma_i = N_i \implies \mu_i = \emptyset \end{aligned}$$

Proof of Proposition 1.

We define the “interim action set” at score ω as the subset of actions that are played with positive probabilities at score ω : $A^j \in IAS(\omega)$ if $p^j(\omega) > 0$.

We state and prove a lemma characterizing interim action sets at any BNE of the AG (symmetric or asymmetric).

Lemma 4. [Interim action sets at BNE]

(i) At any BNE of the AG, $\sigma \in \square^{BNE}(\mathcal{G})$:

• Robust profile:

$$- \exists 1 = \omega^0 > \omega^1 > \omega^2 \geq 0 \text{ s.t. : } \begin{cases} IAS((\omega^1, \omega^0)) = \{A^1\} \\ IAS((\omega^2, \omega^1)) = \{A^1, A^2\} \end{cases} .$$

• Potential profile:

- $\forall k \in \{3, m\}$, if $\exists \omega | p^k(\omega) < 0$, then:

$$\exists 1 = \omega^0 > \omega^1 > \dots > \omega^k \geq 0 \text{ s.t.: } \forall j \in \{1, \dots, k\}, IAS((\omega^j, \omega^{j-1})) = \{A^1, \dots, A^j\}.$$

- If $\exists \omega | p^{m+1}(\omega) < 0$, then:

$$\begin{aligned} & \exists 1 = \omega^0 > \omega^1 > \dots > \omega^m > \omega^{m+1} = \\ \text{0 s.t. } & \begin{cases} \forall j \in \{1, \dots, m\}, IAS((\omega^j, \omega^{j-1})) = \{A^1, \dots, A^j\} \\ \{N\} \subseteq IAS((\omega^m, \omega^{m+1})) \end{cases} \end{aligned}$$

(ii) Moreover: ω^1 is the same in all multiple BNE of a given AG \mathcal{G} .

Proof of Lemma 4. Set $\sigma \in \square^{BNE}(\mathcal{G})$.

• By Lemma 2., (iii), we have:

$$\lim_{\omega \rightarrow 1} \mathbb{E}[u_i | A_i^1, \sigma_{-i}, \omega] = a^1 - c > \begin{cases} a^j - c = \lim_{\omega \rightarrow 1} \mathbb{E}[u_i | A_i^j, \sigma_{-i}, \omega] \\ 0 = \lim_{\omega \rightarrow 1} \mathbb{E}[u_i | N, \sigma_{-i}, \omega] \end{cases} .$$

And by Lemma 2., (i) (continuity): $\exists \omega^1$ s.t. $IAS((\omega^1, 1)) = \{A^1\}$.

• Suppose all players play A^1 at all scores: $IAS([0, 1]) = \{A^1\}$. Then:

$$\begin{aligned} \lim_{\omega \rightarrow 0} \mathbb{P}(S | A_i^1, \omega) = 0 & \implies \lim_{\omega \rightarrow 0} \mathbb{E}[u_i | A_i^1, \sigma_{-i}, \omega] = -c < 0 \\ j \in \{2, \dots, m\} : \lim_{\omega \rightarrow 0} \mathbb{P}(S | A_i^j, \omega) = 1 & \implies \lim_{\omega \rightarrow 0} \mathbb{E}[u_i | A_i^j, \sigma_{-i}, \omega] = a^j - c \\ \lim_{\omega \rightarrow 0} \mathbb{P}(S | N_i, \omega) = 0 & \implies \lim_{\omega \rightarrow 0} \mathbb{E}[u_i | N_i, \sigma_{-i}, \omega] = 0 \end{aligned}$$

Then, by Lemma 2. again, (i) and (ii):

- $\omega \mapsto \mathbb{E}[u_i | A_i^1, \sigma_{-i}, \omega]$ is continuous and strictly increasing on $[0, 1]$ from $-c$ to $a^1 - c$.
- $\omega \mapsto \mathbb{E}[u_i | A_i^j, \sigma_{-i}, \omega]$, $j \in \{2, \dots, m\}$ is constant on $[0, 1]$ and equal to $a^j - c$.
- $\omega \mapsto \mathbb{E}[u_i | N_i, \sigma_{-i}, \omega]$, $j \in \{2, \dots, m\}$ is constant on $[0, 1]$ and equal to 0.

By the bijection theorem, $\exists \omega^1 \in [0, 1]$ s.t.:

$$\mathbb{E}[u_i | A_i^1, \sigma_{-i}, \omega^1] = \mathbb{E}[u_i | A_i^2, \sigma_{-i}, \omega^1] = a^2 - c > \begin{cases} a^j - c = \mathbb{E}[u_i | A_i^j, \sigma_{-i}, \omega], j \in \{3, \dots, m\} \\ 0 = \mathbb{E}[u_i | N, \sigma_{-i}, \omega] \end{cases}$$

And by Lemma 2., (i) (continuity) again: $\exists 0 < \omega^2 < \omega^1$ s.t. $IAS((\omega^2, \omega^1)) \subseteq \{A^1, A^2\}$.

• Suppose no one plays A^k , $k \in \{1, \dots, 2\}$ on (ω^2, ω^1) . Set $k' \in \{1, 2\}$, $\neq k$.

Then, by Lemma 2., (ii), again:

- $\omega \mapsto \mathbb{E}[u_i | A_i^k, \sigma_{-i}, \omega]$ is constant on (s^2, s^1) equals to $a^2 - c$.
- $\omega \mapsto \mathbb{E}[u_i | A_i^{k'}, \sigma_{-i}, \omega]$ is strictly increasing on (s^2, s^1) .

So: $\mathbb{E}[u_i | A_i^k, \sigma_{-i}, \omega] > \mathbb{E}[u_i | A_i^{k'}, \sigma_{-i}, \omega]$, and playing A^k is a profitable deviation. So $A^k \in IAS((\omega^1, \omega^2))$.

Exchanging the the roles objects 1 and 2, we get: $\{A^1, A^2\} \subseteq IAS((\omega^1, \omega^2))$.

In the end: $IAS((\omega^1, \omega^2)) = \{A^1, A^2\}$.

- The proof for the intervals (ω^k, ω^{k-1}) , $j \in \{3, \dots, m\}$ below is similar. The inclusion $IAS(\omega^k, \omega^{k-1}) \subseteq \{A^1, \dots, A^k\}$ relies on the continuity and monotonicity of expected interim payoffs in Lemma 2. enabling an application of the intermediate value theorem. The reverse inclusion comes from the indifference in ω^{k-1} plus the monotonicity of expected interim payoffs in Lemma 2., giving a sharp characterization of the no profitable conditions.
- If ω^{m-1} exists and $\forall j \in \{1, \dots, m\}$ s.t. $\lim_{\omega \rightarrow 0} \mathbb{E}[u_i | A_i^j, \sigma_{-i}, \omega] < 0$, then by the intermediate value theorem again, $\exists 0 < \omega^m < \omega^{m-1}$ s.t. $\{N\} \subseteq SS((0, \omega^m)) \subseteq \{A^1, \dots, A^m, N\}$.

All the preceding proves (i) and (ii).

Notations: $s^k := \omega^k$, $k \in \{0, m+1\}$.

Lemma 4. shows that both actions A^1 and A^2 are played with positive probabilities on (s^2, s^1) . At pure equilibrium, this implies that different players play different actions and the equilibrium profile is asymmetric.

Proof of Theorem 1.

1. Interim action sets

Interim action sets at symmetric equilibrium are given by Lemma 4..

Notations: $t^k := \omega^k$, $k \in \{0, m+1\}$.

2. Constant probabilities

Let us now locate on the interval (t^k, t^{k-1}) and prove that the probability functions $\omega \mapsto p^j(\omega)$, $j \in \{1, m\}$ are constant on each interval (t^k, t^{k-1}) . Due to the complexity of notations, we write down the explicit proof of the probabilities being constant at the inductive step for $k = 2$ and give the way to go for the lower classes.

- Set $k = 2$. Let us locate on the (t^2, t^1) interval.

The strong indifference principle applied at a score $\omega^* \in (t^2, t^1)$ delivers the following differential equation:

$$\begin{aligned}
(E_{[1,2]}^1) : \mathbb{E}[u_i(A_i^1(\omega), \sigma_{-i}([\omega^*, 1])) | \omega^*] &= \mathbb{E}[u_i(A_i^2(\omega), \sigma_{-i}([\omega^*, 1])) | \omega^*] \\
\iff \left(1 - \int_{t^1}^1 f(\omega) d\omega - \int_{\omega^*}^{t^1} p^1(\omega) f(\omega) d\omega\right)^{n-1} a^1 - c &= \left(1 - \int_{\omega^*}^{t^1} p^2(\omega) f(\omega) d\omega\right)^{n-1} a^2 - c \\
\iff \left(1 - F(1) + F(t^1) - F(t^1) + F(\omega^*) + \int_{\omega^*}^{t^1} p^2(\omega) f(\omega) d\omega\right)^{n-1} a^1 & \\
= \left(1 - \int_{\omega^*}^{t^1} p^2(\omega) f(\omega) d\omega\right)^{n-1} a^2 & \\
\iff F(\omega^*) + \int_{\omega^*}^{t^1} p^2(\omega) f(\omega) d\omega = \left(\frac{a^2}{a^1}\right)^{\frac{1}{n-1}} \left(1 - \int_{\omega^*}^{t^1} p^2(\omega) f(\omega) d\omega\right) & \\
\iff (1 + F(t^1)) \int_{\omega^*}^{t^1} p^2(\omega) f(\omega) d\omega = F(t^1) - F(\omega^*) &
\end{aligned}$$

Set: G^2 , a primitive of $p^2 f$. Then:

$$(E_{[1,2]}^1) \iff G^2(t^1) - G^2(\omega^*) = \frac{F(t^1) - F(\omega^*)}{1 + F(t^1)}$$

Deriving on both sides, we get a necessary condition on the probability functions:

$$\begin{aligned}
(E_{[1,2]}^1) \implies p^2(\omega^*) f(\omega^*) &= \frac{-f(\omega^*)}{1 + F(t^1)} \\
\implies \begin{cases} p^2(\omega^*) = \frac{1}{1 + F(t^1)} \\ p^1(\omega) = \frac{F(t^1)}{1 + F(t^1)} \end{cases} &
\end{aligned}$$

We now need to check that those constant probability functions indeed verify equation $(E_{[1,2]}^1)$:

$$\begin{aligned}
(1 + F(t^1)) \int_{\omega^*}^{t^1} \frac{1}{1 + F(t^1)} f(\omega) d\omega &= F(t^1) - F(\omega^*) \\
\iff \frac{(1 + F(t^1))}{(1 + F(t^1))} (F(t^1) - F(\omega^*)) &= F(t^1) - F(\omega^*) \checkmark
\end{aligned}$$

- Set $k \in \{2, \dots, m-1\}$, and suppose $H(1), \dots, H(k)$ hold. Let us locate on the interval (t^{k+1}, t^k) . The strong indifference principle applied at a score $\omega^* \in (t^{k+1}, t^k)$ delivers a system of k differential equations with $k+1$ unknowns. We denote those equations $(E_{[[1, k+1]]}^j)$, $j \in \{1, \dots, k\}$. Each of them is given by:

$$\begin{aligned} (E_{[[1, k+1]]}^j) : \mathbb{E}[u_i(A_i^j(\omega), \sigma_{-i}([\omega^*, 1])) | \omega^*] &= \mathbb{E}[u_i(A_i^{k+1}(\omega), \sigma_{-i}([\omega^*, 1])) | \omega^*] \\ \iff \left(1 - \int_{\omega^*}^{t^k} p^j(\omega) f(\omega) d\omega - \sum_{l=j}^{t^k} \int_{t^l}^{t^{l-1}} p^j(\omega) f(\omega) d\omega\right)^{n-1} a^j - c \\ &= \left(1 - \int_{\omega^*}^{t^k} p^{k+1}(\omega) f(\omega) d\omega\right)^{n-1} a^{k+1} - c \end{aligned}$$

If we replace $p^1(\omega^*)$ by $1 - \sum_{j=2}^{k+1} p^j(\omega^*)$, we end up with only k unknowns, hence a Cramer system. We can use the substitution method, to get in the end a relation between (for instance) only $\int_{\omega^*}^{t^k} p^{k+1}(\omega) f(\omega) d\omega$ and $p_{[[1, k]]}^k \int_{t^k}^{t^{k-1}} f(\omega) d\omega$. Posing primitives and deriving the whole gives constant probabilities $p_{[[1, k+1]]}^j$.

3. Probability levels

Set $k \in \{2, \dots, m-1\}$, and further exploit the differential equations:

$$\begin{aligned} (E_{[[1, k+1]]}^j) : \mathbb{E}[u_i(A_i^j(\omega), \sigma_{-i}([\omega^*, 1])) | \omega^*] &= \mathbb{E}[u_i(A_i^{k+1}(\omega), \sigma_{-i}([\omega^*, 1])) | \omega^*] \\ \iff \left(1 - \int_{\omega^*}^{t^k} p^j(\omega) f(\omega) d\omega - \sum_{l=j}^{t^k} \int_{t^l}^{t^{l-1}} p^j(\omega) f(\omega) d\omega\right)^{n-1} a^j - c \\ &= \left(1 - \int_{\omega^*}^{t^k} p^{k+1}(\omega) f(\omega) d\omega\right)^{n-1} a^{k+1} - c \\ \iff \left(1 - \int_{\omega^*}^{t^k} p^j(\omega) f(\omega) d\omega - \sum_{l=j}^{t^k} \int_{t^l}^{t^{l-1}} p^j(\omega) f(\omega) d\omega\right) \\ &= \left(\frac{a^{k+1}}{a^j}\right)^{\frac{1}{n-1}} \left(1 - \int_{\omega^*}^{t^k} p^{k+1}(\omega) f(\omega) d\omega\right) \\ \iff \left(1 - p_{[[1, k+1]]}^j (F(\omega^*) - F(t^k)) - \sum_{l=j}^{t^k} p_{[[1, l]]}^j (F(t^l) - F(t^{l-1}))\right) \\ &= \left(\frac{a^{k+1}}{a^j}\right)^{\frac{1}{n-1}} \left(1 - p_{[[1, k+1]]}^{k+1} (F(\omega^*) - F(t^k))\right) \end{aligned}$$

Deriving on both sides, we get:

$$\begin{aligned} (E_{[[1, k+1]]}^j) &\implies -p_{[[1, k+1]]}^j f(\omega^*) = -\left(\frac{a^{k+1}}{a^j}\right)^{\frac{1}{n-1}} p_{[[1, k+1]]}^{k+1} f(\omega^*) \\ &\implies p_{[[1, k+1]]}^j = \left(\frac{a^{k+1}}{a^j}\right)^{\frac{1}{n-1}} p_{[[1, k+1]]}^{k+1} \end{aligned}$$

This is a recursive formula for the probability levels within class (t^{k+1}, t^k) . To find the explicit formulas, we use the fact that the $k+1$ probability levels sum up to one:

$$\begin{aligned} \sum_{j=1}^{k+1} p_{[[1, k+1]]}^j &= 1 \implies p_{[[1, k+1]]}^{k+1} \sum_{l=1}^{k+1} \left(\frac{a^{k+1}}{a^l}\right)^{\frac{1}{n-1}} = 1 \\ &\implies p_{[[1, k+1]]}^{k+1} = \left((a^{k+1})^{\frac{1}{n-1}} \sum_{l=1}^{k+1} (a^l)^{\frac{-1}{n-1}}\right)^{-1} \\ &\implies p_{[[1, k+1]]}^j = \left((a^j)^{\frac{1}{n-1}} \sum_{l=1}^{k+1} (a^l)^{\frac{-1}{n-1}}\right)^{-1} \\ &\implies p_{[[1, k+1]]}^j = \left(\sum_{l=1}^{k+1} \left(\frac{a^j}{a^l}\right)^{\frac{1}{n-1}}\right)^{-1} \end{aligned}$$

Finally, on the bottom interval $(0, t^m)$, it cannot be that players apply to $j \in \{1, \dots, m\}$. Otherwise, by Lemma 2. (ii), $\omega \mapsto \mathbb{E}[u|A^j, \sigma, \omega]$ would be increasing so strictly negative at some score $\in (0, t^m)$, hence a profitable deviation to action N . So: p^{m+1} is constant equal to 1 on $(0, t^m)$.

4. Thresholds

The remaining task is to characterize the thresholds t^k , $k \in \{2, \dots, m\}$.

- t^k , $k \in \{2, \dots, m-1\}$ By definition:

$$t^k := \inf \{ \omega^* \in [0, 1] \mid \forall \omega > \omega^*, \min_{l \in \{1, \dots, k\}} \mathbb{E}[u_i(A_i^l(\omega), \sigma_{-i}([\omega, 1])) \mid \omega] \geq \mathbb{E}[u_i(A_i^{k+1}(\omega), \sigma_{-i}([\omega, 1])) \mid \omega] \}$$

By the strong indifference principle, we have that all $\mathbb{E}[u_i(A_i^l(\omega), \sigma_{-i}([\omega, 1])) \mid \omega]$, $l \in \{1, \dots, k\}$ are equal on (t^k, t^{k-1}) . So:

$$\begin{aligned} t^k &= \inf \{ \omega^* \in [0, 1] \mid \forall \omega > \omega^*, \mathbb{E}[u_i(A_i^k(\omega), \sigma_{-i}([\omega, 1])) \mid \omega] \geq \mathbb{E}[u_i(A_i^{k+1}(\omega), \sigma_{-i}([\omega, 1])) \mid \omega] \} \\ &\implies \mathbb{E}[u_i(A_i^k(t^k), \sigma_{-i}([t^k, 1])) \mid t^k] = \mathbb{E}[u_i(A_i^{k+1}(t^{k+1}), \sigma_{-i}([t^{k+1}, 1])) \mid t^{k+1}] \\ &\implies \left(1 - \int_{t^k}^{t^{k-1}} p^k(\omega) f(\omega) d\omega\right)^{n-1} a^k - c = a^{k+1} - c \\ &\implies (1 - p_{[[1, k]]}^k (F(t^{k-1}) - F(t^k)))^{n-1} a^k - c = a^{k+1} - c \\ &\implies p_{[[1, k]]}^k (F(t^{k-1}) - F(t^k)) = 1 - \left(\frac{a^{k+1}}{a^k}\right)^{\frac{1}{n-1}} \\ &\implies F(t^k) = F(t^{k-1}) + \frac{1}{p_{[[1, k]]}^k} \left(-1 + \left(\frac{a^{k+1}}{a^k}\right)^{\frac{1}{n-1}}\right) \end{aligned}$$

Plugging in the formula for the probability levels, we get:

$$\begin{aligned} F(t^k) &= F(t^{k-1}) + (a^k)^{\frac{1}{n-1}} \sum_{l=1}^k (a^l)^{\frac{-1}{n-1}} \left(-1 + \left(\frac{a^{k+1}}{a^k}\right)^{\frac{1}{n-1}}\right) \\ F(t^k) &= F(t^{k-1}) + ((a^{k+1})^{\frac{1}{n-1}} - (a^k)^{\frac{1}{n-1}}) \sum_{l=1}^k (a^l)^{\frac{-1}{n-1}} \end{aligned}$$

This is a recursive formula characterising the thresholds. The explicit formula is therefore:

$$F(t^k) = F(t^0) + \sum_{j=1}^k ((a^{j+1})^{\frac{1}{n-1}} - (a^j)^{\frac{1}{n-1}}) \sum_{l=1}^k (a^l)^{\frac{-1}{n-1}}$$

The second term is close to be a telescopic sum. The second part is $\sum_{j=1}^k \alpha_j (a^j)^{\frac{-1}{n-1}}$ where we compute the terms below:

$$\begin{aligned} - j = 1: & \alpha_1 (a^1)^{\frac{-1}{n-1}} = -(a^1)^{\frac{1}{n-1}} (a^1)^{\frac{-1}{n-1}} = -1 \\ - 2 \leq j \leq k: & \\ & \alpha_j (a^j)^{\frac{-1}{n-1}} = \left(\sum_{l=1}^{j-1} (a^l)^{\frac{-1}{n-1}} - \sum_{l=1}^j (a^l)^{\frac{-1}{n-1}} \right) (a^j)^{\frac{-1}{n-1}} = -(a^j)^{\frac{1}{n-1}} (a^j)^{\frac{-1}{n-1}} = -1 \\ - j = k+1: & \alpha_{k+1} (a^{k+1})^{\frac{-1}{n-1}} = \sum_{l=1}^k (a^l)^{\frac{-1}{n-1}} (a^{k+1})^{\frac{-1}{n-1}} = \sum_{l=1}^k \left(\frac{a^{k+1}}{a^l}\right)^{\frac{-1}{n-1}} \end{aligned}$$

In the end, we get:

$$F(t^k) = 1 - k + \sum_{l=1}^k \left(\frac{a^{k+1}}{a^l}\right)^{\frac{-1}{n-1}}$$

- $k = m$

$$\begin{aligned}
\mathbb{E}[u_i(A_i^m(t^m), \sigma_{-i}([t^m, 1]) | t^m)] &= \mathbb{E}[u_i(N(t^m), \sigma_{-i}([t^m, 1]) | t^m)] \\
&\iff (1 - p_{[1, m]}^m (F(t^m) - F(t^{m-1})))^{n-1} a^m - c = 0 \\
&\iff F(t^m) = F(t^{m-1}) + \frac{1}{p_{[1, m]}^m} \left(-1 + \left(\frac{c}{a^m} \right)^{\frac{1}{n-1}} \right) \\
&\iff F(t^m) = 2 - m + \sum_{l=1}^{m-1} \left(\frac{a^m}{a^l} \right)^{\frac{1}{n-1}} + \sum_{l=1}^m \left(\frac{a^m}{a^l} \right)^{\frac{1}{n-1}} \left(-1 + \left(\frac{c}{a^m} \right)^{\frac{1}{n-1}} \right) \\
&\iff F(t^m) = 1 - m + \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}}
\end{aligned}$$

5. Number of classes

$$\begin{aligned}
k_0(p^*) = k \in \{1, \dots, m\} &\iff F(t^k) \leq 0 < F(t^{k-1}) \\
&\iff 1 - k + \sum_{l=1}^k \left(\frac{a^{k+1}}{a^l} \right)^{\frac{1}{n-1}} \leq 0 < 2 - k + \sum_{l=1}^{k-1} \left(\frac{a^k}{a^l} \right)^{\frac{1}{n-1}} \\
&\iff 1 + \sum_{l=1}^k \left(\frac{a^{k+1}}{a^l} \right)^{\frac{1}{n-1}} \leq k < 2 + \sum_{l=1}^{k-1} \left(\frac{a^k}{a^l} \right)^{\frac{1}{n-1}} \\
k_0(p^*) = m + 1 &\iff F(t^m) > 0 \\
&\iff 2 - m + \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} > 0 \\
&\iff 2 + \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} > m
\end{aligned}$$

Proof of Corollary 1.

(i) We characterize the probability level variations:

- Within classe

By the block structure, we immediately have that for $\omega, \omega' \in [t^k, t^{k-1}]$: $p^j(\omega) = p_{[1, k]}^j = p^j(\omega')$.

- Between classes

From Theorem 1., probability levels write as:

$$p_{[1, k]}^j = \left((a^j)^{\frac{1}{n-1}} \sum_{l=1}^k (a^l)^{\frac{-1}{n-1}} \right)^{-1}$$

$k \mapsto \sum_{l=1}^k (a^l)^{\frac{-1}{n-1}}$ is a sum with positive terms, hence increasing in k . Going to the inverse, we find that $p_{[1, k]}^j$ is decreasing in k .

(ii) We prove First Order Stochastic Dominance:

Denote: $\sigma(\omega)$ the distribution with support $\{A^1, \dots, A^m\}$ and probabilities: $\mathbb{P}(\sigma(\omega) = A^j) = p^j(\omega)$, $j \in \{1, \dots, m\}$.

Using the formulas for probability levels in Theorem 1., we find for $\omega \in (t^k, t^{k-1})$:

$$\begin{aligned}
\sum_{l=1}^j p^l(\omega) &= \sum_{l=1}^j p_{[1, k]}^l = \begin{cases} \sum_{l=1}^j \left((a^j)^{\frac{1}{n-1}} \sum_{l=1}^k (a^l)^{\frac{-1}{n-1}} \right)^{-1} & \text{if } j < k \\ 1 & \text{if } j \geq k \end{cases} \\
&= \begin{cases} \sum_{l=1}^j (a^l)^{\frac{-1}{n-1}} & \text{if } j < k \\ 1 & \text{if } j \geq k \end{cases}
\end{aligned}$$

Set $0 \leq \omega' < \omega \leq 1$.

We seek to demonstrate $\forall j \in \{1, \dots, m\}$:

$$(\star)_j : \sum_{l=1}^j p^l(\omega') \leq \sum_{l=1}^j p^l(\omega)$$

- If ω', ω belong to the same class C^k , then because of the block structure:

$$\sum_{l=1}^j p^l(\omega') = \sum_{l=1}^j p^l(\omega)$$

So $(\star)_j$, $j \in \{1, \dots, m\}$ trivially holds.

- If ω', ω belong to different classes, $\omega' \in C^{k'}$, $\omega \in C^k$, $k < k'$, then there are two subcases:
 - * If $k' = m + 1$ then $\forall j \in \{1, \dots, m\}$, $\sum_{l=1}^j p^l(\omega') = 0$ and $(\star)_j$ is trivially verified.
 - * If $k' \leq m$ then:
 - For $j \leq k$, then: $\sum_{l=1}^j p^l(\omega) = 1$ and $(\star)_j$ is trivially verified.
 - For $j < k < k'$, then:

$$\frac{\sum_{l=1}^j p^l(\omega)}{\sum_{l=1}^j p^l(\omega')} = \frac{\sum_{l=1}^{k'} (a^l)^{\frac{-1}{n-1}}}{\sum_{l=1}^k (a^l)^{\frac{-1}{n-1}}} > 1 \implies (\star)_j$$

(iii) From the proof of Theorem 1., we have the following recursive formula:

$$p_{[[1,k]]}^j = \left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} p_{[[1,k]]}^k$$

And $a^j \mapsto \left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}}$ is decreasing in j , equivalently increasing in j . So, $p_{[[1,k]]}^j$ is increasing in j .

(iv) The ex ante probability of applying to object j writes:

$$p^j := \sum_{k=j}^{m+1} (F(t^{k-1}) - F(t^k)) p_{[[1,k]]}^j$$

For $1 \leq k \leq m - 1$, from Theorem 1. we have:

$$\begin{aligned} (F(t^{k-1}) - F(t^k)) p_{[[1,k]]}^j &= \frac{1}{p_{[[1,k]]}^k} \left(1 - \left(\frac{a^{k+1}}{a^k}\right)^{\frac{1}{n-1}}\right) p_{[[1,k]]}^j \\ &= \sum_{l=1}^k \left(\frac{a^k}{a^l}\right)^{\frac{1}{n-1}} \frac{1}{\sum_{l=1}^k \left(\frac{a^j}{a^l}\right)^{\frac{1}{n-1}}} \left(1 - \left(\frac{a^{k+1}}{a^k}\right)^{\frac{1}{n-1}}\right) \\ &= \left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} \left(1 - \left(\frac{a^{k+1}}{a^k}\right)^{\frac{1}{n-1}}\right) \\ &= \frac{(a^k)^{\frac{1}{n-1}} - (a^{k+1})^{\frac{1}{n-1}}}{(a^j)^{\frac{1}{n-1}}} (\star) \end{aligned}$$

Summing up, we recognize a telescopic sum and we get:

$$\begin{aligned} p^j &= \frac{(a^j)^{\frac{1}{n-1}} - (a^{m+1})^{\frac{1}{n-1}}}{(a^j)^{\frac{1}{n-1}}} \\ p^j &= 1 - \left(\frac{c}{a^j}\right)^{\frac{1}{n-1}} \end{aligned}$$

Proof of Corollary 2.

(i) All agents participate iff there are at most m classes. By Theorem 1. (i), this happens iff:

$$m \geq 1 + \sum_{l=1}^m \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}}$$

(ii) By Proposition 0., at the Nash equilibrium of the Application Game with perfect information, the expected participation is $\frac{m}{n}$.

By Theorem 1., at the symmetric BNE of the Application Game with imperfect information, expected participation is:

$$1 - F(t^m) = m - \sum_{l=1}^m \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}} = \sum_{l=1}^m 1 - \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}} = \sum_{l=1}^m p^l$$

Participation is therefore higher at BNE iff:

$$\begin{aligned} m - \sum_{l=1}^m \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}} &\geq \frac{m}{n} \\ 1 - \frac{1}{n} &\geq \frac{1}{m} \sum_{l=1}^m \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}} \\ m &\geq \frac{n}{n-1} \sum_{l=1}^m \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}} \end{aligned}$$

Proof of Proposition 2.

(B) By Proposition 0., the welfare on the benchmark market is given by:

$$W^B := \mathbb{E}[u(\sigma^*)] = \frac{1}{n} \sum_{k=1}^m (a^k - c)$$

Since $\forall k \in \{1, \dots, m\}$, $a^k - c > 0$, this is the maximum welfare attainable in the Application Game.

(F) We use the formula of the interim welfare, $t^k < w < t^{k-1}$ (see section §5.2):

$$W^F(\omega) = \sum_{j=1}^k p_{[[1,k]]}^j \cdot \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j \right)^{n-1} a^j - c$$

The ex ante welfare aggregates all interim welfare ($k = m$) taking into account the distribution of scores:

$$\begin{aligned} W^F &= \int_0^1 W^F(\omega) f(\omega) d\omega \\ &= \sum_{k=1}^m \left[\int_{t^k}^{t^{k-1}} \sum_{j=1}^k p_{[[1,k]]}^j \cdot \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j \right)^{n-1} a^j f(\omega) d\omega - c(F(t^{k-1}) - F(t^k)) \right] \end{aligned}$$

We denote:

$$\begin{aligned} I_k &:= \int_{t^k}^{t^{k-1}} \sum_{j=1}^k p_{[[1,k]]}^j \cdot \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j \right)^{n-1} a^j f(\omega) d\omega \\ &= \sum_{j=1}^k p_{[[1,k]]}^j a^j \cdot \int_{t^k}^{t^{k-1}} \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j \right)^{n-1} f(\omega) d\omega \end{aligned}$$

We denote:

$$\begin{aligned} L_{kj} &:= \int_{t^k}^{t^{k-1}} \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j \right)^{n-1} f(\omega) d\omega \\ &= \left[\frac{1}{n p_{[[1,k]]}^j} \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j \right)^n \right]_{t^{k-1}}^{t^k} \\ &= \frac{1}{n p_{[[1,k]]}^j} \left[\left(\frac{a^k}{a^j}\right)^{\frac{n}{n-1}} - \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(t^k)) p_{[[1,k]]}^j \right)^n \right] \\ &\stackrel{(*)}{=} \frac{1}{n p_{[[1,k]]}^j} \left[\left(\frac{a^k}{a^j}\right)^{\frac{n}{n-1}} - \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - \frac{(a^k)^{\frac{1}{n-1}} - (a^{k+1})^{\frac{1}{n-1}}}{(a^j)^{\frac{1}{n-1}}} \right)^n \right] \\ &= \frac{1}{n p_{[[1,k]]}^j} \left[\left(\frac{a^k}{a^j}\right)^{\frac{n}{n-1}} - \left(\left(\frac{a^{k+1}}{a^j}\right)^{\frac{1}{n-1}} \right)^n \right] \\ &= \frac{1}{n p_{[[1,k]]}^j} \left[\frac{(a^k)^{\frac{n}{n-1}} - (a^{k+1})^{\frac{n}{n-1}}}{(a^j)^{\frac{n}{n-1}}} \right] \end{aligned}$$

Substituting in I_k , we get:

$$\begin{aligned} I_k &= \sum_{j=1}^k p_{[[1,k]]}^j a^j \cdot \frac{1}{np_{[[1,k]]}^j} \left[\frac{(a^k)^{\frac{n}{n-1}} - (a^{k+1})^{\frac{n}{n-1}}}{(a^j)^{\frac{n}{n-1}}} \right] \\ &= \frac{1}{n} \left((a^k)^{\frac{n}{n-1}} - (a^{k+1})^{\frac{n}{n-1}} \right) \sum_{j=1}^k (a^j)^{\frac{-1}{n-1}} \end{aligned}$$

Substituting in W^F , we get:

$$W^F = \frac{1}{n} \left(\sum_{k=1}^m \left((a^k)^{\frac{n}{n-1}} - (a^{k+1})^{\frac{n}{n-1}} \right) \sum_{j=1}^k (a^j)^{\frac{-1}{n-1}} \right) - c \sum_{k=1}^m (F(t^{k-1}) - F(t^k))$$

The second term is a telescopic sum and the first term is close to be a telescopic sum.

- $k = 1$: $(a^1)^{\frac{n}{n-1}} \cdot (a^1)^{\frac{-1}{n-1}} = a^1$
- $2 \leq k \leq m$: $(a^k)^{\frac{n}{n-1}} \left(\sum_{j=1}^k (a^j)^{\frac{-1}{n-1}} - \sum_{j=1}^{k-1} (a^j)^{\frac{-1}{n-1}} \right) = (a^k)^{\frac{n}{n-1}} (a^k)^{\frac{-1}{n-1}} = a^k$
- $k = m + 1$: $-(a^{m+1})^{\frac{n}{n-1}} \sum_{j=1}^m (a^j)^{\frac{-1}{n-1}}$

In total, we get:

$$\begin{aligned} W^F &= \frac{1}{n} \left(\sum_{k=1}^m a^k - (a^{m+1})^{\frac{n}{n-1}} \sum_{j=1}^m (a^j)^{\frac{-1}{n-1}} \right) - c(1 - F(t^m)) \\ &= \frac{1}{n} \left(\sum_{k=1}^m a^k - c^{\frac{n}{n-1}} \sum_{j=1}^m (a^j)^{\frac{-1}{n-1}} \right) - c \left(m - \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} \right) \\ &= \frac{1}{n} \left(\sum_{k=1}^m a^k \right) - \frac{1}{n} c \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} - c \left(m - \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} \right) \\ &= \frac{1}{n} \left(\sum_{k=1}^m a^k \right) - c \left(m - \frac{n}{n-1} \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} \right) \end{aligned}$$

The welfare gap is given by:

$$\begin{aligned} W^B - W^F &= \frac{1}{n} \sum_{j=1}^m (a^j - c) - \left[\frac{1}{n} \sum_{k=1}^m a^k + c \left(m - \frac{n-1}{n} \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} \right) \right] \\ &= c \left(-\frac{m}{n} + m - \frac{n-1}{n} \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} \right) \\ &= c \frac{n-1}{n} \left(m - \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} \right) \end{aligned}$$

$$\forall l \in \{1, \dots, m\}, c < a^l \Rightarrow \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} < 1 \Rightarrow m > \sum_{l=1}^m \left(\frac{c}{a^l} \right)^{\frac{1}{n-1}} \Rightarrow W^B > W^F$$

Proof of Lemma 3.

- (B) By Proposition 0., on at the Nash equilibrium of the benchmark market, and agent receives utility $(a^i - c)$ iff he is ranked i^{th} in priority. For an agent with score ω , this happens with probability:

$$\mathbb{P}(\omega \text{ ranked } i) = (1 - F(\omega))^{i-1} F(\omega)^{n-i} \binom{n-1}{i-1}$$

In total, we have:

$$W^B(\omega) = \sum_{i=1}^m \binom{n-1}{i-1} (1 - F(\omega))^{i-1} F(\omega)^{n-i} (a^i - c)$$

F cdf hence continuous $\Rightarrow W^B$ continuous.

By definition (or by computation, using the binomial theorem):

$$\sum_{i=1}^n \binom{n-1}{i-1} (1-F(\omega))^{i-1} F(\omega)^{n-i} = 1$$

When ω increases, $\mathbb{P}(\omega \text{ ranked } i)$ increases (decreases) for small (large) i - associated to high (low) utilities ($a^i - c$). So $W^B(\omega)$ strictly increasing with ω .

(F) By Theorem 1., for $t^k < \omega < t^k - 1$, we have:

$$W^F(\omega) := \mathbb{E}[u(p^*)|\omega] = \sum_{j=1}^k p_{[[1,k]]}^j \mathbb{P}(S|A^j, \omega) a^j - c$$

$$\mathbb{P}(S|A^j, \omega) = \left(1 - \sum_{l=j}^{k-1} (F(t^{l-1}) - F(t^l)) p_{[[1,l]]}^j - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j\right)^{n-1}$$

From the proof of Corollary 1., we know:

$$(F(t^{l-1}) - F(t^l)) p_{[[1,l]]}^j = \frac{(a^l)^{\frac{1}{n-1}} - (a^{l+1})^{\frac{1}{n-1}}}{(a^j)^{\frac{1}{n-1}}} \quad (\star)$$

Summing up, we recognize a telescopic sum and we get:

$$\sum_{l=j}^{k-1} (F(t^{l-1}) - F(t^l)) p_{[[1,l]]}^j = \frac{(a^j)^{\frac{1}{n-1}} - (a^k)^{\frac{1}{n-1}}}{(a^j)^{\frac{1}{n-1}}} = 1 - \left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}}$$

Substituting, we get:

$$\mathbb{P}(S|A^j, \omega) = \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j\right)^{n-1}$$

$$W^F(\omega) := \mathbb{E}[u(p^*)|\omega] = \sum_{j=1}^k p_{[[1,k]]}^j a^j \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j\right)^{n-1} - c$$

By Theorem 1. again, for $0 \leq \omega \leq t^m$, we trivially get $W^F(\omega) = 0$ (constant).

F cdf hence continuous $\Rightarrow W^F$ continuous.

For the monotonicity, we differentiate:

$$\frac{\partial \mathbb{P}(S|A^j, \omega)}{\partial \omega} = f'(\omega) (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j (n-1) \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[[1,k]]}^j\right)^{n-2} > 0$$

This immediately proves the monotonicity within class.

When combined with Corollary 1. (ii), it also proves the monotonicity across classes.

Proof of Proposition 3.

We define the interim welfare gap: $g(\omega) = W^B(\omega) - W^F(\omega)$.

By Theorem 1., we have:

$$m > 1 + \sum_{l=1}^k \left(\frac{c}{a^l}\right)^{\frac{1}{n-1}} \iff W^F(0) > 0 \iff g(0) > 0$$

By Proposition 2. (continuity), we get existence of the threshold ω' .

***** Proofs of example (§4.2) *****

Proof of example $n = 3 > m = 2$, $F \sim \mathcal{U}$ for (asymmetric) pure BNE

- Top class $(s^1, 1)$:
By Lemma 4., $\forall i \in \{1, \dots, 3\}$, $\sigma_i((s^1, 1)) = A^1$.

- Middle class (s^2, s^1) :

By Lemma 4.:

$\exists i \in \{1, 2, 3\}, s_i((s^2, s^1)) = A^1$ (set $i=1$).

$\exists i' \in \{1, 2, 3\}, s_{i'}((s^2, s^1)) = A^2$ (set $i'=2$).

Let us characterize the action of player 3 on $[s^2, s^1]$. The interim payoffs of player 3 at $\omega' < s^1$ write:

$$\mathbb{E}[u_3(A_3^1(\omega'), p([\omega', s^1])|\omega')] = \omega' s^1 a^1 - c$$

$$\mathbb{E}[u_3(A_3^2(\omega'), p([\omega', s^1])|\omega')] = (1 - (s^1 - \omega'))a^2 - c$$

The difference between the two is:

$$\Delta(\omega') = [\omega' s^1 a^1 - c] - [(1 - (s^1 - \omega'))a^2 - c] = \omega' s^1 a^1 - (1 - s^1 + \omega')a^2$$

We differentiate with respect to ω' :

$$\frac{\partial \Delta(\omega')}{\partial \omega'} = s^1 a^1 - a^2 = \sqrt{a^1 a^2} - a^2 = \sqrt{a^2}(\sqrt{a^1} - \sqrt{a^2}) > 0$$

So Δ is strictly increasing. We know, by indifference at s^1 : $\Delta(s^1) = 0$. So: $\Delta(\omega') < 0$.

Conclusion: $s_3([s^2, s^1]) = A^2$.

- Bottom interval $(0, s^2)$:

By Lemma 4.: $\exists i \in \{1, 2, 3\}$ s.t. $s^i([0, s^2]) = N$.

By Lemma 2. (ii): $\omega \mapsto \mathbb{E}[u_1(A^1(\omega), s([\omega,])|\omega)]$ is constant..

So $s_1([0, s^2]) = A^1$ and $i \neq 1$.

Fix $i = 2$.

By Lemma 2. (i) and (ii): $\begin{cases} \omega \mapsto \mathbb{E}[u_2(N(\omega), s([\omega,])|\omega)] \text{ is constant.} \\ \omega \mapsto \mathbb{E}[u_3(A^2(\omega), s([\omega, 1])|\omega)] = \mathbb{E}[u_3(N(\omega), s([\omega, 1])|\omega)] \text{ is constant.} \\ \omega \mapsto \mathbb{E}[u_1(A^1(\omega), s([\omega, 1])|\omega)] \text{ is constant.} \end{cases}$

So: $s_2([0, s^2]) = N$ or A^2 .

Proof of example $n = 3 > m = 2, F \sim \mathcal{U}$ for symmetric (interior) BNE

Included in Theorem 1..

Appendix B Supplements

B.1 Pure (asymmetric) Bayes-Nash equilibrium - Partial characterization

The next theorem only partially characterizes the pure BNE of the AG:

Theorem 2. [Pure (asymmetric) BNE]

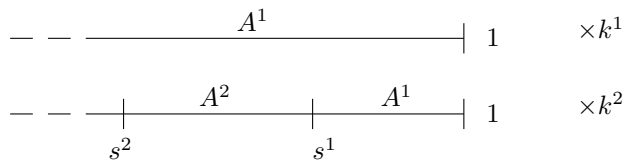
A pure strategy BNE of the AG:

- (i) exists and is unique up to strategies on the $(0, s^m)$ interval, and payoff-unique.
- (ii) exhibits finite number of intervals of scores where the interim action sets are constant.

The proof is in two steps. First, we use Lemma 4 characterizing the interim action sets. Second, for each interval of score with constant interim action set, we characterize the number of agents playing each action in the interim action set. For example, on the interval $[s^2, s^1]$, we determine the pair(s) of two integers (k^1, k^2) , $k^1 + k^2 = n$, where k^1 (k^2) of agents playing A^1 (A^2). We find that the no profitable deviation inequalities between payoffs always defines a (unique) pair (k^1, k^2) . We proceed similarly at lower scores. The theorem still allows many different patterns within the intervals where the interim action set is constant. Whenever we introduce $n \geq 4$ agents, the equilibrium patterns depends finely on the parameters of the AG, hence a low robustness.²⁴ We illustrate this lack of robustness below with an example:

$$n = 4, F \sim \mathcal{U}$$

- Case 1: $a^1 > 8a^2 + 7c \rightarrow (k^1, k^2) = (1, 3)$
- Case 2: $a^1 < 8a^2 + 7c \rightarrow (k^1, k^2) = (2, 2)$



In general, the pure BNE can support quite odd strategy profiles, where some strategies exhibit no sorting (an agent plays higher value objects at lower scores), or sorting with jumps (an agent plays high value objects at high scores, low value objects at intermediate scores but never plays the intermediary value objects). In these profiles, each strategy is virtually unique and highly sophisticated. The profiles are “very asymmetric”. This questions the ability of players to coordinate on these equilibria.

²⁴In this respect, the robustness of the pure equilibrium in section 8.4.2 was a special feature of the example.

B.2 Symmetric (interior) equilibrium - Comparative statics

In this section, we describe how a change in the parameters of the AG affects the symmetric BNE of the AG.

Values and cost

The coming proposition emphasizes that the symmetric equilibrium is invariant to a rescaling of all object values and the application cost.

Proposition 4. *[Invariance to rescaling]*

In an Application Game, if we multiply all object values and the cost by a given constant, the symmetric (interior) Bayes-Nash equilibrium remains unchanged.

Proposition 4 implies that any non-trivial comparative static analysis must first keep the cost fixed as object values fluctuate, and second, normalize the object values while varying the cost.

The relative position of values affects the equilibrium structure in an intuitive way. If objects are highly homogeneous in values, agents almost perfectly randomize between available objects. There is close to a single class, the lowest one, plus the no application class. Coordination is horizontal. When, on the contrary, the values are heterogeneous, people sort by levels of scores, the strategy is close to being pure. Coordination is vertical.

When the cost is high, agents even with intermediate scores resort to the safe no application action, and agents with the highest score coordinate. When the cost is low, incentives to play the no application action disappear, and incentives to coordinate on different objects are reduced.

We provide below a graphical representation of the discussion in the example.

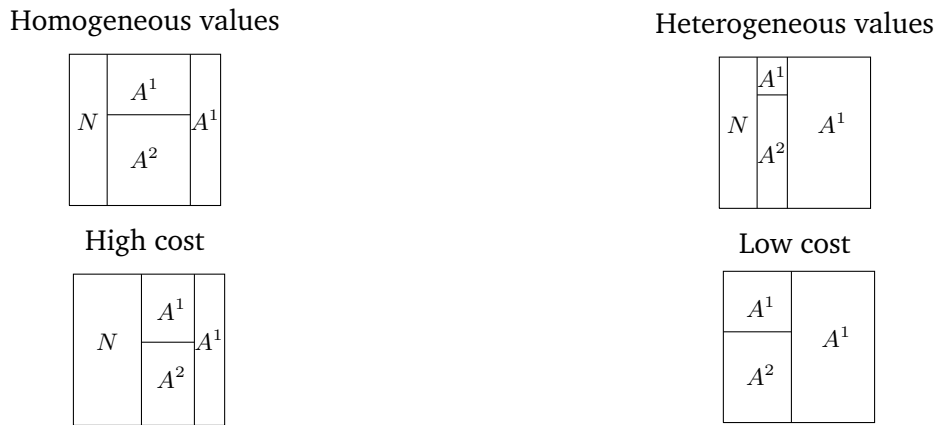


FIGURE 1.11: Comparative statics: Symmetric (interior) BNE for various values and cost - $m = 2$

Market balance

An increase in the number of agents n increases all thresholds, decreases class sizes. The effect is especially strong on bottom classes. It increases (decreases) probability levels for low (high) value objects. In net effect, introducing more agents decreases the ex ante probability with which any action is played, except for the no application action, which is played more frequently. The rationale is that more numerous agents generate competition via an increase in the probability of crowding, hence in the occurrence of failure. This pushes agents to be more cautious: high score agents mix with lower value objects, and low score agents more often decide not to apply.

Due to the recursive structure of the AG, the introduction of an additional object has a very clean effect on the equilibrium structure, described in the coming proposition.

Proposition 5. *[Effect of additional object]*

In an Application Game, the addition of an object with a given value $a^{k_0+1} < a^{new} < a^{k_0}$:

- Only affects equilibrium thresholds t^k and levels $p_{[1,k]}^j$ on the adjacent higher class and on lower classes ($k \in \{k_0, \dots, m\}$).
- Does not affect the equilibrium application probabilities for all application actions A^j , $j \in \{1, \dots, m\}$ that were already available before the addition.

In other words, the addition of an object does not affect individual behavior at levels of scores where agents were all applying to higher-value objects. It does affect individual behavior at levels of scores where some were

applying to lower value objects. The collective behavior remains unchanged in the sense that each lower value object is played as often as before the addition: only the identities of the applicants are modified, not the mass. Playing the new object only happens at the detriment of the no application strategy.

Priority score distribution

Another key property of the equilibrium is that the whole effect of the distribution is captured in the thresholds hence in the class sizes. The next proposition formalizes this remark.

Proposition 6. *[Effect of priority score distribution]*

In an Application Game, the priority score distribution:

- *Does not affect equilibrium levels.*
- *Only affects the equilibrium thresholds, in a way that keeps the mass of each class fixed.*

In summary, we expect a narrow (wide) class at score levels featuring a high (low) concentration of agents. Whether narrow with many agents or wide with few agents, an equilibrium class always features the same mass. In particular, if we change the priority score distribution to a mean-preserving spread distribution, we will get that extreme (middle) classes become narrower (wider). In expectation, the number of agents playing each possible mixture will remain unchanged.

B.3 Extension: Endogenous cost and the dynamic

In reality, decentralized allocation is always a dynamic process. This is also true in the leading example of social housing in Paris, where each week, a new application round opens and new vacant accommodation becomes available.

There are two ways to include a dynamic in the model, thereby increasing its descriptive strength. The first and probably most natural way is to model the dynamic explicitly and study the subsequent dynamic search problem.²⁵ The second and undoubtedly more tractable way is to endogenize some previously exogenous parameters of the static model to display the main effects of the dynamic while remaining in a static and simple framework. In this line of idea, it must be considered that in a dynamic version of the allocation, an agent with a high priority who fails in a given round necessarily keeps a high chance of being allocated an object in future rounds. Her continuation value is high. One way to capture this effect in the static model is to have the cost depend negatively on the priority score. C is now a strictly decreasing function of ω : $c'(\omega) < 0$. We set $c(0) < a^m$ implying $\forall \omega \in [0, 1], j \in \{1, \dots, m\}, c(\omega) < a^j$ (the cost never exceeds the value of any object).

Proposition 7. [BNE with score-dependent cost]

In the application game with score-dependent strictly decreasing cost $c(\omega)$, $c'(\omega) < 0$, a symmetric (interior) Bayes-Nash equilibrium:

- (1) Exists and is unique.
- (2) Is similar to the symmetric Bayes-Nash equilibrium of the Application Game with exogenous cost:
 - (i) If $t^m < 0$, the equilibria are exactly the same with exogenous and score-dependent costs.
 - (ii) If $t^m > 0$, the equilibria are the same except that t^m is higher in with endogenous cost and the m^{th} class $[t^m, t^{m-1}]$ is narrower.

In the proof for the exogenous cost model equilibrium, it was already apparent that thresholds and probability levels in the domain where agents apply with full probability were independent of the cost (with cost functions canceling out on both sides of indifference differential equations). Only the indifference equations between application actions A^j and no application N at t^m feature the application cost on one side.

The implication is that with an endogenous cost rising sharply, the agent applies a little less to all objects by being more prudent at low scores. The lowest value objects suffer the larger decrease in applications.

For illustration, we display the symmetric equilibrium for the small dimensional AG with a score-dependent (linear) cost: $c(\omega) = \frac{3}{2} - \omega$. For comparison, we also display the equilibrium on the same market with exogenous cost $c = 1$.²⁶

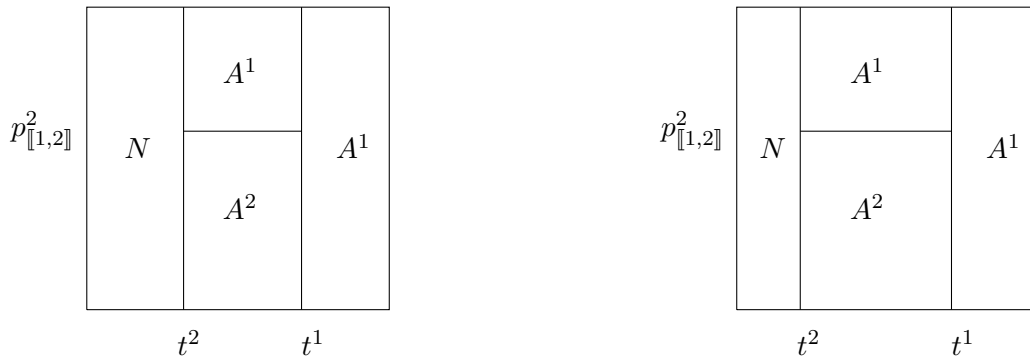


FIGURE 1.12: Endogenous cost $c(\omega) = \frac{3}{2} - \omega$ (left) vs constant cost $c = 1$ (right)
Symmetric (interior) BNE $m = 2$, $n = 3$, $a^2 = 2$, $a^1 = 4$, $F \sim \mathcal{U}$

Corollary 3. [Welfare with endogenous cost]

On the frictional market with score-dependent strictly decreasing cost $c(\omega)$, $c'(\omega) < 0$, the equilibrium interim expected payoff rises faster ($\frac{\partial W^B(\omega)}{\partial \omega}$ larger) on $[t^m, 1]$ than with exogenous cost.

We illustrate this point below:

²⁵ Stationarity assumptions may help: agents and objects leaving the market are replaced by agents with similar priorities and objects with similar values, and agents staying on the market during several periods keep their priorities.

²⁶ Note that the expected cost is the same in both cases (equals 1).

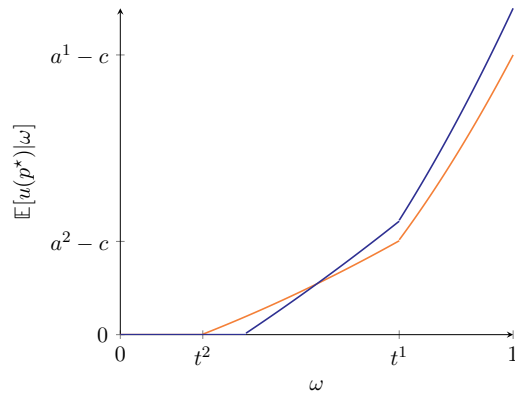


FIGURE 1.13: Endogenous cost $c(\omega) = \frac{3}{2} - \omega$ (blue) vs constant cost $c = 1$ (orange)
 Interim expected payoff at the symmetric (interior) BNE - $m = 2$, $n = 3$, $a^2 = 2$, $a^1 = 4$, $F \sim \mathcal{U}$

In conclusion, endogenous cost induces more discrimination according to score. The market outcome is closer to the outcome of the benchmark market (more discrimination). Accounting for the dynamic will reduce the magnitude of the welfare effects identified in section §5.2.

B.4 Extension: Larger or no truncation

In many real-life matching markets, agents are allowed to apply to more than one object. This is usually the case in school choice or centralized job market. In particular, the French national system for allocating teachers to schools uses a priority point system and a mechanism akin to a serial dictatorship where teachers can apply to several schools.²⁷

Thus, a natural extension of our model consists in relaxing or removing the truncation of the application menu.²⁸ To simplify, we process this extension within the framework of the example with $n = 3$, $m = 2$, $F \sim \mathcal{U}$. With no truncation, the action space includes an additional action B for “both” that consists of applying to both objects on the market (hence paying the application cost twice).

Preliminary results are summarized in the next proposition.

Proposition 8. [BNE with no truncation]

A symmetric (interior) Bayes-Nash equilibrium of the Application Game with $n = 3$, $m = 2$, $F \sim \mathcal{U}$ and no truncation:

- (1) Exists and is unique
- (2) (i) If $\frac{a^2}{a^1} + \frac{c}{a^2} > 1$, the symmetric equilibrium is the same as with the truncation. In particular, agents always apply to at most one object.
- (ii) If $\frac{a^2}{a^1} + \frac{c}{a^2} < 1$, at symmetric equilibrium: agents with large scores apply to object 1, agents with scores lower than a threshold r^1 apply to both objects.

Case (i), where agents disregard the possibility of applying to both objects, arises when the application cost is high relative to the objects’ values. In case (ii), agents use the possibility to apply to all objects, aiming for the high-value object, but hedging against the possibility that it may not be available anymore. Below r^1 , all interim expected payoffs strictly decrease due to competition on both objects, and what is the next shift in action is non-obvious (and non-robust).

The figure below illustrates the two cases:

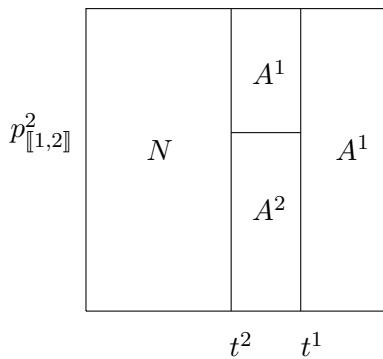


FIGURE 1.14: BNE with no truncation - Case (i)
 $m = 2$, $n = 3$, $a^2 = 2$, $a^1 = 4$, $c = 1.5$, $F \sim \mathcal{U}$

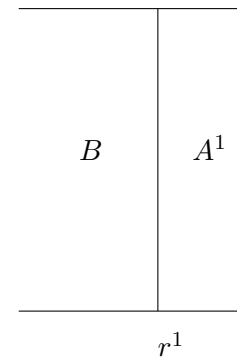


FIGURE 1.15: BNE with no truncation - Case (ii)
 $m = 2$, $n = 3$, $a^2 = 2$, $a^1 = 4$, $c = 0.5$, $F \sim \mathcal{U}$

B.5 Extension: Imperfectly correlated preferences

The assumption that preferences are homogeneous, with any agent assigning exactly the same value as her peers to any object, is rather restrictive. In social housing, for example, some criteria are valuable to all applicants (size of the accommodation, equipment), but applicants may value (for example) different micro-locations differently due, for example, to the location of their jobs. All in all, individual preferences likely combine common and idiosyncratic components.

We model imperfectly correlated preferences in a simple setting with $n = 2$ agents and $m = 2$ objects, and a uniform priority distribution. The objects can have two possible values $v > u > 0$, so that each agent has exactly one most preferred object with value v and one least preferred object with value u . A preference profile $(X_1 X_2)$, $X_i \in \{u, v\}$ means that object 1 has value X_1 to agent 1, X_2 to agent 2.

²⁷ In two steps: The first step manages allocation between regions, and the second stage within regions.

²⁸ Due to homogeneous preferences, the ranking between objects is common, and the action space consists of menus rather than rank-ordered lists.

The prior distribution over preference profiles is such that objects are the same ex-ante (they are equally likely to be each agent's most preferred object), but preferences are correlated (positive correlation when $\theta > \frac{1}{2}$):

$$\mathbb{P}(uu) = \mathbb{P}(vv) = \frac{\theta}{2}, \quad \mathbb{P}(uv) = \mathbb{P}(vu) = \frac{1-\theta}{2}, \quad \theta \in [0, 1]$$

We assume that preferences, just as priority scores, are private information. Thus, a type is two-dimensional: it specifies the priority (score) and the preference (most preferred object), with independence between the two dimensions.

In this setting, a strategy is a mapping of the score support into two possible actions: applying to one's most preferred object (denoted \oplus) or to one's least preferred object (denoted \ominus).

Proposition 9. [BNE with correlated preferences]

A symmetric (interior) Bayes-Nash equilibrium of the Application Game ($n = 2$, $m = 2$, $F \sim \mathcal{U}$):

(1) Exists and is unique.

(2) Has the following block structure:

(i) There are between 1 and 3 classes.

(ii) In the top class $[t^1, 1]$, the agent plays \oplus with full probability, $t^1 = \frac{\theta u - (1-\theta)v}{\theta v + (1-\theta)u}$.

(ii) Conditional on existence, in intermediate class $[t^2, t^1]$, the agent plays \oplus with probability $p(\oplus) = \frac{\theta u - (1-\theta)v}{(2\theta-1)(u+v)}$, \ominus with probability $p(\ominus) = 1 - p(\oplus)$

(iii) Conditional on existence, in the bottom class $[0, t^2]$, the agent does not apply.

At any equilibrium, and as expected, agents with high scores are ambitious and apply to their most preferred object (\oplus). It may become more profitable at lower scores (below a score t^1) to also target one's least preferred object (\ominus) because, in expectation, this object is less demanded. Interestingly, the block structure remains.

The difference with the perfect correlation case is that the shift at t^1 does not necessarily happen. Indeed, the fact all agents play \oplus combined with the imperfect correlation guarantees that both objects receive applications with positive probabilities. Even when both agents are ambitious, there is partial coordination. Thus, it can be that all agents keep on with the same strategy at low scores. The shift happens if and only if the correlation is sufficiently strong and the gap between the two object values is sufficiently small ($\theta > \frac{v}{v+u}$).

The figure below illustrates the two cases:

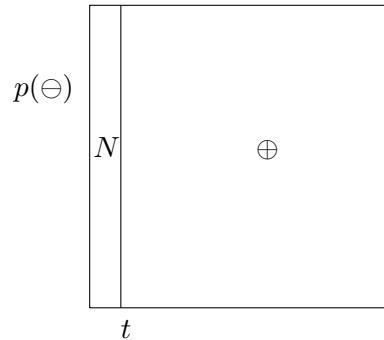
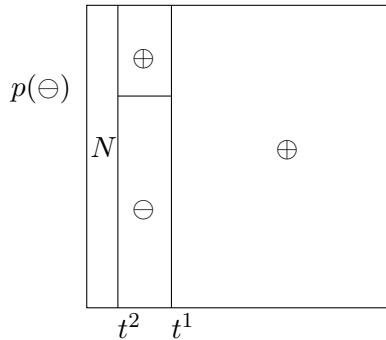


FIGURE 1.16: BNE with correlated preferences

FIGURE 1.17: BNE with correlated preferences

$m = n = 2$, $F \sim \mathcal{U}$, $v = 2$, $u = 1$, $c = 0.5$, $\theta = 0.75$ $m = n = 2$, $F \sim \mathcal{U}$, $v = 5$, $u = 4$, $c = 1$, $\theta = 0.75$

B.6 Proofs for supplements

Proof of Proposition 7.

The proof is similar to the proof of Theorem 1..

- On class 1 to m , the endogenous cost simplifies in the differential equations, and we get the same system than with exogenous cost.
- In the equation for the threshold t^m differs, the cost does not simplify:

$$\begin{aligned} \mathbb{E}[u_i(A_i^m(t^m), \sigma_{-i}([t^m, 1]))|t^m] &= \mathbb{E}[u_i(N(t^m), \sigma_{-i}([t^m, 1]))|t^m] \\ \iff (1 - p_{[1, m]}^m)(F(t^m) - F(t^{m-1}))^{n-1} a^m - c(t^m) &= 0 \\ \iff F(t^m) = 1 - m + \sum_{l=1}^m \left(\frac{c(t^m)}{a^l}\right)^{\frac{1}{n-1}} \end{aligned}$$

The difference in $F(t^m)$ in the endogenous cost model vs exogenous cost model is given by:

$$(c(t^m) - c) \sum_{l=1}^m \left(\frac{1}{a^l}\right)^{\frac{1}{n-1}} > 0 \iff c(t^m) > c$$

Proof of Corollary 3.

The interim expected payoff with endogenous cost is given by $(, \omega \in [t^k, t^{k-1}], k \in \{1, \dots, m\})$:

$$\begin{aligned} W_e^F(\omega) &:= \sum_{j=1}^k p_{[1, k]}^j \cdot \left(\left(\frac{a^k}{a^j}\right)^{\frac{1}{n-1}} - (F(t^{k-1}) - F(\omega)) p_{[1, k]}^j \right)^{n-1} - c(\omega) \\ \frac{\partial W_e^F(\omega)}{\partial \omega} &= \frac{\partial W^F(\omega)}{\partial \omega} - \frac{dc(\omega)}{d\omega} \end{aligned}$$

$c()$ decreasing with ω ($\frac{dc(\omega)}{d\omega} < 0$) implies $\frac{\partial W_e^F(\omega)}{\partial \omega} > \frac{\partial W^F(\omega)}{\partial \omega}$.

Proof of Proposition 8.

The interim expected payoffs at score 1 write:

$$\begin{aligned} \mathbb{E}[u_i(A_i^1(1))|1] &= a^1 - c \\ \mathbb{E}[u_i(A_i^2(1))|1] &= a^2 - c \\ \mathbb{E}[u_i(B_i(1))|1] &= a^1 - 2c \\ \mathbb{E}[u_i(N_i(1))|1] &= 0 \end{aligned}$$

So $\sigma_i^*(1) = A^1$. By continuity, $\exists \omega^1 < 1$ s.t. $\sigma_i^*(\omega^1, 1) = A^1$.

The interim payoffs at lower scores write:

$$\begin{aligned} \mathbb{E}[u_i(A_i^1(\omega), \sigma_{-i}^*[\omega, 1])|\omega] &= \omega^2 a^1 - c \\ \mathbb{E}[u_i(A_i^2(\omega), \sigma_{-i}^*[\omega, 1])|\omega] &= a^2 - c \\ \mathbb{E}[u_i(B_i(\omega), \sigma_{-i}^*[\omega, 1])|\omega] &= \omega^2 a^1 + (1 - \omega^2) a^2 - 2c \\ \mathbb{E}[u_i(N_i(\omega), \sigma_{-i}^*[\omega, 1])|\omega] &= 0 \end{aligned}$$

We solve indifference equations:

$$\begin{aligned} \mathbb{E}[u_i(A_i^1(\omega), \sigma_{-i}^*[\omega, 1])|\omega] &= \mathbb{E}[u_i(A_i^2(\omega), \sigma_{-i}^*[\omega, 1])|\omega] \iff \omega = \sqrt{\frac{a^2}{a^1}} \\ \mathbb{E}[u_i(A_i^1(\omega), \sigma_{-i}^*[\omega, 1])|\omega] &= \mathbb{E}[u_i(B_i(\omega), \sigma_{-i}^*[\omega, 1])|\omega] \iff \omega = \sqrt{1 - \frac{c}{a^1}} \\ \sqrt{\frac{a^2}{a^1}} < \sqrt{1 - \frac{c}{a^1}} &\iff \frac{a^2}{a^1} + \frac{c}{a^2} > 1 \end{aligned}$$

We get two cases:

- $\frac{a^2}{a^1} + \frac{c}{a^2} > 1$: Below a threshold $t^1 = \sqrt{\frac{a^2}{a^1}}$, players switch to playing A^1 and A^2 with indifference. Just below t^1 and similarly to the case with truncation, the agents plays A^2 with probability $p_{[1,2]}^2 = \frac{1}{1+t^1}$. At a score $\omega < t^1$, interim payoffs write:

$$\begin{aligned}\mathbb{E}[u_i(A_i^1(\omega), \sigma_{-i}^*([\omega, 1]))|\omega] &= \mathbb{E}[u_i(A_i^2(\omega), \sigma_{-i}^*([\omega, 1]))|\omega] = (1 - (1 - t^1) - (t^1 - \omega)p_{[1,2]}^1)^2 a^1 - c \\ \mathbb{E}[u_i(B(\omega), \sigma_{-i}^*([\omega, 1]))|\omega] &= (1 - (1 - t^1) - (t^1 - \omega)p_{[1,2]}^1)^2 a^1 + (1 - (t^1 - \omega)p_{[1,2]}^2) a^2 - 2c \\ \Delta(\omega) &:= \mathbb{E}[u_i(A_i^1(\omega), \sigma_{-i}^*([\omega, 1]))|\omega] - \mathbb{E}[u_i(B(\omega), \sigma_{-i}^*([\omega, 1]))|\omega] = c - (1 - (t^1 - \omega)p_{[1,2]}^2)^2 a^2 \\ \frac{\partial \Delta}{\partial \omega} &= -2p_{[1,2]}^2 (1 - (t^1 - \omega)p_{[1,2]}^2)^2 a^2 < 0\end{aligned}$$

So Δ is decreasing until t^1 . And by definition of this case: $\Delta(t^1) = 0$. So on the left of t^1 , $\Delta(\omega) > 0$. The agent does not switch to B . Just as in the model with truncation, he randomizes between A^1 and A^2 potentially until a threshold t^2 where he starts playing N . Below t^2 all interim payoffs stay constant, so the agent keeps on playing N until score 0.

- $\frac{a^2}{a^1} + \frac{c}{a^2} < 1$: Below a threshold $r^1 = \sqrt{1 - \frac{c}{a^2}}$, players switch to playing B . At a score $\omega < r^1$, interim payoffs write:

$$\begin{aligned}\mathbb{E}[u_i(A_i^1(\omega), \sigma_{-i}^*([\omega, 1]))|\omega] &= \omega^2 a^1 - c \\ \mathbb{E}[u_i(B(\omega), \sigma_{-i}^*([\omega, 1]))|\omega] &= \omega^2 a^1 + ((1 - r^1)^2 + 2\omega(1 - \omega)) a^2 - 2c \\ \Delta(\omega) &:= \mathbb{E}[u_i(B(\omega), \sigma_{-i}^*([\omega, 1]))|\omega] - \mathbb{E}[u_i(A_i^1(\omega), \sigma_{-i}^*([\omega, 1]))|\omega] := ((1 - r^1)^2 + 2\omega(1 - \omega)) a^2 - c \\ \frac{\partial \Delta}{\partial \omega} &= 2(1 - 2\omega) a^2 > 0 \iff \omega < \frac{1}{2}\end{aligned}$$

So Δ is decreasing on the left neighborhood of r^1 . And by definition of this case: $\Delta(r^1) = 0$. So on the left neighborhood of r^1 , $\Delta(\omega) > 0$, and the agent does not immediately switch to another action.

Proof of Proposition 9.

The interim expected payoffs at score 1 write:

$$\begin{aligned}\mathbb{E}[u_i(\oplus(1)_i)|1] &= v - c \\ \mathbb{E}[u_i(\ominus(1)_i)|1] &= u - c\end{aligned}$$

So $\sigma^*(1) = \oplus$. By continuity: $\exists t^1 < 1$ s.t. $\sigma^*((t^1, 1]) = \oplus$

The interim expected payoffs at score $\omega < 1$ write:

$$\begin{aligned}\mathbb{E}[u_i(\oplus_i(\omega), \sigma_{-i}^*|\omega)] &= (1 - (1 - \omega)\theta)v - c \\ \mathbb{E}[u_i(\ominus_i(\omega), \sigma_{-i}^*|\omega)] &= (1 - (1 - \omega)(1 - \theta))u - c\end{aligned}$$

We characterize the threshold point t^1 where the agent start being indifferent between the two actions:

$$\begin{aligned}\mathbb{E}[u_i(\oplus_i(t^1), \sigma_{-i}^*|t^1)] &= \mathbb{E}[u_i(\ominus_i(t^1), \sigma_{-i}^*|t^1)] \iff (1 - (1 - t^1)\theta)v = (1 - (1 - t^1)(1 - \theta))u \\ &\iff t^1 = \frac{\theta u - (1 - \theta)v}{\theta v + (1 - \theta)u}\end{aligned}$$

We find $t^1 > 0 \iff \theta > \frac{v}{v+u}$, and $\frac{v}{v+u} > \frac{1}{2}$ so the condition is non trivial.

The interim expected payoffs at score $\omega < t^1$ write:

$$\begin{aligned}\mathbb{E}[u_i(\oplus_i(\omega), \sigma_{-i}^*|\omega)] &= [\omega + ((1 - t^1) + (t^1 - \omega)p(\oplus))(1 - \theta) + (t^1 - \omega)(1 - p(\oplus))\theta]v - c \\ \mathbb{E}[u_i(\ominus_i(\omega), \sigma_{-i}^*|\omega)] &= [\omega + ((1 - t^1) + (t^1 - \omega)p(\oplus))\theta + (t^1 - \omega)(1 - p(\oplus))(1 - \theta)]u - c\end{aligned}$$

The probability level $p(\oplus)$ making the agent indifferent between the two actions is characterized by:

$$\begin{aligned}\mathbb{E}[u_i(\oplus_i(\omega), \sigma_{-i}^*|\omega)] &= \mathbb{E}[u_i(\ominus_i(\omega), \sigma_{-i}^*|\omega)] \\ \iff [\omega + ((1 - t^1) + (t^1 - \omega)p(\oplus))(1 - \theta) + (t^1 - \omega)(1 - p(\oplus))\theta]v &= [\omega + ((1 - t^1) + (t^1 - \omega)p(\oplus))\theta + (t^1 - \omega)(1 - p(\oplus))(1 - \theta)]u\end{aligned}$$

We differentiate this equation and get:

$$\begin{aligned} [1 - p(\oplus)(1 - \theta) - (1 - p)\theta]v &= [1 - p\theta - (1 - p)(1 - \theta)]u \\ \iff p(\oplus) &= \frac{\theta u - (1 - \theta)v}{(2\theta - 1)(u + v)} \end{aligned}$$

We check that this level indeed verifies the indifference equation.

The interim payoff at scores below t^1 strictly decreases and may hit the zero bound at some lower score t^2 , where the agent with lower score would decide not to apply.

Chapter 2

HOW CAN I KNOW HOW MUCH I LIKE YOU?

A Heuristic Approach to Matching and Stability

Laure GOURSAT* – December 2023

We thank Yair ANTLER, Francis BLOCH, Miguel BALLESTER, Catherine BOBTSCHEFF, Pierre BOYER, Olivier COMPTE, Guillaume HAERINGER, Kevin HE, Philippe JEHIEL, Frederic KOESSLER, Antonin MACÉ, Eduardo PEREZ-RICHET, George MAILATH, Debraj RAY, Alex REES-JONES, Larry SAMUELSON, Rani SPIEGLER, Olivier TERCIEUX, Nikhil VELLODI, Rakesh VOHRA and seminar participants at Paris School of Economics and University of Pennsylvania for helpful comments and conversations. We thank Ecole Nationale des Ponts et Chaussées (doctoral funding contract n°20/092) for financial support.

Abstract: On a marriage market with unknown preferences (agents only observe the current matching and realized match utilities), we define a novel and natural heuristic of belief formation (valuation), which incorporates a famous and documented cognitive bias (the projection bias). Under this heuristic, an agent estimates a counterfactual match utility by extrapolating from realized match utilities: his own utility and the weighted average utility of all current partners of the targeted partner's type. We study how this reshuffles the market outcome, as given by pairwise stable matchings when agents have valuation beliefs (v-stability). When restricting our attention to pure matchings, we find that v-stability is equivalent to any two partners holding the same rank according to current utilities (happiness sorting). The predictions under specific preference structures are then straightforward. The alignment of interests across the market governs the size of the v-stable set from empty to maximal. The correlation of preferences by agent or target stabilizes the positive assortative matching. For a generic market, though, we get neither the existence of a pure v-stable matching nor the convergence of a dynamic blocking pair process (predicting persistent moves on the market). The most general version of the model defines a notion of mixed matching, characterizing the proportions of each productive type matched with each partner type. Our main result is a general existence theorem for v-stable matchings in the mixed extension.

Keywords: Matching markets, incomplete information, belief formation, pairwise stability.

JEL codes: C78, D91.

*Ph.D. candidate at Paris School of Economics (PSE) - 48 boulevard Jourdan, 75014 Paris, France (Office R6-41) - laure.goursat@psemail.eu

1 Introduction

Matching and complexity

Nobel prize Alvin Roth defines a matching market as “a market in which prices do not do all the work” (Roth (2017) [56]), implicitly comparing standard markets for goods and matching markets. A standard market connects sellers willing to sell goods and buyers willing to buy and consume goods. All relevant information is summarized in the price, which acts as a signal, and agents just need to solve a decision problem: whether they want to buy / sell, at a given price. A matching market connects agents with other agents willing to form productive partnerships. Prices are either absent or set in advance, they do not result from bargaining, and lose their signaling role. An important corollary is that a matching market participant cannot just choose an object as long as she can afford it, she has to be chosen as well. The unfortunate consequence is that matching markets are utterly complex environment for market participants, in at least two ways.

Limited information on preferences

First, playing optimally demands information that may not be available prior to matching. In particular, matching market participants need knowledge about their own preferences. But intuition suggests it may be hard to assess how happy one would be in a match prior to experiencing the match. Suppose for example that the match utility of agent m when matched with f writes like a function taking as arguments two vectors of characteristics, C_m for the agent m , C_f for the partner f : $u_{mf} := u_m(C_m, C_f)$. We may assume that agent m perfectly knows himself, his characteristics C_m , and what he values in matching (the functional form u_m). In many real-life applications though, assuming prior knowledge of the full vectors of characteristics C_f of potential partners f seems unrealistic, hence too demanding. Indeed, in most empirical contexts, information does not come for free. Pre-matching information acquisition is either impossible (when accepting a job position, some aspects of the jobs (quality of daily interactions with colleagues, effective workload) are not written in the contract and can only be discovered through experience in the job), constrained (a high school student who wants to learn about colleges is time-constrained and will not be able to learn about all the colleges she could rank in the mechanism) or costly (in the school choice example, the best way to acquire information is to visit the colleges’ campuses, hence moving effort and travel costs). In a rational inattention approach,¹ an optimizing agent could decide to remain (partially) uninformed. Empirical evidence from real-life matching markets² shows that participants indeed have limited information about their options, their preferences, and priorities.

However, we argue there is one piece of information that agents have access to. In general, participants observe the current state of the market: who is matched with whom and the realized match utilities. For marriage: one observes who live together and gets a clue of how happy each partner is in the marriage (through facial expressions, body language or small talks). This defines a social learning environment, where one may learn from other agents’ experience. To fix ideas, consider a case of correlated preferences: assume that all males (females) tend to value males (females) similarly. Then if one agent A observe that an agent B is very happy in a given match with another agent C , this is a signal that C has valuable productive characteristics. Agent A should infer that she could also be quite happy in a match with C .

Bounded cognition³

Second, in the contexts where information is indeed imperfect, (social) learning is complex. This

¹See Sims (2003) [63].

²Hoxby and Avery (2013) [32], Wiswall and Zafar (2015) [69], Kapor, Neilson and Zimmerman (2020) [38]. Evidence is provided through survey and treatment (when one gives agents more information, the strategies change).

³We make a central distinction: between instrumentation (the ability given beliefs on the environments, to optimize) and cognition (ability to form correct beliefs on the environment). The present paper does not question perfect instrumentation in matching environments, and rather argue that cognition is the challenging part.

is because the vector of individual characteristics relevant for matching is multi-dimensional.⁴ In this respect, the Bayesian paradigm appears far too demanding. It demands that participants store and process a large quantity of data in a highly sophisticated way.

There is overwhelming evidence (coming from the field or the lab)⁵ that agents are unable to do so. They rather have systematic biases in the way they analyze the data, mostly in a way that simplifies the inference they make. This results in important strategic mistakes (typically, even under the strategy-proof Deferred Acceptance mechanism, they would submit truncated Rank-Ordered List and not necessarily rank their favorite option first).

Research question

The issue is, under the type of environment we describe (with limited information on preferences and bounded cognition), the stability properties of centralized matching mechanisms (such as the celebrated Deferred Acceptance mechanism by Gale and Shapley (1962) [25]) or decentralized matching procedure (such as the dynamic blocking pair process studied by Knuth (1976) [40] and Roth and VandeVate (1990) [58]) no longer hold. Hence the need to develop a new approach to stability.

In this paper, we explicitly model incomplete information on preferences. We set a specific but natural information structure where agents perfectly observe all realized match utilities but do not have access to counterfactual utilities. We define a heuristic way in which agents form beliefs on their own preferences in this environment (the “valuation”).

The question is then: how does this belief formation process affect the (long-run) matching outcome? We answer by applying the usual pairwise stability solution concept to a market where all agents have valuation beliefs (“v-stability”). We characterize v-stable matchings.

Summary model and results

Under the valuation heuristic, an agent builds an estimate of the counterfactual utility that she would get in a blocking pair by extrapolating from current realized utilities. Thus, agents holding valuation beliefs base their blocking decisions on cardinal comparisons of current utilities across their own side of the market. In this sense, the heuristic can be connected to the famous and documented “projection bias” (referring to the tendency to overestimate similarities between oneself and others).

In the pure version of the model, the agent extrapolates from only two signals: his own utility and the utility of the current partner or the targeted partner. A necessary and sufficient condition for v-stability is happiness sorting: any two partners must hold the same rank according to realized match utilities. The predictions under specific preference structures are straightforward. The alignment of interests across the market governs the size of the v-stable set from empty to maximal. The correlation of preferences by agent or target makes the positive assortative matching the unique stable matching. For a generic market, though, we get neither the existence of a v-stable matching nor the convergence of a dynamic blocking pair process (predicting persistent moves on the market).

In the most general, mixed, version of the model, the market is made of a continuum of agents belonging to discrete types. A mixed matching describes the proportions of each male type matched with each female type. The valuation aggregates the utilities of all types matched with the targeted type. We define a v-stable matching as a steady state of the dynamic blocking pair process, where blocks that keep the proportions of types matched unchanged are allowed. We get existence of a v-stable matching in the generic case. The existence theorem proof for mixed v-stable matchings is a standard fixed point proof using Kakutani’s theorem.

The value of this analysis is two-fold. First, it connects in various ways with stylized facts from the empirical marriage market. As an example, the “happiness sorting” condition can be read as a formalization of the famous maxim “You can’t be happier than your wife” which can account to a large

⁴On a marriage market, Becker (1973) [9] quotes some the individual characteristics that he thinks may be determinant for productivity in marriage: health, fertility, cooking abilities...

⁵See Rees-Jones and Shorrer (2023) [52] for a review on education markets.

extent for divorces (Guven, Senik and Stichnoth (2012) [29]). Second, it has an interpretation in terms of the long run dynamic that we can expect on real-life markets. Overall, the analysis predicts significantly more moves than with perfect information.

Outline of paper

The rest of the paper is structured as follows. Section §2 reviews the existing literature on behavioral matching.⁶ Section §3 defines and defends the valuation heuristic in its pure version as a natural belief formation process for matching markets with limited information. Section §4 characterizes pure v-stable matchings in a general environment as well as for standard preference structures. The analysis is both static - answering the “what” (what is v-stable?) question -, and dynamic - answering the “how” (how do we get to a v-stable matching?) question. It emphasizes the lack of existence of a v-stable matching and the lack of convergence of a dynamic blocking pair process in the general case. Section §5 defines a notion of mixed matching over a two-sided population of types. Section §6 applies a standard fixed point theorem in this new setting and restores existence of v-stable matchings. Section §7 proposes various paths to extend the pure model. Section §8 concludes. All proofs are available in appendix §A.

2 Literature review

Limited information

We are not the first ones to raise the issue of limited information on matching markets. Actually, Alvin Roth early on identified modeling of information as an important next step in the analysis of matching mechanisms as this quote can testify: “[Common knowledge of preferences] is not an accurate description of the situation prevailing in the kinds of markets to which these models can mostly be applied. [...] Some of the conclusions reached about strategic decisions are particularly sensitive to the assumption of complete information.” (Roth (1989) [55]). Since then, a literature modeling either private information on preferences⁷ or allowing information acquisition⁸ has emerged (still modelling perfectly rational agents).

Bounded rationality

The literature on behavioral matching remains quite scarce, but growing. Behavioral matching papers typically plug a known behavioral bias (such as non standard preferences)⁹ into the standard matching framework. Their focus is mostly on the strategic interactions (non-cooperative game theory).

On the bounded instrumentation side, Caspari and Khana (2021) [12] model agents who are unable to choose consistently between potential partners when the menu size exceeds two. They derive necessary and sufficient conditions on the choices for existence of a pairwise stable matching. In a large market setting, Artemov, Che and He (2023) [6] propose a solution concept (robust equilibrium) that requires only an asymptotically optimal behavior. They use it to study large random matching markets operated with the applicant-proposing Deferred Acceptance. Although applicants may be non-truthful in Deferred Acceptance, the outcome is arbitrarily close to the stable matching.

On the bounded cognition side, Antler and Bachi (2022) [3] apply cursed equilibrium and analogy-based expectation equilibrium to a dynamic search market to account for an important empirical puzzle

⁶In section §3 and §5, we also connect the heuristic with the economic literature on projection bias and bounded rationality through aggregation.

⁷Wilson (1978) [68], Roth (1989) [55], Nagypal (2004) [49], Dutta and Vohra (2005) [18], Ehlers and Masso (2007) [19], Lee (2009) [43], Hoppe, Moldovanu and Sela (2009) [31], Chakraborty, Citanna and Ostrovsky (2010) [14], Chade, Lewis and Smith (2014) [13], Goursat (2023) [28].

⁸Bade (2015) [8], Hagenbach, Koessler and Tregouët (2017) [30], Immorlica, Lesno, Lo and Lucier (2020) [33], Gleyze and Pernoud (2020) [26], Artemov (2021) [5], Chen and He (2021) [15], Bucher and Caplin (2021) [11].

⁹Meisner (2023) [47] and Meisner and Von Wangenheim (2023) [48] introduce expectation-based loss aversion into a school-choice setting, to account for students playing dominated strategies in strategy-proof mechanisms. Ranking a preferred school below a less preferred one is a way to avoid disappointment. Depending on the preferences, this may result in inefficiencies and instabilities at equilibrium.

in dating markets. They show that when participants are cursed or coarse, the matching technology becomes less efficient as the frictions vanish, with some agents remaining single forever. Li (2017) [44]’s concept of obvious strategy-proofness also has a behavioral interpretation: a mechanism is obviously strategy-proof if and only if an agent who is unable to understand how states affect outcomes could recognize weakly dominated strategies. But since Ashlagi and Gonczarowski (2018) [7] have shown that no stable mechanism is obviously strategy-proof (hence no hope for mechanism design that would be robust to bounded cognition), we need to explicitly model cognition on the market.

Liu, Mailath, Postlewaite and Samuelson (2014) [46]¹⁰ also model non-Bayesian inference. In their job market example, firms infer about workers’ types based on the common knowledge of stability. Firms do not compute a posterior distribution but hold a set of “reasonable” beliefs, beliefs that are consistent with the premise that the observed matching is stable. Firms are modelled as prudent: they are willing to form a blocking pair only if the block is beneficial under any reasonable belief.

Richter and Rubinstein (2023) [53] also has a behavioral flavour. In the context of the roommate problem, they define a notion of unilateral stability, considering that harmony in society may be disrupted by a unilateral proposition to form a blocking pair. Their favorite interpretation is a story of endogenous preferences: When A , currently matched with B , approaches C , B (C) feels betrayed (flattered) and this makes A (less) (more) attractive to B (C).

For the specific field of application of the education markets, Rees-Jones and Shorrer (2023) [52] provide an exhaustive review of empirical works documenting deviations from the rational theory. They also list behavioral theories that have been presented as possible explanations for the deviations. Our model would fall into the category of “application of improper heuristics”. It is consistent with Rees-Jones and Shorrer (2023)’s intuition that most students do not fully attempt to understand the matching procedure and instead merely apply rules that served them well in the past.

When introducing the model, we connect the defined belief formation process to various existing notions of bounded rationality in games.

3 Model: Valuation and v-stability (pure)

3.1 The marriage market

The environment is a marriage market:¹¹ two-sided, discrete (one side of males $m \in M$, one side of females $f \in F$), balanced ($|M| = |F| = n$ agents on each side), and without monetary transfers.

A full matching $\mu \in \Sigma$ is a one-to-one mapping from the set of agents to itself: $\mu : M \cup F \rightarrow M \cup F$ $\mu(m) = f \iff \mu(f) = m$. The set of full matchings is denoted Σ .

A male $m \in M$ receives a utility $u_{mf} \in \mathbb{R}$ when matched with a female $f \in F$.¹²

In the baseline model, we do not allow agents to stay unmatched.¹³ This is equivalent to restricting our attention to markets where there is a positive matching value¹⁴ (and agents understand this). For example, we can normalize all utilities of singles to zero and set all match utilities positive.

3.2 The information

The information available on the market is limited but free and fully symmetric. Agents observe the current state of the market, that is two things. They observe the current matching μ (who is matched with whom) and the set of all realized utilities $u_{m\mu(m)}, \forall m \in M, u_{f\mu(f)}, \forall f \in F$. So if m and f

¹⁰Followers: Liu (2020) [45], Chen and Hu (2020) [15], Bichkandani (2017) [10], Pomatto (2022) [51].

¹¹We appeal to the marriage market vocabulary for illustration. This semantic choice should not conceal the generality of the results and the variety of potential applications (in particular, to wage-regulated job markets).

¹²We define notations for true utilities and beliefs over utilities from the male point of view, but all notations will be symmetric: u_{fm} is the utility of the female in the pair (m, f) .

¹³In section §7.1, we extend the model to study v-stability of any matching, including non full matchings.

¹⁴This case is very frequent in the matching literature. The interpretation for housing is that a household prefers to be assigned any house rather than stay homeless. In education, a student wants to go to college anyway.

are currently matched, anyone on the market observes u_{mf} , u_{fm} , so knows precisely how much each partner is happy in his or her marriage. The interpretation is that feedback information about current satisfaction is easy to get from facial expressions, body language or small (non-strategic) talk.

But agents do not have access to counterfactual match utilities u_{mf} , u_{fm} , $\mu(m) \neq f$. If m and f are not currently matched together, no one knows about u_{mf} , u_{fm} .¹⁵

FIGURE 1 below summarizes the information structure for a market of size 2. If the current matching is μ (parallel lines), the matching and all green utilities are observed by anyone. Any utility arising from the matching μ' (crossed lines) in red is not available data.



FIGURE 2.1: Information structure

In green (red): data available (unavailable) when the current matching is μ

3.3 The valuation

Participants on the defined marriage market are potentially interested in changing partners. To assess whether that would be beneficial, they must form beliefs on counterfactual match utilities.

We define a belief formation process termed “valuation”. More specifically, we put a constraint on the beliefs that agents may have about counterfactual utilities.

We define the valuation from the point of view of male m . To clarify the notation, we represent on FIGURE 2 a subset of the market where one male m is currently matched with a female $\mu(m)$ and considers rematching with a target female f (whose current partner is $\mu(f)$). We denote in green the realized utilities from the male side.

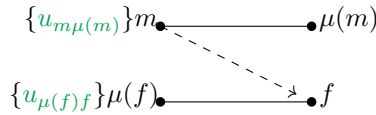


FIGURE 2.2: Sub-market of size 2 - Male m and target f

Definition 1.

The **valuation** is a function mapping a male $m \in M$, a female $f \in F$ and a matching $\mu \in \Sigma$ to a possible belief of male m over the utility u_{mf} that he would get with a targeted partner f when the current matching is μ :

$$v : M \times F \times \Sigma \Rightarrow \mathbb{R}$$

$$v_{mf}^{\mu} \in (u_{m\mu(m)}; u_{\mu(f)f})$$

\leftrightarrow

The belief is constrained to lie between two realized matched utilities: $u_{m\mu(m)}$ is the utility that m currently gets in his own match with his current partner $\mu(m)$. $u_{\mu(f)f}$ is the utility that the current partner of the target $\mu(f)$ gets in his match with the target f . The double arrow simply means that we ignore a priori the order between the two utilities. An important specific case is when $f = \mu(m)$, then v is a singleton of the true match utility u_{mf} . This is consistent with the assumption that agents observe all match utilities, including their own.

This valuation is a simple and, we believe, natural heuristic given the hard informational constraint that feedback data is on realized match utilities only.

¹⁵Differently from standard models of limited information in economics, we do not define a prior distribution over counterfactual match utilities. It would be useless for the analysis as the belief formation process we define is non-Bayesian and prior-free.

Clearly, this is a non-Bayesian estimation of the true utility, and prior-free. It defines a kind of belief support without a distribution.

The valuation beliefs are endogenous to the current state of the market (and only to this - agents do not use memory of past matchings).¹⁶ This endogeneity will play a major role in the negative results in the next section §4.

As simple as it is, this heuristic does not force the preferences to be the same for all agents belonging to one side of the market: the preferences may still differ cardinally and ordinally (due to the idiosyncratic bound $u_{m\mu(m)}$).

An equivalent formulation would say that the belief can be any convex combination of the two utilities:

$$v_{mf}^\mu \in \{(1 - \lambda)u_{m\mu(m)} + \lambda u_{\mu(f)f}, \lambda \in (0, 1]\}.^{17}$$

3.4 Connection with literature

Endogenous preferences

The valuation model introduces endogenous preferences: preferences about future partners depend on the current matching. This is not completely new in the literature.

Liu, Mailath, Postlewaite and Samuelson (2014) [46] also have this endogeneity. In their case, the incomplete information agents (firms on a job market) only use the matching (and their knowledge of stability) to infer about the unknown workers' types. Firms do not have access to workers' realized utilities. The approach to stability is almost opposite to ours. The inference is all about high-order reasoning on selection effects, which a valuation agent completely neglects. Their stability definition crucially rests on the firm having multiple reasonable beliefs (whereas the valuation is a point estimator). Their results are drawn by the specific (asymmetric) information structure, (super-modular) payoff function, and by the possibility for transfers.¹⁸ The “valuation” approach applies to a symmetric environment, and is valid under any payoff functional form.

In Antler (2015) [1], preferences are endogenous to the strategies in a centralized mechanism. This can model situations where an agent finds a target more attractive simply because he observes that the target has ranked him high in her rank-ordered list. Antler (2015) is interested in the strategic aspects and models the associated submission game (whereas agents in our model solve a decision problem). A submission strategy affects the matching through two channels: the direct channel, and an indirect channel (by modifying the preferences and strategies of others). Their main theorem is a negative result, showing that the existence of a stable mechanism (in the exogenous case of Gale and Shapley (1962) [25]) is non-robust to the introduction of endogeneity in the preferences. There always exist strategies and preference functions such that the strategies form a Nash equilibrium of the submission game, but the resulting allocation from the mechanism is unstable with respect to the realized preferences.

Antler (2019) [2] defines a strategic setting between one employer and many agents. Agents form beliefs about how much employers like them from the order in which they approach them, and it affects their preferences and willingness to take the jobs.

Projection bias

In the economic and psychology literature, the term “projection” refers to the ability to assess how similar or different others are from ourselves. Daily life experience suggests this ability can prove helpful, in particular in social learning settings, where we can learn from others' outcomes or actions.

Empirical evidence coming from the psychology and economics literature¹⁹ reports that agents

¹⁶In appendix §B.1, we show that the introduction of memory in the valuation heuristic does not modify the main no existence results for v -stable pure matchings. We therefore suspect that the reason for the no existence is in the behavioral bias rather than in the memory-less assumption. This further motivates our mixed approach in section §5.

¹⁷The weight λ can be different for different agents and pairs, can also be endogenous to the matching. It could be a measure of the distance between agents characteristics: agents would identify more strongly agents with similar characteristics together.

¹⁸As Goursat (2023) [27] shows.

¹⁹See Dawes (1989) [16], Krueger and Clement (1994) [41].

are unable to project correctly, hence a “projection bias”. Specifically, they tend to overestimate the similarities, in a way that is inconsistent with Bayesian rationality.

Most of this literature focuses on “preferences projection” or “taste projection”. For instance, it has been reported that people overestimate how many agents share their tastes for consumption goods (Ross, Greene and House (1977) [54]), political candidates (Delavande and Manski (2012) [17]), risk (Faro and Rottenstreich (2006) [21]), and temporary emotions like hunger and thirst (Van Boven and Loewenstein (2003) [65]).

In a theory paper, Gagnon-Bartsch (2017) [23] incorporates taste projection into canonical models of observational learning, where individuals choose between options, and payoffs depend both on an unknown state of the world (common preferences) and on an idiosyncratic taste (private preferences). In this environment, the private preferences are most often private information. Therefore, an agent willing to learn about the state must disentangle between the role of the idiosyncratic taste and the role of the private information about the state in accounting for different actions. A “taste projector” wrongly interprets any difference in action as coming from differences in the information. Gagnon-Bartsch (2017) studies the consequence in terms of beliefs and actions: “taste projection” leads to too little dispersion in actions. Gagnon-Bartsch et al. (2022) [24] illustrate the point on a market for goods between consumers uncertain about the quality and a monopoly.

The valuation heuristic implicitly assumes inter-personal utility comparisons.²⁰ The agent extrapolates based on two realized utilities.

Using the realized utility $u_{\mu(f)f}$ as a signal of the counterfactual utility u_{mf} is correct only if the two male agents $\mu(f)$ and m have the same preferences. This can be interpreted as a projection bias, in the usual sense of taste projection, happening within one’s own side of the market.²¹

Similarly,²² using $u_{m\mu(m)}$ as a signal of u_{mf} is correct only if the two female agents $\mu(m)$ and f have the same productive characteristics. This can be interpreted as a projection bias, where the agent overestimates the similarity of characteristics of agents from the other side of the market.

3.5 V-stability

We define “v-stability” as the usual pairwise stability notion for matchings markets (Gale and Shapley (1962) [25]) when everyone on the market has valuation beliefs.

Definition 2.

μ is **v-stable** if: $\exists v$ valuation s.t.: $(NBP)_\mu^v : \forall (m, f): \begin{cases} v_{m\mu(m)}^\mu \geq v_{mf}^\mu \\ \text{or} \\ v_{f\mu(f)}^\mu \geq v_{fm}^\mu \end{cases}$

We demand that there is no blocking pair: there should be no pair of agents that would both strictly prefer to match together than to stay with their current partners.

We have defined v-stability in definition 2. asking that just one valuation function v satisfies the (NBP) condition. The next lemma shows that this is equivalent to all possible functions v satisfying the inequalities. Said differently whether a matching is v-stable or not is robust to the exact specification of the weight λ in the convex combination $v_{mf}^\mu = (1 - \lambda)u_{m\mu(m)} + \lambda u_{\mu(f)f}$.

²⁰In section §7.2, we provide a generalization of the heuristic that drops the interpersonal comparisons, while preserving the extrapolation from realized market outcomes.

²¹Note however that the term “projection” implicitly assumes a direction : that the agent (wrongly) projects his own preferences on another other agent. This makes sense in the social learning setting with asymmetric information structure where each agent knows her own preferences but not others’ preferences. In our setting, the information structure is perfectly symmetric so that an agent has access to exactly the same information (realized utilities) on his own preferences and on others’ preferences. It therefore makes less sense to constrain the direction of the projection: we model an agent who thinks than himself m and another male $\mu(f)$ have similar preferences for whatever reason (projection, reverse projection). The word “identification bias” would probably be more general in this respect.

²²The valuation is defined as an interval of possible (point) beliefs, rather than as a point belief. This is for mostly for consistency (if an agent identifies some agents belonging to the same side of the market together, it makes sense to assume that he does this identification on both sides of the market); it has no consequence in the analysis (using a point belief version of the heuristic with $v_{mf}^\mu = u_{\mu(f)f}$ would lead to the same set of stable matchings).

Corollary 1.

$$\mu \text{ is } \mathbf{v}\text{-stable} \text{ if: } \forall v \text{ valuation: } (NBP)_\mu^v : \forall (m, f): \begin{cases} v_{m\mu(m)}^\mu \geq v_{mf}^\mu \\ \text{or} \\ v_{f\mu(f)}^\mu \geq v_{fm}^\mu \end{cases}$$

The proof consists in transforming the conditions on the v into conditions on the true utilities. We can write the valuation as a convex combination of $u_{m\mu(m)}$ and $u_{\mu(f)f}$. The condition for m being willing to block with f simplifies as a comparison between those two utilities. In total, conditions for stability depend only on the true utilities.

FIGURE 3 below illustrates the way to check the (NBP) condition on an example matching.



FIGURE 2.3: Example: Checking $(NBP)_\mu^v$

We first consider the potential blocking pair $\{m_1, f_2\}$. m_1 currently gets 6 but thinks he would get less by rematching with f_2 , so m_1 is unwilling to block with f_2 and $\{m_1, f_2\}$ is not a blocking pair. We then consider the pair $\{m_3, f_1\}$. Both agents think they would strictly increase their payoffs by rematching together, so they block the matching. μ is v -unstable.

The usual interpretations of stability remain with a slight twist. At a v -stable matching, there is *no feeling of justified envy* (a desirable fairness property). A v -stable matching is an absorbing state of the market, hence a credible interpretation as a long-run outcome.

In terms of notation, Σ^v will be the set of v -stable matchings. Σ^u will be the set of “ u -stable” matchings, so matching that are stable with perfect information.

4 Results: V -stable matchings (pure)

We characterize v -stable matchings, first in the general case, then for specific market structures. We use both static and dynamic approaches to stability.

4.1 Characterization: Happiness sorting

To state a concise characterization of v -stable matchings, we need the following vocabulary.

“Jealousy” is a local condition comparing two realized utilities within a given side of the market.

Definition 3.

$m \in M$ is **jealous** of $m' \neq m \in M$ in $\mu \in \Sigma$ if: $u_{m\mu(m)} < u_{m'\mu(m')}$.

Male m is jealous of another male m' when m' is currently happier than m given their current matches.

“Happiness sorting” is a global condition constraining the whole matching.

Definition 4.

Set $M = \{m_1, \dots, m_n\}$, $F = \{f_1, \dots, f_n\}$.

Set $\mu \in \Sigma$, s.t. $\forall i \in \{1, \dots, n\}$, $\mu(m_i) = f_i$, and $u_{m_1 f_1} \geq \dots \geq u_{m_n f_n}$.

μ satisfies **happiness sorting** if: $u_{f_1 m_1} \geq \dots \geq u_{f_n m_n}$.

A matching satisfies happiness sorting if when we order within each side of the market all agents according to their realized utilities, any two partners hold exactly the same rank. The happiest of all

males is currently matched with the happiest of all females, the second happiest of all males is matched with the second happiest of all females and so forth. Happiness sorting can be interpreted an ordinal version²³ of the famous maxim "you cannot be happier than your wife".

FIGURE 4 below illustrates happiness sorting.

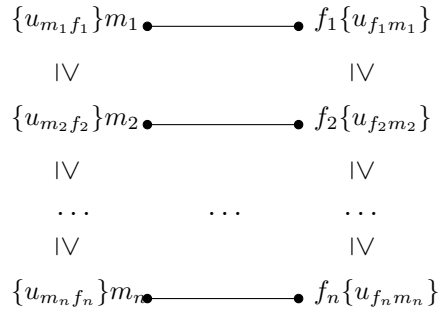


FIGURE 2.4: Happiness sorting

The next proposition fully characterizes v-stable matchings.

Proposition 1.

Set $\mu \in \Sigma$.

μ is v-stable $\iff \mu$ satisfies **happiness sorting**.

A necessary and sufficient condition for the absence of blocking pair is happiness sorting. In the proof, we rewrite the condition for an agent being willing to block with a target as a function of true match utilities u . We observe that m blocks with f if and only if he is jealous of f 's partner. Thus, for v-stability, we need to avoid any pair of jealous agents. We first prove that this is equivalent to happiness sorting on a small-dimensional market of size 2. We then generalize to a market of any size n by observing that imposing happiness sorting for any subgroup of size 2 is equivalent to imposing happiness sorting on the whole market as in definition 4..

A short intuition for this result could say: blocking decisions are based on jealousy. But there are always some jealous agents. There is no way one can eliminate jealousy from the market. The way to ensure stability is rather to organize jealousy in a hierarchical way so that jealousies never coincide to trigger a blocking pair. And this is exactly what happiness sorting does.²⁴

These results, underlying the role of jealousy and ordinal happiness in stability seem to match stylized facts about our real-life markets.

Jealousy, which is an irrational motive from a purely economic point of view, has attracted much attention in psychology. In a book "Jealousy: Theory, Research, and Clinical Strategies" (White (1989) [67]), Gregory White argues that jealousy has been a recurring theme in human relationships, and a major motive for action (in the case of romantic jealousy, it may cause murder, spousal violence, and marital breakdown).

In the economic empirical literature Guven, Senik and Stichnoth (2012) [29] try to account for divorces, using French survey data on marriage, divorce and self-reported happiness. They find a large correlation between divorces and happiness gaps between spouses. In particular, many divorces are initiated by women being less happy than their husband. The prediction from the valuation model is consistent with this finding, in an ordinal way: agents compare utility ranks rather than utility levels with their spouses.

²³Happiness sorting is silent on comparisons of utilities across sides: it could be that all men (women) are quite happy (unhappy) in marriage and still the matching could satisfy the condition.

²⁴In a matching complying with the happiness sorting condition, an agent m_i (f_i) is jealous of agents m_j (f_j) and wants to block with f_j (m_j) if and only if $j < i$. In any potential blocking pair, exactly one agent is willing to block, hence no block.

4.2 Structures on preferences and existence

Proposition 1. is a general characterization of v -stable matchings. An interesting application consists in explicitly computing the set of v -stable matchings after imposing some structure on the market, in particular on the preferences.

Alignment

The first result is about the role of the alignment in preferences. In matching, we say that preferences are aligned whenever when m likes f , then f also likes m . The simplest way to model alignment is to write (without loss of generality) match utilities with two components that depend only on the identity of the pair:

$$\begin{cases} u_{mf} = a_{\{m,f\}} s_{\{m,f\}} \\ u_{fm} = (1 - a_{\{m,f\}}) s_{\{m,f\}} \end{cases}$$

$s_{\{m,f\}}$ is the surplus generated by the pair $\{m, f\}$. $a_{\{m,f\}}$ is the sharing rule - a share $a_{\{m,f\}}$ of the surplus goes to the male m , a share $1 - a_{\{m,f\}}$ to the female f .

We then look at two polar cases:

Proposition 2.

1. **Full alignment:** $a_{\{m,f\}} = a$

$$\Sigma^v = \Sigma$$

2. **Full misalignment:** $s_{\{m,f\}} = s$

$$\Sigma^v = \emptyset$$

In a first extreme case (1.), we fix the sharing rule to a constant so that the heterogeneity in the utilities comes from the surplus. This defines fully aligned preferences.²⁵ Then, any full matching is v -stable.²⁶

The next figure displays an example on a market of size 2, where the sharing rule is always $a = \frac{2}{3}$. On this small market, there are only two possible full matchings: the one with parallel lines and the one with crossed lines, that we both display on the same figure, jointly with the resulting match utilities. Match utilities from a given pair are displayed on the line connecting the pair, each utility closer to the agent receiving it.

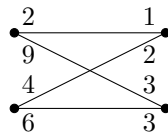


FIGURE 2.5: Numerical example with fully aligned preferences

One can check that for any of the two possible matchings, happiness sorting is indeed verified. The intuition is that in any matching, no agent belonging to the pair generating the highest surplus is willing to block. Conditional on that, agents in the second-highest-surplus pair will not block either and one can continue the induction until the pair currently generating the second-lowest surplus.

²⁵This is, for instance, the trick used in Arnosti (2016) [4] with $a = \frac{1}{2}$.

²⁶This result would still hold with a more general version of aligned preferences:

“Top-top pair” preferences as in Lee and Yariv (2014) [42] $\begin{cases} u_{mf} = \phi(s_{\{m,f\}}) \\ u_{fm} = \psi(s_{\{m,f\}}) \end{cases}$ ϕ, ψ increasing.

Whatever the subset of the market that one considers, one can find a pair such that the agents are each other’s most preferred partner.

At the other extreme (2.), we fix the surplus so that the heterogeneity comes from the sharing. Then, no full matching is v -stable.

The next figure is a numerical example, where the surplus is fixed to 7.

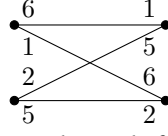


FIGURE 2.6: Numerical example with fully misaligned preferences

No matching satisfies happiness sorting. The intuition in the example is the following. m_1 receives a high realized utility so attracts jealousy from m_2 . But the reason why m_1 is very happy is because he extracts much surplus from his partner f_1 who gets a low utility, hence jealousy from f_1 on f_2 . The agents getting a bad deal block together.

Correlation

In the real world, preferences are often correlated across different agents on the market.²⁷ We study what the v -stability theory can predict in this case.²⁸

The next proposition defines homogeneous (heterogeneous) preferences - when there is an exogenous quality of each partner (agent) in a quite strong sense and characterizes the v -stable set.

Proposition 3.

1. Homogeneous preferences

$$\text{Suppose: } \forall i < j \in \{1, \dots, n\}, \begin{cases} \forall m, m' : u_{mf_i} > u_{m'f_j} \\ \forall f, f' : u_{fm_i} > u_{f'm_j} \end{cases}$$

Then:

$$\boxed{\Sigma^v = \{\mu^\oplus\}}, \forall i \in \{1, \dots, n\} : \mu^\oplus(m_i) = \mu^\oplus(f_i)$$

2. Heterogeneous preferences

$$\text{Suppose: } \forall i < j \in \{1, \dots, n\}, \begin{cases} \forall f, f' : u_{m_i f} > u_{m_j f'} \\ \forall m, m' : u_{f_i m} > u_{f_j m'} \end{cases}$$

Then:

$$\boxed{\Sigma^v = \{\mu^\oplus\}}, \forall i \in \{1, \dots, n\} : \mu^\oplus(m_i) = \mu^\oplus(f_i)$$

In statement (1.), preferences are almost cardinally homogeneous (or common): we can order females so that, for instance, f_1 delivers higher utilities to any males than whatever f_2 would deliver. A consequence is that all males hold the same preferences ordinally (even if two different males may still value the same female differently cardinally). We find that there is a unique v -stable matching, it is the positive assortative matching, that is, the matching where the most-preferred female is matched with the most-preferred male, and so forth. In this case, the prediction with perfect and imperfect information perfectly coincide ($\Sigma^v = \Sigma^u = \{\mu^\oplus\}$).²⁹

The intuition for this case is that the valuation beliefs perform very well in a context of homogeneous preferences. The taste projection is correct. It enables agents to recover their true preferences at least ordinally,³⁰ and to make the same blocking decisions they would make had they access to the true match utilities.

In statement (2.), we introduce correlation by agents rather than by target. Then preferences are heterogeneous; some agents are more cheerful than others and always receive higher utility. The

²⁷This correlation is often explicitly modeled in the matching literature as it is what creates competition on the market.

²⁸Note, however, that the assumption of common knowledge of preferences is easier to defend in this case.

²⁹Becker (1973) [9] has shown that with (ordinally) homogeneous preferences, the positive assortative matching is the unique stable matching in perfect information.

³⁰The valuation interval is not necessarily a singleton.

unique stable matchings sorts agents according to their degree of cheerfulness. the intuition is that at any other matching, the less cheerful agents wrongly identify to the more cheerful ones and block the matching.

This result contrasts with perfect information where we have no precise prediction³¹ for the u-stable set.³²

4.3 Existence of v-stable matchings

No existence

The case with misaligned preferences provides a sound example where the set of full v-stable matchings is empty:

Corollary 2.

A v-stable matching may not exist ($\Sigma^v = \emptyset$).

This is a major contrast with the environment with perfect information, where Gale and Shapley (1962) [25] have shown that a u-stable matching always exists.³³ This is a direct consequence of the fact that beliefs are endogenous to the current matching. Conversely, the case with aligned preferences illustrates that potentially all possible matchings can be stable.

In summary, the v-stable set is significantly more flexible / changeable than the u-stable set.³⁴ The degree of alignment is what governs the size of the v-stable set, from empty to maximal.

Frequency of no existence

We run Monte Carlo simulations to assess how frequent this issue of no existence is. We draw a set of idiosyncratic utilities, for instance $u_{m,f}, u_{f,m}$ i.i.d. $\sim \mathcal{U}([0,1])$. The computer checks for any possible matching whether it is stable and eventually reports whether a stable matching exists. We iterate multiple times, and plot the proportion of preference realizations for which a full matching is stable.

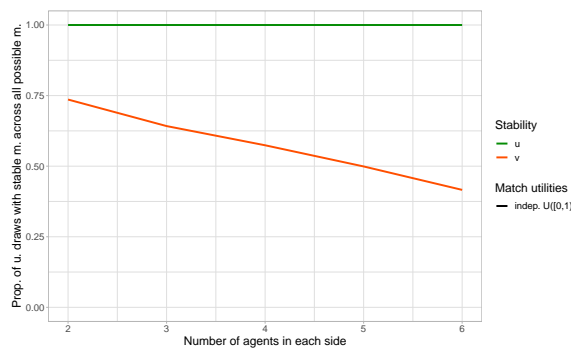


FIGURE 2.7: Proportion of stable matching existence cases with idiosyncratic preferences

We notice that the proportion of inexistence cases is quite large and increasing with market size.³⁵

³¹The issue is that heterogeneous preferences describes preferences across different agents but does not say anything about each individual's preference order.

³²A fortunate consequence is that one could build an additional empirical test of the valuation theory in the environment of heterogeneous preferences.

³³And in the positive matching value specification, it is necessarily full.

³⁴An intuition for this flexibility could be following. With perfect information, the fact that a pair $\{m, f\}$ blocks a given matching μ , making μ unstable implies that in any matching μ' where m and f are matched, they both prefer to stay together than to go with their partners in μ . Said differently, the fact that μ is unstable helps μ' being stable. With v-stability, as with any belief heuristic based on current match utilities only, one loses the dependence: to study the stability of μ and μ' , one runs comparisons among disjoint sets of utilities (except if μ and μ' share some common pairs): $(u_{m\mu(m)})_{m \in M}, (u_{f\mu(f)})_{f \in F}$ for μ , $(u_{m\mu'(m)})_{m \in M}, (u_{f\mu'(f)})_{f \in F}$ for μ' . The fact that μ is v-unstable does not predict anything about the v-stability of μ' .

³⁵With uniform distribution, the probability that a given (full) matching is stable on a market of size n is just the probability that n i.i.d. uniform random variables are ranked in a specific order, so $\frac{1}{n!}$. This converges fast to 0 as $n \rightarrow \infty$. By contrast,

4.4 Blocking pair dynamic and convergence

Closely related to the issue of existence (in the static framework) is the issue of convergence (in a dynamic framework). The question is: if we let agents on the market rematch in a decentralized way, what happens on the long run?

Said differently, after addressing the “what” question (“what are the v-stable matchings?”), we want to address the “how” question (“how does the market reach v-stable matchings?”).

The process

Formally, we define a dynamic blocking pair process as follows:

Definition 5.

A dynamic blocking pair process (DBPP) is a dynamic process $(\mu_t)_{t \in \mathbb{N}}$ where:

- A state of the process is a (full) matching $\mu \in \Sigma$.
- The process moves from a state μ_t to the next state μ_{t+1} in the following way:
 - If there is no blocking pair in μ_t : $\mu_{t+1} = \mu_t$.
 - If there is a single blocking pair $\{m, f\}$ in μ_t :

$$\begin{cases} \mu_{t+1}(m) = f, \mu_{t+1}(f) = m \\ \mu_{t+1}(\mu_t(m)) = \mu_t(f), \mu_{t+1}(\mu_t(f)) = \mu_t(m) \\ \forall m' \in M - \{m, \mu(f)\} : \mu_{t+1}(m') = \mu_t(m'), \forall f' \in F \setminus \{f, \mu(m)\} : \mu_{t+1}(f') = \mu_t(f') \end{cases}$$
 - If there are multiple blocking pairs, the process picks one of the blocking pairs $\{m, f\}$ and builds μ_{t+1} as above.

The interpretation is that we start from a matching and potentially move to a subsequent matching. There are two different cases. When there is no blocking pair in the current matching, the process stabilizes on the matching. Whenever there are blocking pairs, the process implements one of the blocking pairs by forming the pair, and immediately reassigns the former partners of the two blockers together.³⁶

We illustrate the move associated with a blocking pair $\{m, f\}$ below:



FIGURE 2.8: DBPP: Move associated to blocking pair $\{m, f\}$

Note that we do not specify a way to select among blocking pairs, in case the matching features multiple blocking pairs, so that there may be various dynamic blocking pair processes, each associated with a given (deterministic or stochastic) selection rule.³⁷ We do not specify either an initial matching μ_0 from which the process could start, because we want to characterize the long run outcome of the process independently of initial conditions.

Convergence definition

We are interested in the asymptotic behavior of the process, in particular in its convergence.³⁸

the same probability in the roommate problem has lower bound $(\frac{4e^3}{\Pi n})^{\frac{1}{2}}$ (Pittel (1993) [50]), which decreases quite slowly with n .

³⁶The fact that we immediately reassign the rejected agents together guarantees that the states are always full matchings.

³⁷An example of a natural stochastic selection rule: uniform distribution on the set of all blocking pairs.

³⁸We recall only informally the meaning of convergence and related notions (cycle, closed cycle, open cycle, basin of attraction) because these are standard definitions. The formal definitions can be found in any maths textbook on dynamic processes, in the economic and biology literature on dynamic systems (Foster and Young (1990) [22], Kandori, Mailath and Rob (1993) [37]) or in the dynamic matching literature with perfect information (Knuth (1976) [40], Roth and VandeVate (1990) [58]).

A dynamic process is said to *converge* if after some periods it remains in the same state forever. Convergence is heavily linked to the existence of cycles of states and their nature. A (blocking pair) *cycle* is a finite sequence of matchings such that each matching in the sequence is obtained from the previous one where we implement a blocking pair. A cycle is *closed* if, at each stage of the cycle, the implementation of any of the (potentially multiple) blocking pairs leads to matchings within the cycle. A cycle is *open* if there exists at least one matching in the cycle and one blocking pair in this matching, such that if one implements this blocking pair, the subsequent matching does not belong to the cycle. The relation between the static notion of existence of a v -stable matching and the dynamic notion of convergence of a DBPP is straightforward. If there does not exist a (full) stable matching, then no DBPP can converge. If there exists stable matchings, and there is no blocking pair cycle, any DBPP converges (to v -stable matchings). If there exists stable matchings, and any blocking pair cycle is open, then any DBPP equipped with a stochastic selection rule for blocking pairs with full support converges (to stable matchings). The intuition is that thanks to the full support assumption on the selection of blocking pairs, there is a strictly positive probability that the DBPP takes the convergence path that gets out of cycles. Even if this probability is small and potentially after much cycling, in the long run, this path will be taken for sure. Finally, if there exists v -stable matchings, and a closed blocking cycle, some DBPP will not converge.³⁹

The notion of convergence defined is a notion of global convergence a la Roth and Vande Vate (1990) [58], hence demanding. When global convergence fails, it is interesting to characterize local convergence, by studying the size of the *basin of attraction* of a v -stable matching μ , defined as the set of matchings such that when we start from one of these matchings and implement blocking pairs, we reach μ for sure.

The question is: Does a DBPP always converge to stable matchings when they exist? Equivalently: conditional on existence, are stable matchings *attractors* of a DBPP?

No convergence

In a world of perfect information, Knuth (1976) [40] first exhibited an example of a cycle.⁴⁰ Yet later on, Roth and Vande Vate (1990) [58] showed that any cycle is open, hence some convergence.

For v -stability, we find negative results in both respects as soon as market size is larger or equal to 3:

Proposition 4.

Set a marriage market.

(i) If the size of the market is 1 or 2, then any DBPP converges.

(ii) If the size of the market is n , $n \geq 3$, then:

- A DBPP may not converge.
- In addition, the basin of attraction of v -stable matchings may be very small, even empty.

We prove (ii) by exhibiting an example of a closed cycle. To build the example, we consider again the counter example we used for existence on a market of size 2. We then add a third pair with very high utilities for both partners, so that these agents are never willing to enter a blocking pair:

³⁹ Consider for instance a DBPP starting on a matching within the closed cycle.

⁴⁰ In Knuth (1976) and followers, the definition of the dynamic process differs in that former partners rather stay unmatched for one period before they are allowed to enter a blocking pair (with anyone). But in the specific counter example used by Knuth (1976), the process moves only across full or almost-full matchings where the two singles always rematch in the next stage. The counter example thus works with any of the two definitions of the process.

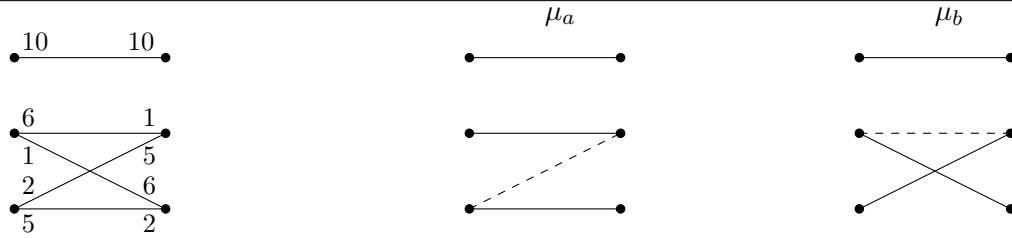


FIGURE 2.9: Example of closed blocking pair cycle $\mu_a \leftrightarrow \mu_b$

The cycle is of minimal length 2. In each of the two matchings, there is a unique blocking pair, so that there is no concern of blocking pair selection. Any DBBP with initial state μ_a or μ_b will cycle forever.

We can define the rest of the match utilities so that a third matching μ complies with happiness sorting and is v-stable. μ and μ_a or μ_b could differ from each other by just one block, or two pairs. Being one block away from the v-stable matching does not even guarantee that we move to it - hence the result on the small-size / empty basin of attraction.

The conclusion, is that, even in case of existence of a v-stable matching, a DBPP may not converge to it, neither globally, nor locally.

Frequency of no convergence

We run Monte Carlo simulations to assess how frequent the issue of no convergence is. We draw a set of idiosyncratic utilities. We randomly (uniformly) pick an initial matching. The computer checks for potential blocking pairs and implement them (with uniform selection rule in case of multiple blocking pairs), for a finite but large number of periods (up to 300).⁴¹ Any time the process stops on a matching for two successive periods, the computer acknowledges a convergence case. We iterate multiple times (on the initial matching and on the set of utilities). We report the proportions of NBPP trials with convergence and the average number of periods before convergence, conditional on convergence.

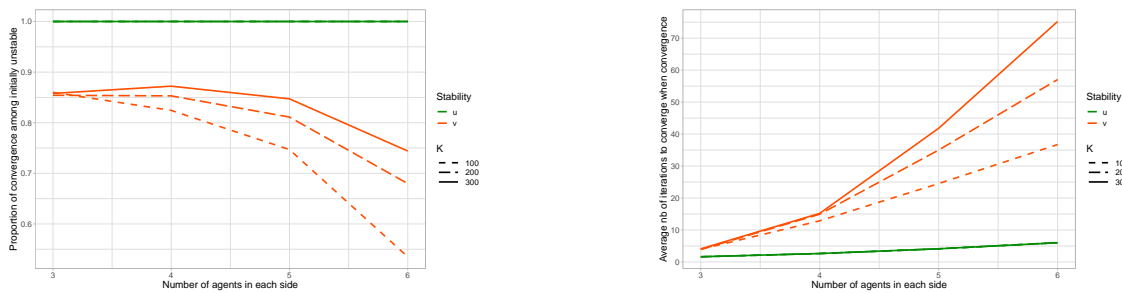


FIGURE 2.10: Proportion of convergence cases and speed of convergence of DBPP with idiosyncratic preferences

The results show evidence that the no convergence issue arises frequently and that even in case of convergence, it is significantly slower than with the true utilities.

In total, the prediction from the baseline model is clear. If indeed agents have valuation beliefs, the market may keep on moving and never stabilize, for two different reasons: lack of existence of a v-stable matching, and, even in case of existence, lack of convergence of a DBPP.

⁴¹When doubling the number of time periods does not change much the number of convergence cases, we consider that the number of time periods is a good approximation of infinite time horizon.

5 Model: Valuation and v-stability (mixed)

5.1 Mixed matchings

The continuum marriage market

We consider a two-sided continuum population, and n discrete types of males ($M \in \mathcal{M}$) and females ($F \in \mathcal{F}$) of equal mass 1.⁴² The type summarizes all relevant characteristics for match utilities: all males of types M matched with females of type F receive the same match utility u_{MF} .

Mixed and behavioral matchings

We define three notions of (full) matchings:

Definition 6.

(i) **Pure matching:**

$$\begin{aligned} \mu &: \mathcal{M} \cup \mathcal{F} \rightarrow \mathcal{M} \cup \mathcal{F} \\ \forall M \in \mathcal{M}, \mu(M) &\in \mathcal{F} \\ \forall F \in \mathcal{F}, \mu(F) &\in \mathcal{M} \\ \mu(M) = F &\iff \mu(F) = M \end{aligned}$$

(ii) **Mixed matching:**

$$\begin{aligned} p &: \Sigma \rightarrow [0, 1] \\ \forall \mu \in \Sigma &: p(\mu) > 0 \\ \sum_{\mu \in \Sigma} p(\mu) &= 1 \end{aligned}$$

(iii) **Behavioral matching:**

$$\begin{aligned} p &: \mathcal{M} \times \mathcal{F} \rightarrow [0, 1] \\ \forall M \in \mathcal{M}, \forall F \in \mathcal{F} &: p(M, F) > 0 \\ \forall M \in \mathcal{M}, \sum_{F \in \mathcal{F}} p(M, F) &= 1 \\ \forall F \in \mathcal{F}, \sum_{M \in \mathcal{M}} p(M, F) &= 1 \end{aligned}$$

A pure⁴³ matching is a standard matching as in section §3, where each type is matched exclusively to one partner type. A mixed⁴³ matching is a full-support probability distribution over pure matchings $p \in \Delta(\Sigma)$, where $p(\mu) > 0$ is the weight on pure matching μ . A behavioral⁴³ matching is a full-support measure over pairs of types, where $p(M, F) > 0$ describes both the mass of males of type M matched with females of type F and the mass of females of type F matched with males of type M . We focus on full matchings where for each pair of a male and a female type, there is at least one pair of individuals belonging to these types who are currently matched.⁴⁴

The question immediately arises: are the two representations (mixed and behavioral) equivalent? The next proposition answers positively.

Proposition 5.

A mixed matching can always be represented as a behavioral matching and conversely.

From now on, we denote $P := \Delta(\Sigma)$, the set of mixed or behavioral matchings. Once a mixed matching is set, different agents belonging to the same type have different match statuses. From now on, we call “ex-post” type and denote $(M - F)$, the pair of an agent’s type and his current partner’s type. If there are n ex-ante types on each side of the market, and the current matching is full, there are n^2 ex-post types on each side of the market.

⁴²We still denote individuals with small cases $m \in \mathcal{M}, f \in \mathcal{F}$. $M(m)$ ($F(f)$) denotes the type of individual m (f).

⁴³The analogy in the vocabulary with strategies in games is made on purpose and will later help in the analysis.

⁴⁴In section §7.1, we extend the model study v-stability of any mixed or behavioral matchings.

We illustrate below the notion of mixed / behavioral matching in its mixed version⁴⁵ on a market with two ex-ante types ($2^2=4$ ex-post types) on each side. Because there are only two full pure matchings in this environment, a mixed matching is summarized by just one number $p \in [0, 1]$, the weight on the pure matching where $\{M_1, F_1\}, \{M_2, F_2\}$ are matched (in green).

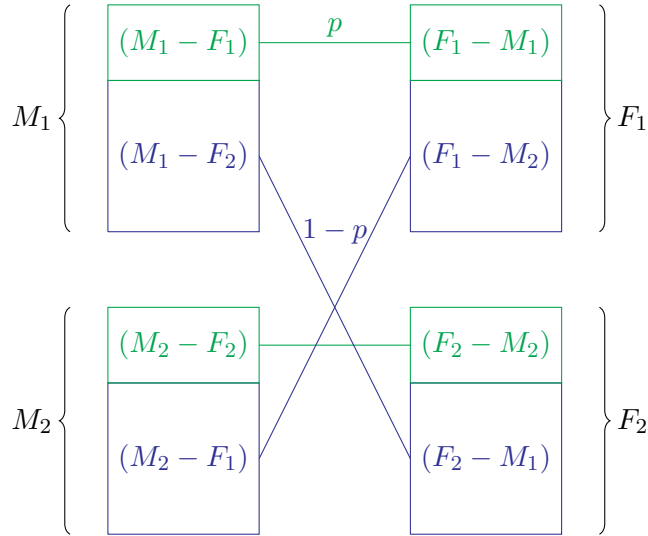


FIGURE 2.11: Mixed matching p on a market of size 2

The set of mixed matching P is a subset of \mathbb{R}^n . We compute its dimension.

Lemma 1.

The set of mixed or behavioral matchings P on a market of size n has dimension $1 + n - 2n$.

Mixed matchings in the literature

Our notion of mixture over matchings is not entirely new. Vande Vate (1989) [66], Rothblum (1992) [59], and Roth, Rothblum and Vande Vate (1993) [57] define a notion of “fractional” matching (in our vocabulary: mixed or behavioral matching).

Yet, their motivation is mainly technical: they want to unify the assignment model introduced by Shapley and Shubik (1971) [62] and the marriage model by Gale and Shapley (1962) [25]. With respect to efficiency, they show that “monogamous marriage is the best”, that is, socially optimal matchings are necessarily “integral matchings” (in our vocabulary: pure matchings).

The interpretations also differ. Their favorite interpretation is that one type is one individual sharing her time, as in polygamy or on job markets where workers can hold several jobs. Our interpretation is that the population is made of multiple agents with the same productive characteristics and preferences bundled in types.

They do define stability of a fractional matching (strong stability, ex-post stability, fractional stability), but all the various notions implicitly assume perfect information on the utilities in any integral matching. By contrast, we are interested in belief formation under imperfect information. Specifically, we argue that valuation beliefs make a lot of sense in a mixed environment.

The information

The information available on the market is such that agents do not have fine access to counterfactual match utilities, even the ones that involve their own types. If male m of type M is currently matched to a female f of type F , he can observe u_{MF} , but is ignorant of any $u_{MF'}$, $F' \neq F$.

Agents perfectly observe the current mixed matching p . This is not very demanding as p is an aggregate

⁴⁵The associated behavioral matching would be: $p(M_1, F_1) = p(M_2, F_2) = p$, $p(M_1, F_2) = p(M_2, F_1) = 1 - p$.

specifying the matching in proportions of types resulting from the (potentially unknown) individual matches.

5.2 The valuation

Valuation in mixed matching

We assume that agents form beliefs on counterfactual utilities in the following way.

Definition 7.

The valuation is a function mapping a male m , a female type F , and a mixed matching $p \in P$ to a belief over the utility u_{mF} that a male m of type M currently assigned to a female of type F' would get with any targeted partner f of type F when the current matching is p :

$$v : M \times \mathcal{F} \times P \rightarrow \mathbb{R}$$

$$v_{mF}^p \in (u_{mF'}, \sum_{M_0 \in \mathcal{M}} p(M_0, F) u_{M_0 F'}) \text{ if } F \neq F' \quad (1)$$

$$v_{mF}^p = u_{mF'} \quad \text{if } F = F' \quad (2)$$

On line (1), the valuation of agent m about type F extrapolates from two signals: the utility that he gets in his current match and a weighted average of the realized utilities of all types of males currently matched with females of type F . The weights on utilities are given by the true proportions in the current mixed matching p , so that if F is mostly matched with males of type M , u_{mF} heavily influences the belief (bound). In other words, to assess how productive a potential partner is, an agent observes how happy the target's average current partner is. Line (2) says that the agent perfectly observes his current utility. In total, the valuation depends on the exact individual m only through his match status. The emotion associated with this belief formation process is envy, more than the jealousy from section §3.3. Jealousy is directed towards one fellow male, whereas envy refers to longing for a situation that is not currently one's own.

We find this belief formation process very natural and consistent with empirical examples where people seem to aggregate data in a similar fashion.

Interpretations

There are several ways we can interpret the valuation.

The first story is about a cognitive limitation, which we can term in two different ways. Either the agent is naive; she cannot understand the link between partners' types and utilities. Or the agent is coarse; she cannot finely distinguish between utilities arising from the same targeted type but different agents' types.

A second story, in the spirit of rational inattention, would argue that if there is a positive search cost to collect each u_{mF} , one might not want to pay this cost.

A third interpretation (and our favorite one) is about a constraint on the feedback data. Data could be available only in the form of the aggregate statistics $\sum_{M \in \mathcal{M}} p(M, F) u_{mF}$ ⁴⁶ and not with whole sample of $(u_{mF})_{M \in \mathcal{M}}$. This is often the case on large and centralized markets. In college admissions, students applying to colleges on an online platform often receive feedback on former students' satisfactions or achievements at each college, as measured by employment and wages after college. This data can be provided either by the platform or available in specialized press. It is almost always presented in aggregate form, with averages over the whole population of former students from each college. The fact that students are of different types and that these students' types partially account for the outcomes is neglected.

⁴⁶One example for the marriage market could be that potential partners of type F brag about the aggregate satisfaction they deliver to their current partner $\sum_{M \in \mathcal{M}} p(M, F) u_{mF}$. Bragging is often made in a coarse way, focusing on one aggregate statistic of attractiveness that should speak to anyone on the other side of the market.

Connection to literature on bounded rationality with aggregation

The agent in the valuation model bears similarities with some of the cornerstone models of bounded cognition. In particular, a valuation agent can be seen as a naive agent as in the behavioral equilibrium by Esponda (2008) [20] or as a coarse agent as in the analogy-based expectation equilibrium by Jehiel (2005) [35].

The behavioral equilibrium is a prior-free approach: in the original example, the buyer forms his belief about the quality of the good based on observed data on payoff in the case of transaction. She is naive in the sense that she neglects the link between the quality of the good and the seller's action. She therefore sees the observed qualities when there is a transaction as representative of the true quality distribution. The valuation agent's naivete is in her failure to appreciate the link between a utility generated by a target and the target's current partners' types.

In the analogy-based expectation equilibrium, coarse players form beliefs on the strategies of other players by computing an average of the true equilibrium strategies, weighted by the true probabilities of each state. A valuation agent is coarse because she is unable to finely distinguish between the several types of partners of a target type.⁴⁷

The valuation beliefs could also be interpreted as resulting from a wrong causal model (Spiegler (2020) [64]), which we represent on a directed acyclic graph:

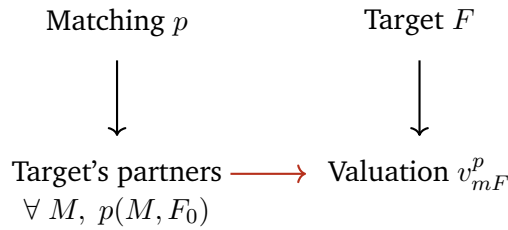


FIGURE 2.12: True causal model for valuation of male m about female F in matching p . The red arrow is omitted in the wrong causal model.

5.3 V-stability

Blocking pairs

The definition of a blocking pair is slightly more subtle in the mixed matching environment. The question is not only: with whom should one type trade his current partners? It is also: which ones exactly of the current partners does the type want to trade? A blocking pair is therefore defined both with respect to the blocking type and to the dropped partners' types (equivalently, with respect to ex-post types):

Definition 8.

- (i) A **strict blocking pair** $\{(M - F'), (F - M')\}$ in a mixed matching $p \in P$ is a pair of a male ex-post type $(M - F')$ and a female ex-post type $(F - M')$ s.t.:

$$\begin{cases} v_{mF}^p > u_{MF'}, m \in (M - F') \\ v_{fM}^p > u_{FM'}, f \in (F - M') \end{cases}$$

- (ii) A **weak blocking pair** $\{(M - F'), (F - M')\}$ in a mixed matching $p \in P$ is a pair of a male ex-post type $(M - F')$ and a female ex-post type $(F - M')$ s.t.:

$$\begin{cases} v_{mF}^p \geq u_{MF'}, m \in (M - F') \\ v_{fM}^p \geq u_{FM'}, f \in (F - M') \end{cases}, \text{ with at least one equality.}$$

⁴⁷Note that others heuristics could also be in the spirit of analogy. For instance, a very coarse heuristic would say that male agents only have access to the average happiness of all matched males agents on the market. Abusing vocabulary, the valuation corresponds to a reasonably fine analogy partition of the set of possible pairs of types where all pairs where the female type is fixed are in the same class. The average happiness heuristic corresponds to the coarsest analogy partition.

The u on the right side of the inequalities reflects that an agent of type M currently matched with a female of type F perfectly observes u_{MF} .

In a weak blocking pair, at least one agent is indifferent between blocking or not.

Dynamic blocking pair process

We define a quadruple as a group of two males and two females types. Quadruples play a major role in the dynamic blocking pair process defined just below as the local scale of the market where we check and implement blocking pairs.

Definition 9.

A dynamic blocking pair process (DBPP) is a dynamic process $(p_t)_{t \in \mathbb{N}}$ where:

- A state of the process is a (full) mixed matching $p \in P$.
- The process moves from a state p_t to the next state p_{t+1} in the following way:
 - It draws at random (with full support distribution) a quadruple $\{M, M', F, F'\}$.
 - If there is no blocking pair in the quadruple $\{M, M', F, F'\}$, the matching remains unchanged: $p_{t+1} = p_t$.
 - If there are some blocking pairs in the quadruple $\{M, M', F, F'\}$:
 - * All strict blocking pairs are implemented until the mass of one ex-post type in p_t is exhausted.
 - * Weak blocking pairs may not be implemented, fully implemented, or partially implemented (implemented for only part of the mass of the indifferent ex-post type(s)).

This process implements blocking pairs focusing on one quadruple per period. Importantly, indifferent agents are given full flexibility.⁴⁸

We illustrate one step of the process below. FIGURE 13 is a focus on the quadruple $\{M, M', F, F'\}$ within a market of size larger than 2. We represent uneven masses of ex-post types on purpose. The rest of the market, which does not appear on the figure, compensates so that the mass of any ex-ante type is 1.

The figure corresponds to a case with one strict blocking pair $\{M, F\}$ in quadruple $\{M, M', F, F'\}$ in p_t . The blocking ex-post type ($F - M'$) has the lowest mass and forces the block to stop after exactly this mass of each ex-post type has been reassigned.

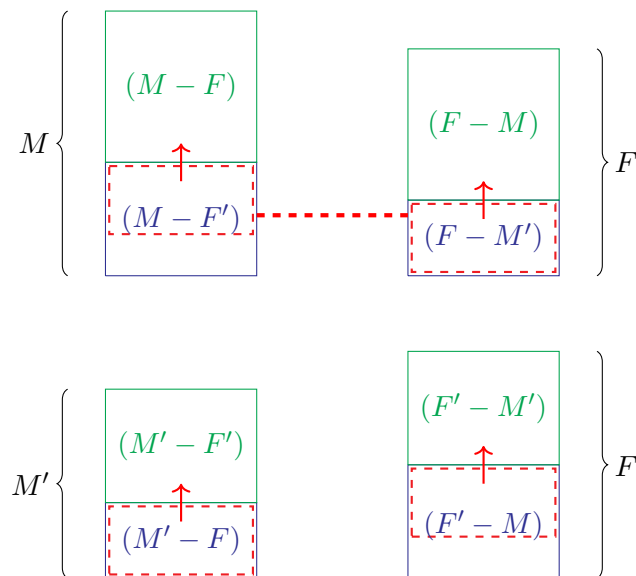


FIGURE 2.13: A DBPP step where the implemented blocking pair is $\{(M - F'), (M' - F)\}$.

⁴⁸ Similarly to the indifference principle in Nash equilibrium, where players may use any mixture when they are indifferent over actions.

Steady state

The solution concept is a notion of steady state of the dynamic process. There exists a steady state of the DBPP if, after some time, the matching remains unchanged forever.

Definition 10.

$p \in P$ is v -stable if it is a steady state of the DBPP.

The notion of steady state is quite permissive, in the sense that there exist various ways to achieve it in the DBPP.

The first, most standard way would be to ensure that there is no blocking pair at all (“stability without flows”).

But there could also be blocking pairs in a way that ensures that when the blocking pairs are implemented, the measure p is unchanged (“stability without flows”). At the micro, individual level, agents may switch partners. But at the macro type level, masses of ex-post types remain constant.

When there are flows and blocking pairs are only strict, there is little flexibility on the matching because the agents willing to block and trade partners must have exactly the same mass (“stability with flows and only strict blocking pairs”). For instance, on a market of size 2, this constrains $p = \frac{1}{2}$. When a blocking pair is rather weak, it gives the possibility that the arbitrary decisions of indifferent agents about blocking exactly balance the difference in masses (“stability with flows and a weak blocking pair”).

6 Results: v -stable matchings (mixed)

6.1 General result: Existence

Existence

The shift to the mixed matching environment restores existence.

Theorem 1.

There always exists a v -stable matching in P .

Note that the existence would work with any belief continuous in the measure p . It does not crucially hinge on the functional form of the valuation. The average just stands as the most natural way of aggregating signals.⁴⁹

Similarly, the existence result and the set of steady states is robust to the distribution with which the process draws quadruples (as long as this distribution is full support). The most natural distribution would draw more often the quadruples that are more frequent in a uniform way.

Lemma 2.

The probability distribution drawing quadruples uniformly according to their frequencies draws each quadruple $\{M, M', F, F'\}$ with probability:

$$X_{\{M, M', F, F'\}} = \frac{p(M, F) + p(M', F') + p(M, F') + p(M', F)}{n(n-1)^2}$$

Sketched proof

The proof is a standard fixed-point proof. It formalizes the “transformation function” as a set-valued function F , mapping a former mixed matching p_t to new possible matchings p_{t+1} . The transformation function is built through many auxiliary functions at the local level (for one quadruple $\{M, M', F, F'\}$). We show that the graph of the transformation function is closed, implying by the closed graph theorem that the function is upper hemi-continuous. We then apply the fixed point theorem by Kakutani (1961) [36], to prove that there exists a fixed point or steady state of the DBPP.

The fact that we give full flexibility to indifferent agents is key in the proof; it is what closes the graph of the transformation function.

⁴⁹In this respect, we also expect the very coarse average happiness heuristic to produce v -stable matching.

To illustrate this point, we define and plot the transformation function on a market of size 2, where a matching is summarized by one number $p \in \mathbb{R}$ so that the graph of the function can be plotted in two dimensions.

- $g : [0, 1] \rightrightarrows \{0, 1, 2\}^2$ counts possible blocking pairs in any matching p . An image $g(p)$ has two coordinates: each one reports the number of blocking pairs in each of the 2 pure matchings.
- $k^p : \{0, 1, 2\}^2 \rightarrow [0, 1]$ starting from a given matching p , maps a number of effective blocking pairs into a new matching. The transformation amounts to implementing the blocking pairs.
- $h^p : \mathcal{P}(\{0, 1, 2\}^2) \rightrightarrows [0, 1]$ starting from a given matching p , maps a set of numbers of possible blocking pairs into a set of possible new matchings, consistently with k^p . With this function, we can accommodate cases with weak blocking pairs and treat all the cases depending on how indifferences resolve.
- $f : P \rightrightarrows P$ with $f(p) = h^p(g(p))$ is the transformation function. It maps a matching to a set of possible new matchings.

Images of k^p can only be 0, 1, p , $1 - p$. Images of h^p can only be $\{0\}$, $\{1\}$, $\{p\}$, $\{1 - p\}$, $[0, p]$, $[p, 1]$, $[0, 1 - p]$, $[1 - p, 1]$.

Thus, if $] \alpha, \beta [$ is an open interval where there is no indifference point, the graph of f can only be continuous and linear in p :

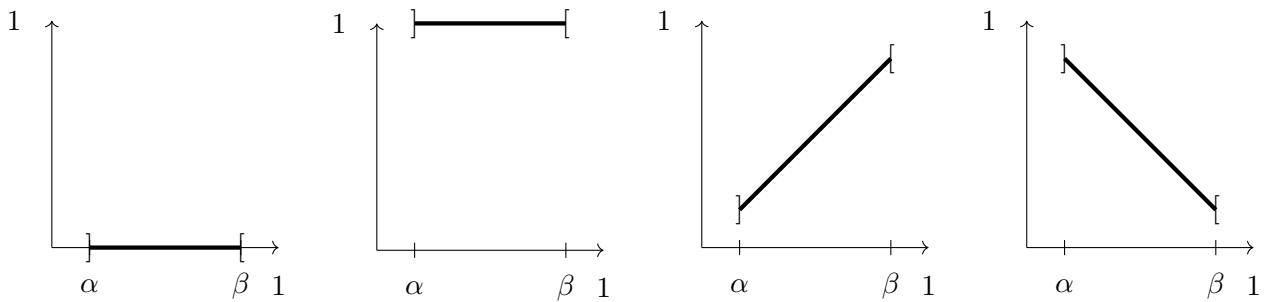


FIGURE 2.14: Transformation function f on domains with no indifference

Now consider a mixed matching p^* , where $\{(M - F), (F - M)\}$ is a strict blocking pair, and $\{(M - F'), (F - M')\}$ is a weak blocking pair because $(M - F')$ is indifferent to block with F at p^* . In the neighborhood of p^* , $(M - F')$ is not indifferent. Without loss of generality, we set that for $p > p^*$ ($p < p^*$), $(M - F')$ is willing (not willing) to block with F . The graph of f in the neighbourhood of p^* is then:

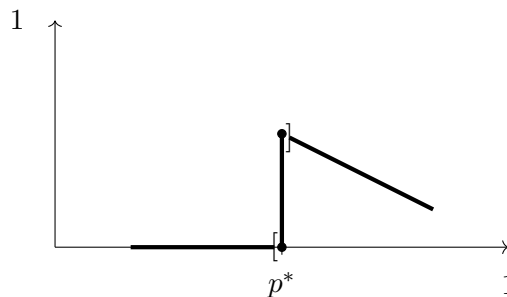


FIGURE 2.15: Transformation function at indifference point

The flexibility we give to indifferent agents (captured in the definition of h^p) is what connects the two linear curves with the vertical line.

We get that the graph of f , $Gr(f) = \{(p, f(p)), p \in P\}$, is closed. Consequently, we can apply the closed graph theorem and Kakutani's fixed point theorem.

6.2 Misaligned preferences: Characterization

When preferences are misaligned, matched agents have opposite interests to capture as much as they can of a fixed surplus:

$$\begin{cases} u_{MF} = a_{\{M,F\}}s \\ u_{FM} = (1 - a_{\{M,F\}})s \end{cases}$$

In section §4.2, we have shown that misalignment is a case where no pure matching is v-stable. In theorem 1., though, it appears that the mixed extension solves the no existence issue. It is therefore interesting⁵⁰ to characterize the v-stable matchings in this case, providing a constructive proof of existence. For readability, we focus on a market of size 2, with types M_1, M_2, F_1, F_2 . We simplify the notation $a_{\{M_i, F_j\}}$ to a_{ij} . Our problem has four parameters (the sharing rules): $a_{11}, a_{12}, a_{21}, a_{22} \in [0, 1]$.⁵¹ The surplus s does not matter for stability and can be normalized to $s = 1$.

Proposition 6.

When preferences are misaligned on a market of size 2, v-stable matchings are the following p :

(1) $a_{12} < a_{21} < a_{22} < a_{11}$

– $p = p_0 := \frac{a_{21} - a_{12}}{a_{11} - a_{12}}$ (with flows, weak BP)

– If $\frac{1}{2} < p_0$: $p = \frac{1}{2}$ (with flows, strict BP)

(2) $a_{12} < a_{22} < a_{21} < a_{11}$

(a) $(a_{22} - a_{21})(a_{11} - a_{12}) < (a_{21} - a_{12})(a_{11} - a_{12})$

– $p = p_1 := \frac{a_{22} - a_{21}}{a_{11} - a_{21}}$ and $p = p_2 := \frac{a_{21} - a_{12}}{a_{11} - a_{12}}$ (with flows, weak BP)

– If $p_1 < \frac{1}{2} < p_2$: $p = \frac{1}{2}$ (with flows, strict BP)

(b) $(a_{21} - a_{12})(a_{11} - a_{12}) < (a_{22} - a_{21})(a_{11} - a_{12})$

– $p_2 < p < p_1$ (without flows)

(3) $a_{22} < a_{21} < a_{12} < a_{11}$

– $p = p_3 := \frac{a_{12} - a_{21}}{a_{12} - a_{22}}$ (with flows, weak BP)

– If $p_3 < \frac{1}{2}$: $p = \frac{1}{2}$ (with flows, strict BP)

(4) $a_{22} < a_{12} < a_{21} < a_{11}$

– $p = p_2$ (with flows, weak BP)

– If $\frac{1}{2} < p_2$: $p = \frac{1}{2}$ (with flows, strict BP)

The characterization shows that all three kinds of stability may arise, depending on the parameters. In particular, this proves that the flexibility in the treatment of weak blocking pairs is, in some cases at least, necessary for existence.

7 Discussions and extensions

7.1 Unmatched agents

In section §3 (§5), we have focused on full pure (mixed) matchings. We have consistently defined the valuation $v_{m_f}^\mu$ ($v_{m_f}^p$), positing that both the agent and the target (types) were currently matched in μ (p). In many real situations, though, agents may consider remaining unassigned.

In this section, we extend all definitions to allow for single agents. We define v-stability of a non-full matching. We denote $\bar{\Sigma}(\bar{P})$, the set of all pure (mixed) matchings. We denote $\bar{\mathcal{M}} := \mathcal{M} \cup \emptyset$, $\bar{\mathcal{F}} := \mathcal{F} \cup \emptyset$.

⁵⁰And also less computational than the general case, as the misalignment divides the number of parameters of the problem by two.

⁵¹In the coming proposition, we fix (without loss of generality) $a_{11} = \max_{i,j} a_{ij}$.

Definition 11.(i) **Pure matching:**

$$\begin{aligned} \mu &: \mathcal{M} \cup \mathcal{F} \rightarrow \mathcal{M} \cup \mathcal{F} \cup \emptyset \\ \forall M \in \mathcal{M}, \mu(M) &\in \bar{\mathcal{F}} \\ \forall F \in \mathcal{F}, \mu(F) &\in \bar{\mathcal{M}} \\ \mu(M) = F &\iff \mu(F) = M \end{aligned}$$

(ii) **Mixed matching:**

$$\begin{aligned} p &: \bar{\Sigma} \rightarrow [0, 1] \\ \sum_{\mu \in \Sigma} p(\mu) &= 1 \end{aligned}$$

(iii) **Behavioral matching:**

$$\begin{aligned} p &: \bar{\mathcal{M}} \times \bar{\mathcal{F}} \rightarrow [0, 1] \\ \forall M \in \mathcal{M}, \sum_{F \in \bar{\mathcal{F}}} p(M, F) &\leq 1 \\ \forall F \in \mathcal{F}, \sum_{M \in \bar{\mathcal{M}}} p(M, F) &\leq 1 \end{aligned}$$

We define a notion of v -stability for the most general class of matchings \bar{P} :

Definition 12.

$p \in \bar{P}$ is almost v -stable if $\exists (p_\epsilon)_\epsilon \in P$ s.t.: $\begin{cases} p_\epsilon \xrightarrow{\epsilon \rightarrow 0} p \\ p_\epsilon \text{ is } v\text{-stable} \end{cases}$

We define a small perturbation p_ϵ of the non-full matching that makes it a full matching. In a full matching, feedback data about any targeted type is available. We can therefore study the v -stability of the perturbed matching p_ϵ using standard definitions.

This approach is in the spirit of the “trembling hands” in perfect Nash equilibrium by Selten (1985) [60]. The trembling gives feedback data to agents targeting types that are completely unmatched in p . The data on types that are currently matched with positive mass is little perturbed.

As an example, we may consider a candidate perturbation putting a ϵ weight uniformly on any pair, and $(1 - \epsilon)$ weight on all the pairs as prescribed in p :

$$p_\epsilon = (1 - \epsilon).p \oplus \epsilon.q, \text{ where } \forall M, F, q(M, F) = \frac{1}{n}$$

Then, the valuation of a type m of type M about type F in the matching where m is unassigned and F has zero mass matched is:

$$v_{mf}^p \in (u_{M\emptyset}; \frac{1}{n} \sum_{M_0 \in \mathcal{M}} u_{M_0F}].$$

If m is rather currently assigned to a female of type F' and the targeted type F has positive mass matched, the valuation of f about M is:

$$(u_{MF'}; \sum_{M_0 \in \mathcal{M}} ((1 - \epsilon)p(M_0, F) + \frac{\epsilon}{n})u_{M_0F}]$$

We can now redefine the DBBP in a more natural way by making rejected partners (agents whose former partner has engaged in a blocking pair) unassigned before they potentially form new matches. We would like to show a general existence theorem for almost v -stable matchings. Existence in this setting does not directly stem from our previous existence theorem 1.. In the cases where existence rests on “stability with flows”, the flows now transition through a state where some agents remain unmatched and where potentially new moves could happen.

7.2 Valuation with bundles of goods

A more general model assumes that a match generates a (multi-dimensional) bundle of goods for each partner.⁵² If $m \in M$ and $f \in F$ are currently matched, male m enjoys bundle $x_{mF} \in \mathbb{R}^n$. The mapping from bundles to utility function is described by individual utility functions: $u_M, u_F : \mathbb{R}^n \rightarrow \mathbb{R}$. The index M, F indicates that different types may value the same bundle differently.

Definition 13.

The valuation is a function mapping a male m , a female type F , and a (full) mixed matching $p \in P$ to a belief over the utility u_{mF} that a male m of type M currently assigned to a female of type F' would get with any targeted partner f of type F when the current matching is p :

$$v : M \times \mathcal{F} \times P \rightarrow \mathbb{R}$$

$$v_{mF}^p \in (u_M(x_{mF'}), \sum_{M_0 \in \mathcal{M}} p(M_0, F) u_M(x_{M_0 F})) \text{ if } F \neq F' \quad (1)$$

$$v_{mF}^p = u_M(x_{mF'}) \text{ if } F = F' \quad (2)$$

When forming his belief about counterfactual utility u_{mF} , male m looks at the bundles that he enjoys in his match with his partner of type F' and at the bundles that current partners of the targeted type F enjoys in their matches. He integrates these bundles into his own utility function, and then aggregates in a convex combination.

In this setting, blocking decisions are based on some form of jealousy (in the sense of comparison within peers), but this jealousy is outcome-based (rather than utility-based). It could be that almost all male types M matched with type F_1 are very happy in their matches because in x_{mF_1} they get much leisure time, and they value leisure time a lot according to u_M . But if $m_2 \in M_2$ does not value leisure time that much according to u_{M_2} , it could be that m_2 is unwilling to block with F_1 . At the macro level, jealousy is not directed towards the same agents anymore.

With some assumptions on preferences or on bundles, we should be able to characterize v -stable matchings in this more general environment, as the next proposition tentatively does.

Proposition 7.

- (i) If preferences are cardinaly homogeneous ($u_M = v, u_F = w$), then the v -stable matchings are as in the baseline model.

In particular, a pure matching is v -stable if and only if it satisfies happiness sorting.

- (ii) If preferences are ordinaly homogeneous ($u_M(x) > u_M(x') \Leftrightarrow u_{M'}(x) > u_{M'}(x')$) then: A pure matching is v -stable if and only if the male who gets the current k^{th} most preferred male bundle is matched to the female who gets the current k^{th} most preferred female bundle, $\forall k \in \{1, \dots, n\}$.

Case (i) fixes the same utility function v for all males. It can be connected to the baseline model by setting $u_{mF} := v(x_{mF})$. Case (ii) assumes that male utility functions are homogeneous only ordinaly. We get the same hierarchical structure as in section §4 where on any subset of two matched pairs on the market, one should find one pair of non-jealous agents. However, this does not summarize in a single and easy-to-write macro condition such as happiness sorting due to the cardinal heterogeneity in preferences.

8 Conclusion

This paper defines a new cooperative solution concept for matching markets that incorporates behavioral elements. It models a heuristic belief formation process where agents wrongly project preferences and characteristics between different agents belonging to the same side of the market. It studies pairwise stability in this context, in the standard setting of pure matchings and in the more general environment of mixed matchings.

⁵²Becker [9] gives caricatural examples of typical goods generated in marriage: quality of meals, share of household work...

Overall, the value of the results is two-fold.

First, it accounts for empirical puzzles about real-life markets. The happiness sorting condition can be read as a formalization of the famous maxim in marriage: “You can’t be happier than your wife”, and by contraposition can account for divorces. The analysis with specific preferences shows how misalignment in preferences maintains instabilities, and how assortativity may arise even more frequently with imperfect information than with perfect information.

Second, it gives robust predictions about the long-run dynamic on matching markets. The overall prediction says that the market moves significantly more with imperfect information than with perfect information. The exact prediction differs depending on the granularity of the market, in line with our model in two levels: pure and mixed matchings. When types are divisible / each agent is virtually unique in its productive characteristics, the lack of existence of a v -stable matching and the lack of convergence of a dynamic blocking pair process predict persistent moves. When types are indivisible / there are many agents with the same productive characteristics, the general existence of a mixed v -stable matching predicts that the market will stabilize. The characterization of mixed v -stable matchings under misaligned preferences shows that at the steady states, agents may keep on exchanging partners in a way that does not disrupt the production.

We see many directions for future theoretical work.

We are interested in modeling how agents with different information / belief formation processes interact on a market. As an extreme example, we wonder about stability when some agents on the market have complete information, some other agents have incomplete information and valuation beliefs.

We also wonder about the behavior of strategic agents who understand the belief formation process. In the valuation model, if an agent is currently very happy in his match, this makes his current partner attractive to other agents. This likely results in blocks and the happy agent losing his partner. Thus, happy agents have a clear incentive to hide their happiness. This is a challenge for strategy-proofness of a mechanism that would ask to report happiness levels.

We would also like to introduce a notion of farsightedness and to compute a farsighted stable set with valuation beliefs. Indeed, agents in the baseline model are myopic in two ways. They are myopic in the usual sense because when they participate in a blocking pair, they do not anticipate that the new matching produced might also be blocked and that the series of blocks could finally lead to a matching that could be worse than the initial one. They are also myopic in a new behavioral sense: they do not anticipate that their preferences will change in the new match and that they could be better off in the initial matching according to the new preferences.

We are also interested in the design challenge. The question is: How can one design matching markets to maximize social objectives, taking into account that participants may form valuation beliefs? In particular, can we propose a v -stable mechanism - a mechanism such that, when applied to any preference structure where a v -stable matching exists, it finds this matching? In appendix §B.2 we argue that the endogeneity in the valuation beliefs creates technical design difficulties. But we show that at least, a v -stable matching is a fixed point of the following procedure: take a given matching and the preferences that are induced by valuation beliefs in this matching and naively apply deferred acceptance.

In matching with imperfect information, pairwise and group stability are no longer equivalent. It would therefore be interesting to study group stability in the valuation setting, potentially producing even fewer existence cases. The nice feature of the model in this respect is the symmetry of the information structure: because all agents have access to the same feedback data on the current matching, we need not make arbitrary assumptions about how agents share information within coalitions.

Yet, we consider that the most pressing exercise is an empirical test of the theory. In a laboratory experiment, we could match participants to perform an incentivized task in pairs and provide feedback information on payoffs. We would then ideally elicit beliefs about counterfactual utilities, and give the participants the opportunity to be reassigned through blocking proposals. We would test the consistency of the beliefs and blocking decisions with the valuation heuristic and check whether the market stabilizes on v -stable matchings.

References

- [1] Y. Antler. Two-sided matching with endogenous preferences. *American Economic Journal: Microeconomics*, 7(3):241–258, 2015.
- [2] Y. Antler. No one likes to be second choice. *The Economic Journal*, 129(619):1119–1138, 2019.
- [3] Y. Antler and B. Bachi. Searching forever after. *American Economic Journal: Microeconomics*, 14(3):558–590, 2022.
- [4] N. Arnosti. Centralized clearinghouse design: A quantity-quality tradeoff. Available at SSRN 2571527, 2016.
- [5] G. Artemov. Assignment mechanisms: common preferences and information acquisition. *Journal of Economic Theory*, 198:105370, 2021.
- [6] G. Artemov, Y.-K. Che, and Y. He. Stable matching with mistaken agents. *Journal of Political Economy Microeconomics*, 1(2):270–320, 2023.
- [7] I. Ashlagi and Y. A. Gonczarowski. Stable matching mechanisms are not obviously strategy-proof. *Journal of Economic Theory*, 177:405–425, 2018.
- [8] S. Bade. Serial dictatorship: The unique optimal allocation rule when information is endogenous. *Theoretical Economics*, 10(2):385–410, 2015.
- [9] G. S. Becker. A theory of marriage: Part i. *Journal of Political economy*, 81(4):813–846, 1973.
- [10] S. Bikhchandani. Stability with one-sided incomplete information. *Journal of Economic Theory*, 168:372–399, 2017.
- [11] S. F. Bucher and A. Caplin. Inattention and inequity in school matching. Technical report, National Bureau of Economic Research, 2021.
- [12] G. Caspari and M. Khanna. Non-standard choice in matching markets. *arXiv preprint arXiv:2111.06815*, 2021.
- [13] H. Chade, G. Lewis, and L. Smith. Student portfolios and the college admissions problem. *Review of Economic Studies*, 81(3):971–1002, 2014.
- [14] A. Chakraborty, A. Citanna, and M. Ostrovsky. Two-sided matching with interdependent values. *Journal of Economic Theory*, 145(1):85–105, 2010.
- [15] Y.-C. Chen and G. Hu. Learning by matching. *Theoretical Economics*, 15(1):29–56, 2020.
- [16] R. M. Dawes. Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, 25(1):1–17, 1989.
- [17] A. Delavande and C. F. Manski. Candidate preferences and expectations of election outcomes. *Proceedings of the National Academy of Sciences*, 109(10):3711–3715, 2012.
- [18] B. Dutta and R. Vohra. Incomplete information, credibility and the core. *Mathematical Social Sciences*, 50(2):148–165, 2005.
- [19] L. Ehlers and J. Massó. Incomplete information and singleton cores in matching markets. *Journal of Economic Theory*, 136(1):587–600, 2007.
- [20] I. Esponda. Behavioral equilibrium in economies with adverse selection. *American Economic Review*, 98(4):1269–1291, 2008.
- [21] D. Faro and Y. Rottenstreich. Affect, empathy, and regressive mispredictions of others' preferences under risk. *Management Science*, 52(4):529–541, 2006.

- [22] D. Foster and P. Young. Stochastic evolutionary game dynamics. *Theoretical population biology*, 38(2):219–232, 1990.
- [23] T. Gagnon-Bartsch. Taste projection in models of social learning. Technical report, working paper, 2017.
- [24] T. Gagnon-Bartsch and A. Rosato. Quality is in the eye of the beholder: Taste projection in markets with observational learning. *mimeo*, 2022.
- [25] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- [26] S. Gleyze and A. Pernoud. Informationally simple incentives. *mimeo*, 2020.
- [27] L. Goursat. Robust incomplete information stability: For matching markets without transfers. *mimeo*, 2023.
- [28] L. Goursat. Whether and where to apply: Information and discrimination in matching with priority scores. *mimeo*, 2023.
- [29] C. Guven, C. Senik, and H. Stichnoth. You can't be happier than your wife. happiness gaps and divorce. *Journal of Economic Behavior and Organization*, 82(1):110–130, 2012.
- [30] J. Hagenbach, F. Koessler, and T. Trégouët. Truthtelling in matching markets. *The Scandinavian Journal of Economics*, 119(4):882–909, 2017.
- [31] H. C. Hoppe, B. Moldovanu, and A. Sela. The theory of assortative matching based on costly signals. *The Review of Economic Studies*, 76(1):253–281, 2009.
- [32] C. M. Hoxby and C. Avery. The missing "one-offs": The hidden supply of high-achieving, low income students. Technical report, National Bureau of Economic Research, 2012.
- [33] N. Immorlica, J. Leshno, I. Lo, and B. Lucier. Information acquisition in matching markets: The role of price discovery. Available at SSRN 3705049, 2020.
- [34] R. W. Irving. An efficient algorithm for the “stable roommates” problem. *Journal of Algorithms*, 6(4):577–595, 1985.
- [35] P. Jehiel. Analogy-based expectation equilibrium. *Journal of Economic Theory*, 123(2):81 – 104, 2005.
- [36] S. Kakutani. A generalization of brouwer's fixed point theorem. *mimeo*, 1941.
- [37] M. Kandori, G. J. Mailath, and R. Rob. Learning, mutation, and long run equilibria in games. *Econometrica: Journal of the Econometric Society*, pages 29–56, 1993.
- [38] A. J. Kapor, C. A. Neilson, and S. D. Zimmerman. Heterogeneous beliefs and school choice mechanisms. *American Economic Review*, 110(5):1274–1315, 2020.
- [39] A. Kloosterman and P. Troyan. School choice with asymmetric information: Priority design and the curse of acceptance. *Theoretical Economics*, 15(3):1095–1133, 2020.
- [40] D. E. Knuth. *Mariage stables et leurs relations avec d'autres problèmes combinatoires*. Montreal: Les Presses de l'Universite de Montreal, 1976.
- [41] J. Krueger and R. W. Clement. The truly false consensus effect: an ineradicable and egocentric bias in social perception. *Journal of personality and social psychology*, 67(4):596, 1994.
- [42] S. Lee and L. Yariv. On the efficiency of stable matchings in large markets. Available at SSRN 2464401, 2014.

- [43] S.-H. Lee. Jumping the curse: Early contracting with private information in university admissions. *International Economic Review*, 50(1):1–38, 2009.
- [44] S. Li. Obviously strategy-proof mechanisms. *American Economic Review*, 107(11):3257–3287, 2017.
- [45] Q. Liu. Stability and bayesian consistency in two-sided markets. *American Economic Review*, 110(8):2625–2666, 2020.
- [46] Q. Liu, G. J. Mailath, A. Postlewaite, and L. Samuelson. Stable matching with incomplete information. *Econometrica*, 82(2):541–587, 2014.
- [47] V. Meisner. Report-dependent utility and strategy-proofness. *Management Science*, 69(5):2733–2745, 2023.
- [48] V. Meisner and J. von Wangenheim. Loss aversion in strategy-proof school-choice mechanisms. *Journal of Economic Theory*, 207:105588, 2023.
- [49] E. Nagypal. Optimal application behavior with incomplete information. *Manuscript, Econ. Dept., Northwestern Univ*, 1, 2004.
- [50] B. Pittel. The "stable roommates" problem with random preferences. *The Annals of Probability*, 21(3):1441–1477, 1993.
- [51] L. Pomatto. Stable matching under forward-induction reasoning. *Theoretical Economics*, 17(4):1619–1649, 2022.
- [52] A. Rees-Jones and R. Shorrer. Behavioral economics in education market design: A forward-looking review. *mimeo*, 2023.
- [53] M. Richter and A. Rubinstein. Unilateral stability in the roommate problem. *mimeo*, 2023.
- [54] L. Ross, D. Greene, and P. House. The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, 13(3):279–301, 1977.
- [55] A. E. Roth. Two-sided matching with incomplete information about others' preferences. *Games and Economic Behavior*, 1(2):191–209, 1989.
- [56] A. E. Roth. Al roth on matching markets. *Social Science Bits Podcast*, 2017.
- [57] A. E. Roth, U. G. Rothblum, and J. H. Vande Vate. Stable matchings, optimal assignments, and linear programming. *Mathematics of operations research*, 18(4):803–828, 1993.
- [58] A. E. Roth and J. H. Vande Vate. Random paths to stability in two-sided matching. *Econometrica: Journal of the Econometric Society*, pages 1475–1480, 1990.
- [59] U. G. Rothblum. Characterization of stable matchings as extreme points of a polytope. *Mathematical Programming*, 54:57–67, 1992.
- [60] R. Selten and R. S. Bielefeld. *Reexamination of the perfectness concept for equilibrium points in extensive games*. Springer, 1988.
- [61] L. Shapley. Some topics in two-person games. *Advances in game theory*, 52:1–29, 1964.
- [62] L. S. Shapley and M. Shubik. The assignment game i: The core. *International Journal of game theory*, 1(1):111–130, 1971.
- [63] C. A. Sims. Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690, 2003.

- [64] R. Spiegler. Behavioral implications of causal misperceptions. *Annual Review of Economics*, 12:81–106, 2020.
- [65] L. Van Boven and G. Loewenstein. Social projection of transient drive states. *Personality and social psychology bulletin*, 29(9):1159–1168, 2003.
- [66] J. H. V. Vate. Linear programming brings marital bliss. *Operations Research Letters*, 8(3):147–153, 1989.
- [67] G. L. White and P. E. Mullen. *Jealousy: Theory, research, and clinical strategies*. Guilford Press, 1989.
- [68] R. Wilson. Information, efficiency, and the core of an economy. *Econometrica: Journal of the Econometric Society*, pages 807–816, 1978.
- [69] M. Wiswall and B. Zafar. How do college students respond to public information about earnings? *Journal of Human Capital*, 9(2):117–169, 2015.

Appendix A Proofs

Proof of corollary 1.

$$(NBP)_\mu^v \begin{cases} v_{m\mu(m)}^\mu = u_{m\mu(m)} \\ v_{mf}^\mu = (1 - \lambda)u_{m\mu(m)} + \lambda u_{\mu(f)f} \end{cases}$$

m unwilling to block with $f \Leftrightarrow u_{m\mu(m)} \geq u_{\mu(f)f}$.
 v -stability is independent of exact v .

Proof of proposition 1.

$(NBP)_\mu^v$ m willing to block with f : $u_{\mu(f)f} > u_{m\mu(m)}$.

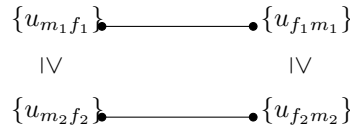
So m is willing to block with f iff he is jealous of $\mu(f)$.

So $(NBP)_\mu^v$ is equivalent to no pair of 2 jealous agents.

We prove that this is equivalent to happiness sorting, in two steps.

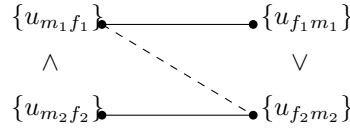
– Small market of size 2. We prove separately each direction of the equivalence.

\Leftarrow Suppose we do have happiness sorting as in the figure below (m_1 happier than m_2 , f_1 happier than f_2).



Then m_1 and f_1 are not jealous and do not want to enter a blocking pair with f_2 , m_2 respectively. One cannot find a pair of jealous agents who would block together.

\Rightarrow We prove this direction by contraposition. Suppose we do not have happiness sorting, as in the figure below (m_2 strictly happier than m_1). Then m_1 and f_2 are jealous and block.



– Market of any size n .

On a market of size n , blocking pairs could form within any subgroup of size 2 on the market.

So $(NBP)_\mu^v$ is equivalent to happiness sorting on any subgroup of size 2 on the market.

But imposing happiness sorting on any subgroup of size 2 on the market is equivalent to imposing happiness sorting on the whole market as in definition 4.

For a formal proof, consider again the two directions of the equivalence:

\Leftarrow Suppose we have happiness sorting at order n : $u_{m_1 f_1} \geq \dots \geq u_{m_n f_n}$ and $u_{f_1 m_1} \geq \dots \geq u_{f_n m_n}$.
 Let us consider a subgroup of pairs i and j ($i < j$). $u_{m_i f_i} > u_{m_j f_j}$ and $u_{f_i m_i} > u_{f_j m_j}$.
 Hence happiness sorting at order 2.

\Rightarrow Suppose, by contradiction that we do not have happiness sorting at order n : $u_{m_1 f_1} \geq \dots \geq u_{m_n f_n}$ but we do not have $u_{f_1 m_1} \geq \dots \geq u_{f_n m_n}$.

So there must be at least one $i \in \{1, \dots, n-1\}$ such that $u_{f_{i+1} m_{i+1}} > u_{f_i m_i}$ (otherwise, we could use transitivity to get back to $u_{f_1 m_1} \geq \dots \geq u_{f_n m_n}$).

This contradicts happiness sorting at order 2 for subgroup of pairs i and $i+1$.

Proof of proposition 2.

1. Full alignment: $a_{\{m,f\}} = a$

$$\text{So: } \begin{cases} u_{mf} = a s_{\{m,f\}} \\ u_{fm} = (1-a) s_{\{m,f\}} \end{cases}$$

Set $\mu \in \Sigma$.

Denote $\{m_k, f_k\}$, $k \in \{1, \dots, n\}$ the pair generating the k^{th} higher surplus in μ : $\begin{cases} \mu(m_k) = f_k \\ s_{m_1 f_1} \geq \dots \geq s_{m_n f_n} \end{cases}$

Multiplying the surplus by a positive constant keeps the order: $\begin{cases} a s_{m_1 f_1} \geq \dots \geq a s_{m_n f_n} \\ (1-a) s_{m_1 f_1} \geq \dots \geq (1-a) s_{m_n f_n} \end{cases}$

We get: $\begin{cases} u_{m_1 f_1} \geq \dots \geq u_{m_n f_n} \\ u_{f_1 m_1} \geq \dots \geq u_{f_n m_n} \end{cases}$

So happiness sorting is verified.

By proposition 1.: $\Sigma \subseteq \Sigma^v$, $\Sigma^v = \Sigma$.

2. Full misalignment: $s_{\{m,f\}} = s$

$$\text{So: } \begin{cases} u_{mf} = a_{\{m,f\}}s \\ u_{fm} = (1 - a_{\{m,f\}})s \end{cases}$$

Set $\mu \in \Sigma$.

Denote m_0 (f_0) the male (female) with the lowest share of the surplus in μ : $\begin{cases} m_0 := \arg \min_{m \in M} a_{\{m,f\}} \\ f_0 := \arg \min_{f \in F} 1 - a_{\{m,f\}} \end{cases}$

By definition, m_0 and f_0 are not currently matched together: $\mu(m_0) \neq f_0$.

$$\text{So: } \begin{cases} a_{\{m_0, \mu(m_0)\}} < a_{\{\mu(f_0), f_0\}} \\ 1 - a_{\{\mu(f_0), f_0\}} < 1 - a_{\{m_0, \mu(m_0)\}} \end{cases}$$

Multiplying the sharing rule by a positive constant keeps the order:

$$\text{So: } \begin{cases} a_{\{m_0, \mu(m_0)\}}s < a_{\{\mu(f_0), f_0\}}s \\ (1 - a_{\{\mu(f_0), f_0\}})s < (1 - a_{\{m_0, \mu(m_0)\}})s \end{cases}$$

$$\text{We get: } \begin{cases} u_{\{m_0, \mu(m_0)\}} < u_{\{\mu(f_0), f_0\}} \\ u_{\{f_0, \mu(f_0)\}} < u_{\{m_0, \mu(m_0)\}} \end{cases}$$

So happiness sorting is not verified.

By proposition 1.: $\Sigma^v = \emptyset$.

Proof of proposition 3.

1. Homogeneous preferences

$$\forall i < j \in \{1, \dots, n\}, \begin{cases} \forall m, m' : u_{mf_i} > u_{m'f_j} \\ \forall f, f' : u_{fm_i} > u_{f'm_j} \end{cases}$$

- Set $\mu = \mu^\oplus$, $\forall i \in \{1, \dots, n\} : \mu^\oplus(m_i) = \mu^\oplus(f_i)$.

$$\text{By definition of preferences: } \begin{cases} u_{m_1f_1} > \dots > u_{m_nf_n} \\ u_{m_nf_n} > \dots > u_{m_1f_1} \end{cases}$$

So happiness sorting is verified.

By proposition 1.: $\mu \in \Sigma^v$.

- Set $\mu \neq \mu^\oplus \in \Sigma$.

$$\text{So } \exists i, i', j, j' \text{ s.t. } \begin{cases} \mu(m_i) = f_j, \mu(m_{i'}) = j' \\ i < i', j > j' \end{cases}$$

$$\text{By definition of preferences: } \begin{cases} u_{m_i f_j} < u_{m_{i'} f_{j'}} \\ u_{f_j m_i} > u_{f_{j'} m_{i'}} \end{cases}$$

So happiness sorting is not verified ($\{m_i, f_{j'}\}$ blocking pair).

By proposition 1.: $\mu \notin \Sigma^v$.

In total: $\Sigma^v = \{\mu^\oplus\}$.

2. Heterogeneous preferences

$$\forall i < j \in \{1, \dots, n\}, \begin{cases} \forall f, f' : u_{mf_i} > u_{m_j f'} \\ \forall m, m' : u_{f_i m} > u_{f_j m'} \end{cases}$$

- Set $\mu = \mu^\oplus$, $\forall i \in \{1, \dots, n\} : \mu^\oplus(m_i) = \mu^\oplus(f_i)$.

$$\text{By definition of preferences: } \begin{cases} u_{m_1f_1} > \dots > u_{m_nf_n} \\ u_{m_nf_n} > \dots > u_{m_1f_1} \end{cases}$$

So happiness sorting is verified.

By proposition 1.: $\mu \in \Sigma^v$.

- Set $\mu \in \Sigma, \neq \mu^\oplus$.

$$\text{So } \exists i, i', j, j' \text{ s.t. } \begin{cases} \mu(m_i) = f_j, \mu(m_{i'}) = j' \\ i < i', j > j' \end{cases}$$

$$\text{By definition of preferences: } \begin{cases} u_{m_i f_j} > u_{m_{i'} f_{j'}} \\ u_{f_j m_i} < u_{f_{j'} m_{i'}} \end{cases}$$

So happiness sorting is not verified ($\{m_{i'}, f_j\}$ blocking pair).

By proposition 1.: $\mu \notin \Sigma^v$.

In total: $\Sigma^v = \{\mu^\oplus\}$.

Proof of corollary 2.

Using counter-example in FIGURE 6 or result in proposition 2., statement 2..

Proof of proposition 4.

(i) Convergence

– Market size 1

There is a unique matching. Any DBPP starts and stays on this matching. If this matching is v-stable, we trivially have convergence.

– Market size 2

There are two different matchings.

If both matchings are v-stable, the proof is similar to case of market of size 1. Hence convergence.

If only one of the two matchings is v-stable, the DBPP can either start from the v-stable matching (in which case it converges) or start from the v-unstable matching (in which case, whatever the blocking pair, we move to the other (v-stable) matching and stay there). Hence convergence.

(ii) No convergence

– Market size 3

Using counter-example in FIGURE 9.

Proof of proposition 5.

- Set p mixed matching.

Define:

$$\forall M, F : p(M, F) := \sum_{\mu | \mu(M)=F} p(\mu)$$

Then:

$$\begin{aligned} \forall M : \sum_F p(M, F) &= \sum_F \sum_{\mu | \mu(M)=F} p(\mu) \\ &= \sum_{\mu} p(\mu) \\ &= 1 \end{aligned}$$

$$\forall M : \sum_M p(M, F) = 1 \text{ (by symmetry)}$$

- Set p behavioral matching.

Define:

$$\forall \mu : p(\mu) = \frac{1}{n(n-1)!} \sum_{M, F | \mu(M)=F} p(M, F)$$

Then:

$$\begin{aligned} \sum_{\mu} p(\mu) &= \sum_{\mu} \frac{1}{n(n-1)!} \sum_{M, F | \mu(M)=F} p(M, F) \\ &= \frac{1}{n(n-1)!} \sum_{M, F} (n-1)! p(M, F) \quad ((n-1)! \text{ pure matchings where } M \text{ and } F \text{ are matched}) \\ &= \frac{1}{n} \sum_M \sum_F p(M, F) \\ &= \frac{1}{n} \sum_M 1 \\ &= 1 \end{aligned}$$

Proof of lemma 1.

- Notations

There are n pairs of types (M, F) .

The weights are constrained by a system of $2n$ equations, each involving n terms:

$$\begin{aligned}\forall M \in \mathcal{M}, \sum_F p(M, F) &= 1 \\ \forall F \in \mathcal{F}, \sum_M p(M, F) &= 1\end{aligned}$$

Let us number the types $\mathcal{M} = \{M_1, M_n\}$, $\mathcal{F} = \{F_1, F_n\}$ and denote $p_{ij} = p(M_i, F_j)$.

The system writes:

$$\begin{aligned}\forall i \in \{1, \dots, n\}, \sum_j p_{ij} &= 1 (M_i) \\ \forall j \in \{1, \dots, n\}, \sum_i p_{ij} &= 1 (F_j)\end{aligned}$$

- Induction formula

Let us fix p_{i1} , $i \in \{1, \dots, n-1\}$ ($n-1$ variables). This fixes by (F_1) : $p_{n1} = 1 - \sum_{i=1}^{n-1} p_{i1}$.

The system now writes:

$$\begin{aligned}\forall i \in \{1, \dots, n\}, \sum_{j=1}^n p_{ij} &= 1 - p_{i1} (M_i) \\ \forall j \in \{2, \dots, n\}, \sum_{i=1}^n p_{ij} &= 1 (F_j)\end{aligned}$$

Let us fix p_{1j} , $j \in \{2, \dots, n-1\}$ ($n-2$ variables). This fixes by (M_1) : $p_{1n} = 1 - \sum_{j=1}^{n-1} p_{1j}$.

The system now writes:

$$\begin{aligned}\forall i \in \{2, \dots, n\}, \sum_{j=2}^n p_{ij} &= 1 - p_{i1} (M_i) \\ \forall j \in \{2, \dots, n\}, \sum_{i=2}^n p_{ij} &= 1 - p_{1j} (F_j)\end{aligned}$$

So, by fixing $(n-1) + (n-2) = 2n-3$ variables, we get to as system of $2n-2 = 2(n-1)$ equations with $n-1$ terms, just as on a market of size $(n-1)^2$.

So:

$$\dim(P_n) = \dim(P_{(n-1)^2}) + 2n - 3$$

- Induction proof

We iterate on n and get the explicit formula:

$$\dim(P_n) = \dim(P_2) + \sum_{k=2}^n (2k - 3)$$

With $n = 2$, the system writes:

$$\begin{cases} p_{11} + p_{12} = 1 \\ p_{21} + p_{22} = 1 \\ p_{11} + p_{21} = 1 \\ p_{12} + p_{22} = 1 \end{cases}$$

If we set a single variable, this fixes the three other variables.

So: $\dim(P_2) = 1$

We get:

$$\begin{aligned}\dim(P_n) &= 1 + 2 \sum_{k=2}^n k - 3(n-2) \\ &= 1 + 2 \left(\frac{n(n+1)}{2} - 3 \right) - 3(n-2) \\ &= 1 + n(n+1) - 3n \\ &= 1 + n - 2n\end{aligned}$$

Proof of theorem 1.

Set market parameters: market size n , and match utilities $u_{MF}, u_{FM}, \forall M, F$.

1. PRELIMINARIES

• Notations and vocabulary

– Matchings:

p : Mixed matching.

$P_{\mathcal{M} \times \mathcal{F}}$: Set of mixed matchings.

\mathcal{Q} : Quadruple $\{M, M', F, F'\}$.

“Local”= Restricted to quadruple \mathcal{Q} .

$p_{\mathcal{Q}}$: Local mixed matching.

$P_{\mathcal{Q}}$: Set of local mixed matchings.

μ_a, μ_b : Pure local matchings ($\mu_a(M) = F, \mu_a(M') = F'$ and $\mu_b(M) = F', \mu_b(M') = F$).

– Subsets:

$\mathcal{P}(E)$: Set of subsets of E .

– Blocking pairs:

“Strict blocking pair”: Both agents strictly want to block.

“Weak blocking pair”: At least one agent is indifferent between blocking or not.

“Possible blocking pairs”: Any set of blocking pairs that contains all strict blocking pairs and potentially some weak blocking pairs.

“Possible matchings”: Matching obtained after implementing blocking pairs taking into account all strict blocking pairs and potentially some weak blocking pairs.

“Effective blocking pairs”: Set of blocking pairs, once ties are broken and weak blocking pairs are considered as effectively blocking or not.

“Effective matching”: Matching obtained after implementing blocking pairs taking into account effective blocking pairs.

• Remark

- When we examine a blocking pair locally, we do not specify the identity of the pair. We only specify to which pure matching the blocking pair applies. The transformation will be the same whatever the exact blocking pair.

2. TRANSFORMATION FUNCTION

• Local functions

(g) **Counting blocking pairs in a local matching.**

Set-valued function mapping a local matching to sets of possible numbers of blocking pairs.

$$g_{\mathcal{Q}} : P_{\mathcal{Q}} \rightrightarrows \{0, 1, 2\}^2$$

$$p \mapsto E$$

Where:

If there are $s_i \in \{0, 1, 2\}$ ($w_i \in \{0, 1, 2\}$) strong (weak) blocking pairs in μ_i : $E = \{(x_a, x_b) | s_i \leq x_i \leq s_i + w_i\}$.

Set E contains $l + 1$ elements that count zero / some / all weak blocking pairs, plus always all strict blocking pairs.

Special cases, where the set E is a singleton:

- * No blocking pairs ($s_i = w_i = 0$): $E = \{(0, 0)\}$.
- * Only strict blocking pairs ($w_i = 0$): $E = \{(s_a, s_b)\}$.

(k) **Local transformation given effective blocking pairs.**

Single-valued function mapping a number of local blocking pairs to a new local matching.

$$k_{\mathcal{Q}} : \{0, 1, 2\}^2 \rightarrow P_{\mathcal{Q}}$$

$$e \mapsto r$$

Where:

- * If same number of blocking pairs in both local pure matchings ($e_a = e_b$):
eg.: no blocking pairs in quadruple ($e_a = e_b = 0$).
eg.: one blocking pair in each pure matching ($e_a = e_b = 1$).

Then: k_Q implements the blocking pairs until the lowest mass ex-post blocking type is exhausted. In total, the local matching remains the same:

$$r = p_Q$$

- * If imbalance in the number of blocking pairs in the two pure matchings (set $e_a > e_b$):
 eg.: one blocking pair from μ_a to μ_b ($e_a = 1 > e_b = 0$).
 eg.: two blocking pairs from μ_a to μ_b , one from μ_b to μ_a ($e_a = 2 > e_b = 1$).
 Then: k_Q implements the blocking pairs until the lowest mass ex-post blocking type is exhausted. In total, the local matching is modified in the following way:

$$\begin{aligned} \cdot \text{ If } p(M, F') \leq p(M', F), \text{ then } & \begin{cases} r(M, F) := p_t(M, F) + p_t(M, F') \\ r(M', F') := p_t(M', F') + p_t(M, F') \\ r(M, F') := 0 \\ r(M', F) := p_t(M', F) - p_t(M, F') \end{cases} \\ \cdot \text{ If } p(M, F') \geq p(M', F), \text{ then } & \begin{cases} r(M, F) := p_t(M, F) + p_t(M', F) \\ r(M', F') := p_t(M', F') + p_t(M', F) \\ r(M, F') := p_t(M, F') - p_t(M', F) \\ r(M', F) := 0 \end{cases} \end{aligned}$$

(h) **Local transformations given possible blocking pairs.**

Set-valued function from the set of possible numbers of local blocking pairs to new possible local matchings.

$$h_Q^p : \mathcal{P}(\{0, 1, 2\}^2) \rightrightarrows P_Q \\ E \mapsto R$$

- * If set E is a singleton ($|E| = 1$, $E = \{e\}$), then: $R = \{k_Q(e)\}$
- * If set E contains more than one element ($|E| > 1$), then: R contains all convex combinations of the elements in E .

(f) **Local transformations from a matching to possible matchings.**

Set-valued function mapping a local matching to a set of new possible local matchings. By composition:

$$f_Q^p : P_Q \rightrightarrows P_Q \\ f_Q^p(p_Q) = h_Q^p(g_Q(p_Q))$$

• Integral functions

– **Global transformations from a matching to possible matchings with changes only in one quadruple.**

Set-valued function mapping a matching to a set of new possible matching obtained after implementing the blocking pairs in quadruple Q .

$$F_Q : P_{\mathcal{M} \times \mathcal{F}} \rightrightarrows P_{\mathcal{M} \times \mathcal{F}} \\ p \mapsto R$$

Where:

$$R(M_0, F_0) = \begin{cases} f_Q^p(p)(M_0, F_0) & \forall M_0, F_0 \in Q \\ \{p(M_0, F_0)\} & \text{if } M_0 \text{ or } F_0 \notin Q \end{cases}$$

– **DBPP with random draws of the quadruple.**

At each step, the DBPP randomly draws a quadruple Q and transform the matching according to F_Q .

3. FIXED POINT THEOREM

(a) The graph of F_Q is closed

The condition for $(M - F')$ being willing to block with F writes:

$$v_{mF}^p > u_{MF'} \iff \sum_{M_0 \in \mathcal{M}} p(M_0, F) u_{M_0 F} > u_{MF'}$$

Due to the linearity, the space $P_{\mathcal{M} \times \mathcal{F}}$ is divided into regions where an agent is willing to block, unwilling to block, and a sub-space of dimension $[1 + n - 2n] - 1 = n(n - 2)$ where he is indifferent.

Suppose, for instance, that at a given point in the indifference sub-space p^* , $(M - F')$ is indifferent to block with F . In a neighbourhood of p^* , $(F - M')$ is willing to block with M (and other pairs do not block).

By definition of our auxiliary functions:

$$\begin{aligned} g_{\mathcal{Q}}(p_{\mathcal{Q}}^*) &= \{\{0, 1\}, \{0, 0\}\} \\ k_{\mathcal{Q}}(\{\{0, 1\}\}) &= r \text{ (as above)} \\ k_{\mathcal{Q}}(\{0, 0\}) &= p_{\mathcal{Q}}^* \\ h_{\mathcal{Q}}^{p^*}(\{\{0, 1\}, \{0, 0\}\}) &= \{\lambda \cdot r + (1 - \lambda) \cdot p_{\mathcal{Q}}^*, \lambda \in [0, 1]\} \\ f_{\mathcal{Q}}^{p^*}(p_{\mathcal{Q}}^*) &= \{\lambda \cdot f_{\mathcal{Q}}^{p_1} + (1 - \lambda) \cdot f_{\mathcal{Q}}^{p_2}, \lambda \in [0, 1]\} \\ &\text{where at } p_1 \text{ (} p_2 \text{), } (F - M') \text{ is willing (unwilling) to block with } M \end{aligned}$$

So the set $Gr(f_{\mathcal{Q}}^p) = \{(p, f_{\mathcal{Q}}^p(p)), p \in P_{\mathcal{Q}}\}$ is closed.

The integral function $F_{\mathcal{Q}}$ is defined as doing the changes in quadruple \mathcal{Q} as in $f_{\mathcal{Q}}^p$ and nothing in other quadruples.

So the set $Gr(F_{\mathcal{Q}}) = \{(p, F_{\mathcal{Q}}(p)), p \in P_{\mathcal{M} \times \mathcal{F}}\}$ is also closed.

In summary, the definition of $h_{\mathcal{Q}}^p$ for a value E s.t. $|E| > 1$ allows for full flexibility. At the local level, an indifferent individual in a weak blocking pair can be allowed to block or not to block. At the macro level, any proportion of the indifferent type can form a blocking pair. This closes the graph of $F_{\mathcal{Q}}$.

(b) F is upper hemi-continuous

Closed graph theorem

If $\Gamma : A \rightrightarrows B$ is an upper hemi-continuous set-valued function with closed domain and closed values if and only if the graph of Γ is closed.

We apply the closed graph theorem with $\Gamma = F_{\mathcal{Q}}$, $A = B = P_{\mathcal{M} \times \mathcal{F}}$.

So $F_{\mathcal{Q}}$ is upper hemi-continuous.

(c) Kakutani theorem applied to $F_{\mathcal{Q}}$

Kakutani fixed point theorem

Set S a non-empty, compact, convex subset of \mathbb{R}^m .

$\phi : S \rightrightarrows$ upper hemi-continuous set-valued function with the property that $\phi(x)$ is non-empty, closed and convex for all $x \in S$.

We apply Kakutani fixed point theorem with $m = n$, $S = P_{\mathcal{M} \times \mathcal{F}}$, $\phi = F_{\mathcal{Q}}$.

We conclude that $F_{\mathcal{Q}}$ has a fixed point.

By definition 10., this fixed point is a v-stable matching.

Conclusion: There always exists a v-stable matching.

Proof of lemma 2.

The mass of a quadruple $\{M, M', F, F'\}$ is given by the sum of the measures of the pairs in this quadruple:

$$S_{\{M, M', F, F'\}} := p(M, F) + p(M', F) + p(M, F') + p(M', F')$$

A uniform distribution draws quadruples according to their S .

To get a probability distribution, we need some normalization:

$$X_{\{M, M', F, F'\}} = \frac{S_{\{M, M', F, F'\}}}{C}$$

We equalize the sum of the probabilities to 1 - In the sum, each pair $\{M, F\}$ appears exactly $(n - 1)^2$ times.

$$\begin{aligned} \sum_{\{M, M', F, F'\}} X_{\{M, M', F, F'\}} &= \frac{1}{C} \sum_{M, F} (n - 1)^2 p(M, F) \\ &= \frac{1}{C} (n - 1)^2 \sum_M \sum_F p(M, F) \\ &= \frac{1}{C} (n - 1)^2 n \\ \frac{1}{C} (n - 1)^2 n &= 1 \iff C = \frac{1}{n(n - 1)^2} \end{aligned}$$

Proof of proposition 6.

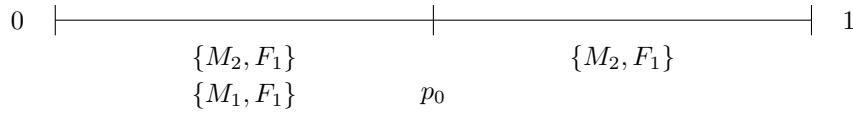
We write the blocking pairs conditions for the whole market:

$$\begin{aligned} \{M_1, F_2\} & \begin{cases} a_{11} < pa_{22} + (1-p)a_{12} \\ 1 - a_{22} < p(1 - a_{11}) + (1-p)(1 - a_{12}) \end{cases} \\ \{M_2, F_1\} & \begin{cases} a_{22} < pa_{11} + (1-p)a_{21} \\ 1 - a_{11} < p(1 - a_{22}) + (1-p)(1 - a_{21}) \end{cases} \\ \{M_1, F_1\} & \begin{cases} a_{12} < pa_{11} + (1-p)a_{21} \\ 1 - a_{21} < p(1 - a_{11}) + (1-p)(1 - a_{12}) \end{cases} \\ \{M_2, F_2\} & \begin{cases} a_{21} < pa_{22} + (1-p)a_{12} \\ 1 - a_{12} < p(1 - a_{22}) + (1-p)(1 - a_{21}) \end{cases} \end{aligned}$$

We rewrite inequalities as a function of p and determine for each $p \in [0, 1]$ what are the blocking pairs.

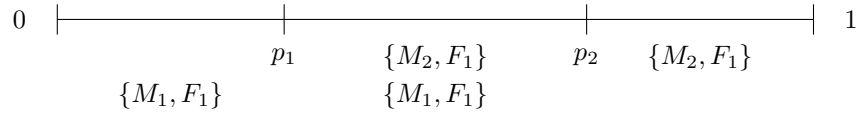
We get:

(1) $a_{12} < a_{21} < a_{22} < a_{11}$



(2) $a_{12} < a_{22} < a_{21} < a_{11}$

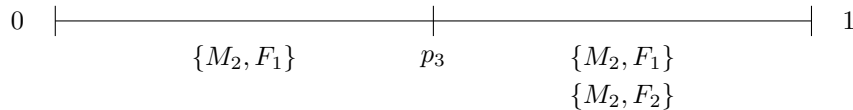
(a) $(a_{22} - a_{21})(a_{11} - a_{12}) < (a_{21} - a_{12})(a_{11} - a_{12})$



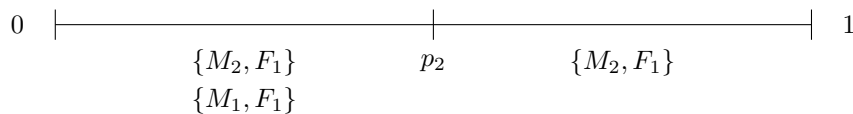
(b) $(a_{21} - a_{12})(a_{11} - a_{12}) < (a_{22} - a_{21})(a_{11} - a_{12})$



(3) $a_{22} < a_{21} < a_{12} < a_{11}$



(4) $a_{22} < a_{12} < a_{21} < a_{11}$

**Proof of proposition 7.**

(i) We set $u_{MF} := v(x_{MF})$ and we are back to the case in section §4.

(ii) The ordinal preferences enable to define an order on realized bundles. So we do exactly the same proof than in the proof of proposition 1., except that we compare realized bundles instead of realized match utilities.

Proof of proposition 9.

(i) By u-stability of DA.

(See u-stable matching constructive existence proof in Gale and Shapley (1962) [25]).

(ii) Set $\mu \in \Sigma^v$.

So μ satisfies happiness sorting:

$$\forall i \in \{1, \dots, n\}, \mu(m_i) = f_i, \text{ with } \begin{cases} u_{m_1 f_1} \geq \dots \geq u_{m_n f_n} \\ u_{f_1 m_1} \geq \dots \geq u_{f_n m_n} \end{cases}$$

- The valuation preferences ($\succ [V]^\mu$) are ordinally homogeneous: $\begin{cases} m_1 \succ_f \dots \succ_f m_n \\ f_1 \succ_m \dots \succ_m f_n \end{cases}$
 Indeed: $\forall m \in M, i < j \in \{1, \dots, n\} : u_{m_i f_i} > u_{m_j f_j} \implies v_{m_f}^\mu > v_{m_f'}^\mu \implies f_i \succ_m f_j$.
- DA applied to homogeneous preferences has a unique solution, the unique stable matching: the positive assortative matching μ^\oplus (and terminates in n rounds).
 Precisely, in (male-proposing) DA: In round 1, all males m_1, \dots, m_n propose to f_1 , f_1 accepts m_1 , and rejects all other males.
 In further round $i \geq 2$, all males m_i, \dots, m_n propose to the same female f_i , f_i accepts m_i , rejects all other males.
- Here, by definition of valuation preferences: $\mu^\oplus := \mu$.

Appendix B Additional content

B.1 Valuation with memory

Memory and dynamic process

We extend the model introducing some memory in the dynamic. Agents keep memory of past experience: they know about match utilities they enjoyed in the past, even after changing partners. The motivation for this extension is two fold. First, it seems very reasonable to assume that agents learn, especially in the closed cycles of the DBPP identified in section §3.4 where agents explore the same matchings infinitely often. Second, with memory the valuation becomes less sensitive to the current matching and more sticky over time; hence some hope that memory could restore existence of / convergence to v-stable matchings.

The memory puts weights on matchings that were explored in the past according to their historical frequencies. Formally, it is a probability distribution over matchings: $\alpha \in \Delta(\Sigma)$. In this context, the valuation definition naturally extends:

Definition 14.

The **valuation** is a function mapping a male $m \in M$, a female $f \in F$ and a memory $\alpha \in \Delta(\Sigma)$ to a belief of male m over a targeted partner f when the current matching is μ :

$$v : \Delta(\Sigma) \times M \times F \times \Sigma_f \rightarrow \mathbb{R}$$

$$v_{mf}^\alpha = \sum_{\mu \in \Sigma} \alpha(\mu) u_{\mu(f)f}$$

The valuation of agent m about target f under memory α is given by the aggregate satisfaction of f 's historical partners.

We again define a dynamic process based on blocking pairs:

Definition 15.

A **memory dynamic blocking pair process (MDBPP)** is a dynamic process s_t where:

- A state of the process s_t is given by a matching $\mu_t \in \Sigma$ and a memory $\alpha_t \in \Delta(\Sigma)$.
- The process moves from a state $s_t = (\mu_t, \alpha_t)$ to another state $s_{t+1} = (\mu_{t+1}, \alpha_{t+1})$ in the following way:
 - Matching:
 - * If there is no blocking pair in μ_t : $\mu_{t+1} = \mu_t$
 - * If there is a blocking pair $\{m, f\}$ in state μ_t :

$$\begin{cases} \mu_{t+1}(m) = f, \mu_{t+1}(f) = m \\ \text{if } \mu_t(m) \neq \emptyset, \mu_{t+1}(\mu_t(m)) = \mu_t(f), \text{ if } \mu_t(f) \neq \emptyset, \mu_{t+1}(\mu_t(f)) = \mu_t(m) \\ \forall m' \in M - \{m, \mu(f)\} : \mu_{t+1}(m') = \mu_t(m'), \forall f' \in F - \{f, \mu(m)\} : \mu_{t+1}(f') = \mu_t(f') \end{cases}$$
 - Memory: $\alpha_{t+1} = \frac{t}{t+1}\alpha_t + \frac{1}{t+1}\mu_{t+1}$

The moves of the matching are defined just as in the standard DBPP. Each time we move to a new matching, the history stores one more matching, and the memory updates.

Steady states

The solution concept is a notion of steady state of the dynamic process. There exists a steady state of the MDBPP if after some time, the pair of the memory and current matching remains unchanged.

For illustration, we characterize this steady state on a market on size 2. On this market, there are only two possible matchings μ and μ' so that memory can be described by just one scalar α (the weight on μ). This is summarized on the following graph:



FIGURE 2.16: Memory in a market of size 2

For example, m_1 is willing to block with f_2 in matching μ under memory α if his current utility a_1 is strictly lower than the convex combination $\alpha a_2 + (1 - \alpha)x_1$.

No existence

We can use this small dimension framework to show that non existence of a steady state is generic:

Proposition 8.

A steady state of the MDBPP does not always exist.

The proof uses the following counter-example:



FIGURE 2.17: Counter-example for existence of steady state of the MDBPP

Consider μ . m_1 gets the highest utility a male can achieve so never blocks.⁵³ Conversely, f_1 receiving the lowest possible utility always block.⁵⁴ The only possibility for a blocking pair is $\{m_2, f_1\}$. m_2 compares 5 to a convex combination between 6 and 2. So he blocks only if α is very high, close to one (the threshold is $\frac{3}{4}$). Conversely, in μ' , $\{m_1, f_1\}$ is a blocking pair if α is low (the threshold is $\frac{1}{4}$). For intermediate memory levels, there is no blocking pair. The next figure summarizes blocking pairs for different levels of memory:

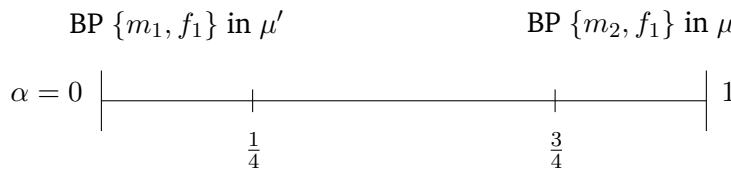


FIGURE 2.18: Blocking pairs as a function of the memory in the counter example from FIGURE 17

Suppose we start from an initial state $s_0 = (\mu, 1)$. Because for $\alpha = 1$, there is a blocking pair in μ , the process moves to state $s_1 = (\mu', \frac{1}{2})$. Because for intermediate α , there is no blocking pair in μ' , the process moves to state $s_2 = (\mu', \frac{1}{3})$, then to $s_3 = (\mu', \frac{1}{4})$ then to $s_4 = (\mu', \frac{1}{5})$. When $\alpha = \frac{1}{5}$, there is a blocking pair in μ' , so that the process moves to $s_5 = (\mu, \frac{1}{3})$, again in the intermediate memory region, where there is no blocking pair in μ . In the next states, the matching is μ , and the memory increases, until it crosses the $\frac{3}{4}$ threshold. Memory cycles between the two extreme regions, including in the limit $t \rightarrow \infty$.

This is the case for any the initial starting point. The limit cycle is similar to the no convergence cases in fictitious play (Shapley (1964) [61]). Due to the memory accumulation (no discount, no stop), the process spends more and more time on the same matching before switching.

The conclusion is that memory does not induce v-stability of a pure matching. This further motivates the shift to the general mixed framework.

B.2 V-stable mechanism

From a design perspective, the main challenge is to find a v-stable mechanism: a mechanism, such that, when applied to any preference structure where a v-stable matching exists, it finds this matching (or it reports that no stable matching exists as in Irving (1985) [34] for the roommate problem).

The difficulty in this task comes from the fact that beliefs are endogenous to the state of the market. Thus, if the mechanism operates with rounds and tentative matchings, it would make sense to let agents revise their valuations at each new tentative matching. Hence the possibility for regret: some agents would rather go back to the previous round and propose / accept differently.⁵⁵ Our intuition is that it should be possible to prove an impossibility result as in Kloosterman and Troyan (2020) [39].

A closely related question is: what happens if a market designer naively applies a u-stable mechanism, such as deferred acceptance, to preferences that have been formed given an initial (pure, full) matching and a valuation belief-formation process? The next proposition answers this question.

⁵³We materialize this on the figure with the cross.

⁵⁴We materialize this on the figure with the circle.

⁵⁵Oppositely, the standard constructive proof in perfect information uses the fact that in Deferred Acceptance, there can be no regret.

Proposition 9.

Set $\mu \in \Sigma$, and denote $\succ [v]^\mu$ the induced ordinal preferences.

Apply the Deferred Acceptance mechanism (Gale and Shapley (1962) [25]) to the $\succ [v]^\mu$ preferences.

Denote $DA(\succ [v]^\mu)$ the resulting matching.

Then:

(i) $DA(\succ [v]^\mu) = \mu'$, μ' u-stable with respect to preferences v^μ

(ii) μ v-stable $\implies DA(\succ [v]^\mu) = \mu$ (n rounds)

Statement (i) is a direct implication of the fact that DA is a u-stable mechanism.

Statement (ii) says that if the initial matching μ is v-stable, DA delivers exactly the same matching μ , and is quite quick (n rounds). The proof recalls that valuation beliefs always generate homogeneous ordinal preferences. And by proposition 3., the outcome of DA on homogeneous preferences is the unique stable matching, in this case, μ .

So a v-stable matching is a fixed point of this procedure. The contraposition provides us with a test for v-stability that can be even quicker to implement than checking happiness sorting. We run DA, keeping valuation preferences fixed. If the resulting matching is different from the initial one, this signals that the initial matching was v-unstable.

Ideally we would like to be able to predict the outcomes or more standard mechanisms (such as variants of deferred acceptance as in Antler (2015) [1]).

Chapter 3

ROBUST INCOMPLETE INFORMATION STABILITY

In matching markets without monetary transfers

Laure GOURSAT* – December 2023 – [*Preliminary version*]

We thank Francis BLOCH, Olivier COMPTE, Philippe JEHIEL, Andrew POSTLEWAITE, George MAILATH, Larry SAMUELSON, Olivier TERCIEUX for helpful comments and conversations.

We thank the European Research Council (Philippe JEHIEL's advanced grant 2018-2022 LTCSEI) and Ecole Nationale des Ponts et Chaussées (doctoral funding contract n°20/092) for financial support.

Abstract: We consider a matching market with no transfers and incomplete asymmetric information - on one side, agents do not observe types of potential partners; they just observe the type of their current partner. The model can represent civil servants' job markets where wages are regulated and where employers have trouble learning about workers' productivity prior to hiring. We apply the definition of incomplete-information stable matchings by Liu, Mailath, Postlewaite, and Samuelson (2014) - a pair is blocking if both partners strictly want to block under any reasonable beliefs they may have using their private information and common knowledge of stability. Even under monotonic payoffs, the incomplete-information stable set may be large - it depends finely on the market structure and the prior belief support. If the unknown workers' type function is a bijection, the stable sets with complete and incomplete information perfectly coincide (to include only positive assortative matchings). We show, using examples, that the robust approach can reach precise predictions even beyond the monotonic case.

Keywords: Matching markets, incomplete information, belief formation, pairwise stability.

JEL codes: C78, D91.

*Ph.D. candidate at Paris School of Economics (PSE) - 48 boulevard Jourdan, 75014 Paris, France (Office R6-41) - laure.goursat@psemail.eu

1 Introduction

Motivations: Incomplete asymmetric information and ban on monetary transfers

"Matching" problems study the formation of productive partnerships, with numerous empirical applications: marriage, labor, housing, college admissions, organ donation, and many more. In most real-life matching markets, there are no monetary transfers and there is incomplete information.

The ban on monetary transfers is often motivated by ethical or fairness reasons. The ban can mean two different things in practice. Prices are either absent. In developed economies, it is forbidden to buy or sell organs, and brides do not offer dowry when they get married. Or prices can be set in advance. In social housing, rents are fixed. In civil servants' job markets (for the assignment of teachers to schools, of doctors to hospitals), wages are functions of qualification and experience and cannot be negotiated. In college admissions, students cannot bargain on tuition fees.

Incomplete information (on one's own preferences and payoffs) often arises due to a lack of experience. Formations of the kind of partnerships we have in mind are rare events, and the same partnerships are maintained over long time horizons. At the extreme, some people get married, go to college, receive an organ only once, and occupy only a few different houses or job positions in their whole lives. The standard learning premise from game theory that players can repeatedly play the same game until they reach a perfect understanding of the game structure, in particular of the payoffs, cannot hold. This implies that agents should be able to assess the outcome of a given unrealized match ex-ante. Before effectively being married / hired, they should be able to tell how happy they would be in each match. But intuition suggests guessing about one's happiness ex-ante is a complex multi-dimensional theoretical problem. Reciprocally, once two agents are matched, the matching experience enables them to collect data on their partner and to directly observe their own satisfaction from the match.

In two-sided markets where the two sides are not exactly of the same nature, the incomplete information often has an asymmetric structure. Typically, one side is informed, and the other side is uninformed or only partially informed about the characteristics of agents on the other side. In civil servants' job markets, the hospitals / schools know reasonably well about the teachers; they have centralized access to data about their skills and performance, whereas doctors / teachers have little clue about the working experience in each unit before starting there. In school choice, colleges receive exhaustive applications with vitae, certificates, and letters, whereas students cannot really guess how successful or satisfied they would be before effectively engaging in the educational track.

Research question: Stability and stable matchings

Most of the theoretical matching literature (both with transferable utilities (TU) and with non-transferable utilities (NTU)) assumes perfect information.¹ This is for tractability, and as emphasized above, is quite unrealistic.

The central solution concept there is pairwise stability.² A matching is pairwise stable if no matched agent would prefer to be single (individual rationality) and if there is no pair of two matched agents who would both prefer to rematch together rather than stay with their current partners (no blocking pairs).

Pairwise stability has a strong descriptive value as the long-run outcome in both decentralized and centralized markets. For decentralized markets, Roth and Vande Vate (1990) [16] have shown that the dynamic process of rematchings converges to stable matchings. In centralized markets, Roth (1984) [14], using data from markets for doctors in the US, have found that regions that used stable mechanisms turned out to be successful while other mechanisms turned out to be heavily criticized by the public / candidate blocking pairs and to break down.

In a cornerstone paper, Gale and Shapley (1962) [7] have shown that with complete information,

¹See Roth and Sotomayor (1990) [15] for a review of this literature.

²The focus on pairwise stability with complete information is without loss of generality: the sets of pairwise stable matchings and the core matchings (matchings that are immune to blocking by coalitions of size larger than a pair) always coincide.

there always exists a stable matching. In the constructive proof for existence, they introduce the stable Deferred Acceptance mechanism.

In this paper, we are interested in pairwise stability³ with incomplete asymmetric information and transferable utilities. We face a major conceptual challenge: the definition of stability with complete information does not immediately translate to the incomplete information setting. We need to define how uninformed agents form beliefs; there could be as many stability notions as belief formation processes. Hence a two-fold research question: With incomplete information - What does it mean for a matching to be pairwise stable? What are the pairwise stable matchings?

The robust approach in the TU case

Liu, Mailath, Postlewaite, and Samuelson (2014) [11] (henceforth "LMPS") have developed a highly interesting approach for markets with TU. Their motivating example is a job market where workers have common knowledge about the firms' productive types, but a firm does not observe the workers' productive types (except for the type of the worker she is currently matched to). In their model, firms make inference about the unknown workers' types based on common knowledge that the observed matching is stable.⁴ A firm is considered willing to block if she strictly wants to block under any belief she may have about the workers' types, given this knowledge. Their main result⁵ says that: When the workers' values and the surpluses are strictly increasing and super-modular in types (sufficient condition), the set of incomplete-information stable outcomes⁶ coincide with the set of complete-information stable outcomes (hence efficiency).⁷

Summary: Model and results

In this paper, we apply a similar approach to incomplete-information stability, but forbidding monetary transfers. We adopt the setting of LMPS except that firms make no side payments to workers when they match. We use their definition of stability and follow the same agenda.

We find that an incomplete-information stable matching always exists. Under monotonic payoffs, some inference happens, and the stable set can be quite small (it may include only positive assortative outcomes). How small depends on the support structure (whether the worker type function is bijective, injective, or surjective).

Interestingly, the mechanic behind is quite different from LMPS. In our NTU setting, the absent payments cannot play the signaling role they have in LMPS. A worker wants to block if and only if the firm has a higher type than his current partner, so that blocking proposals do not signal anything about the workers' types. Our inductive proof is simpler in that respect.

We also argue, through examples, that robust stability is able to give a precise prediction with more general (non-common) preferences. This again contrasts with LMPS, where robust stability has no bite once we relax the super-modularity assumption.

Outline of paper

The rest of the paper is structured as follows. Section §2 reviews the literature, in particular, the handful of papers that directly proceed from LMPS. Section §3 defines incomplete-information stability, through an illustrative example and a formal definition, translating LMPS' concept to a NTU market.

³The focus on pairwise stability with incomplete information is with loss of generality. Allowing coalitions of size larger than a pair could potentially dismiss more matchings than when we consider uniquely blocking pairs. We stick to pairwise stability to remain as objective as possible: we would like to avoid making arbitrary assumptions about what information is shared within coalitions). More fundamentally, we think that in the applications we have in mind, the obvious blocking coalition is the pair: doctors or teachers cannot communicate with their peers and decide to exchange partners.

⁴In the authors' view, common knowledge that a matching is stable arises from the repeated observation of this matching.

⁵PROPOSITION 3 in LMPS.

⁶The outcome is the matching in type. Payments at stable outcomes may still differ from the complete information benchmark.

⁷We treat in details the main example from LMPS in section §4.2.

Section §4 gives the main analytical result, under monotonic match values. It also argues for the value of this approach for different preference structures. Section §5 concludes.

2 Literature review

Liu, Mailath, Postlewaite, and Samuelson (2014)

We emphasize three key forces in the original paper that we would like to preserve in our own analysis.

The notion of incomplete-information stability is robust in the sense that the inference made by firms just restricts the support of the beliefs over workers' types. "Reasonable beliefs" are beliefs consistent with common knowledge of stability. The definition does not make arbitrary assumptions about how firms might select specific beliefs in the support.⁸ As a consequence, the results rely only on the structure of the economy and the hypothesis of stability. In addition, little restrictions are made on workers' types function (except that it has discrete values).

The concepts and proofs are very elegant. The paper defines an iterative belief formation reminiscent of rationalizability (Bernheim (1984) [2] and Pearce (1984) [12]) in games, where knowledge of stability at increasing order dismisses wrong beliefs. In the proof for the main result (PROPOSITION 3), it is very clear how the super-modularity in payoffs helps the inference. With super-modular payoffs, high types of workers can afford to accept lower blocking payment proposals. Because payments are continuous and types are discrete, there always exists a level of blocking payment that enables the firm to separate the worst case belief about a worker's type from the more favorable beliefs, making the firm willing to block. The proof is by induction, starting from the lowest types.

Finally, the definition of blocking pair in LMPS is very conservative. To be considered willing to block, a firm needs to strictly benefit from blocking under any reasonable belief. Their favorite interpretation is that firms do not effectively maintain a set of admissible beliefs but rather hold a posterior distribution over the belief support. They still view firms as expected profit maximizers. The goal of using the distribution-free approach is simply to characterize the outcomes that are immune to prudent inference of the firms. This approach is expected to produce few blocks and therefore, a large incomplete-information stable set. But it reaches a surprisingly precise prediction. In the super-modular case, the stable set is minimal (and the only belief reasonable at any order is the true belief).

On the weaker side, the analysis of the NTU case is really lacking. Even in the original example from LMPS of a job market, it can be argued that many job markets feature fixed wages, such as the markets for civil servants (accounting for more than 20% of the working population in many developed economies). And public economics provides many more examples of allocation problems where money is banned in practice.

In addition, the paper gives no clue about what could be the stable set with non super-modular payoffs.⁹ Our guess is that the stable set is very large, even maximal, because the inference by firms cannot even start. The super-modularity is really what makes blocking proposals at certain prices signals of the workers' types.¹⁰

Followers on LMPS

Several papers have extended, refined or questioned the robust approach in the TU case.

Liu (2020) [10] reformulates the model from LMPS. A matching function maps workers' types into a pair of a matching and a transfer scheme. Borrowing vocabulary from non-cooperative game theory,

⁸ Similarly to iterated elimination of never-best responses in non-cooperative games.

⁹ Or non sub-modular payoffs, as the paper also investigates this case. They get that under monotonic and sub-modular payoffs, every incomplete information stable matching outcome is negative assortative. But a stable outcome could be inefficient due to too many matched agents.

¹⁰ See section §4.2 for the argument.

the paper distinguishes between on-path beliefs (formed when based on knowledge that no one blocks the matching) and off-path beliefs (beliefs conditional on counterfactual pairwise deviations). This reformulation makes it easier to answer the classical question in the economics of uncertainty: How do agents compute posterior beliefs? The main refinement proposed in the paper requires that a firm's belief is updated using Bayes' rule from the prior belief conditional on what the firm observes and knows. The paper analyzes the differences between pairwise stability, the core, efficiency, competitive equilibrium in this setting.

Chen and Hu (2020) [5] generalize LMPS by allowing firms to have heterogeneous information about workers' types that do not necessarily derive from private observation of the current partner's type. A firm's information is described by a profile of partitions over possible type profiles of the workers. They show that in the more general setting, the robust approach becomes very permissive.

Pomatto (2022) [13] models the proposal and acceptance game and proposes a forward-induction logic. In case a player makes an offer, the offer is interpreted according to the highest degree of strategic sophistication that can be ascribed to its proponent. The main result is an equivalence between this notion and robust incomplete-information stability.

Bikhchandani (2017) - NTU and continuous workers' types

Bikhchandani (2017) [3] locate in a LMPS environment, except for two major differences: monetary transfers and a continuous support of possible workers' types ($W = [\underline{w}, \bar{w}] \subset \mathbb{R}$).

The main result is that even under monotonic payoffs, the set of stable matchings a la LMPS (or the set of "ex-ante stable matchings" in their vocabulary) is large (maximal). The problem is that the worst possible belief \underline{w} always qualifies as reasonable, and prevents any firm from blocking.

They conclude that "ex-ante stability has little predictive power". We find this conclusion a little unfair because of the continuous workers' types. This is a deviation from the original framework, and it fully drives the negative results. Actually, the robust approach in the TU case would reach a similar negative conclusion if it used continuous types rather than discrete types. In the proof in LMPS, it is obvious that firms need a discrete gap between the remuneration values of two workers facing the same type. The firm can then make a blocking proposal at a price in this gap to disentangle between the two workers' types.

We believe the assumption of continuity of workers' types is mostly a way¹¹ to motivate the alternative (non-robust) notion of "Bayesian stability" defined in the second part of the paper. A pair is considered to be a "Bayesian blocking pair" if the worker strictly increases his payoff in the block, and the firm has a higher expected payoff in the block, where the expected payoff is computed with the Bayes rule on the set of reasonable beliefs (in their vocabulary "admissible beliefs"). The stable set is again defined with iterations: for a matching μ , the probability distribution over workers' types is updated by eliminating types that would create Bayesian blocking pairs. Under monotonic payoffs, Bayesian stable matchings exist. Under even stronger conditions (super-modularity), stable matchings are assortative and efficient.

In many ways, the present paper is more faithful than Bikhchandani (2017) to LMPS. We show that assortativity and efficiency can be reached in imperfect information with discrete types and no transfers, putting less structure on how firms compute posterior than what is needed with continuous types. Overall, our work provides evidence that the ban on transfers does not threaten the precision of the prediction from the robust stability - the continuous support assumption does.

Incomplete information stability with NTU

A few papers propose different approaches to pairwise stability in incomplete information. Lazarova and Dimitrov (2013) [9] focus on the question of convergence: can we reach a complete-information stable outcome in a decentralized way when information is imperfect? They model the

¹¹Whether types in real life are discrete or continuous is a different question. It depends on the application one has in mind, in particular, on the size of the market.

interaction between dynamic blocking and learning behaviors. Another difference from LMPS is that the type defines the value to the partner but not the preferences, so that two agents with the same type may have different preferences.

Chakraborty, Citana, and Ostrovsky (2010) [4] model college admissions where colleges infer the values of students based on what they observe of the current matching and awareness of selection effects. They show that when colleges observe the full matching, stable mechanisms do not usually exist. A stable mechanism exists when colleges observe only their own matches and students have common preferences.

Goursat (2023) [8] studies a marriage market with incomplete symmetric information where agents infer their unknown preferences through a heuristic based on realized utilities in the current matching. The heuristic completely neglects the selection effect. In this sense, this is the polar case from LMPS where there is common knowledge of preferences and firms guess about workers' types based on how the workers select among the firms.

This paper also belongs to the literature on the core with incomplete information, initiated by Wilson (1978) [17] and surveyed in Forges, Mertens, and Vohra (2002) [6].

3 Model: Matching with incomplete information and no transfers

3.1 The market

We define a two-sided, agent-agent, one-to-one matching market, without monetary transfers, often summarized in the literature under the words "marriage market". To be consistent with our motivating example and with the LMPS benchmark, we rather consider the market as a (wage-regulated) job market.

There is a finite set of workers (firms) I (J). An individual worker (firm) is denoted $i \in I$ ($j \in J$). We use male (female) pronouns for workers (firms).

Each agent has a cardinal type, which summarizes his or her productive characteristics. The finite set of workers' (firms') possible types is denoted $W \subset \mathbb{R}$ ($F \subset \mathbb{R}$). A function $w : I \rightarrow W$ ($f : J \rightarrow F$) maps each worker's (firm's) index into a worker (firm) type.

A matching is a function $\mu : I \rightarrow J \cup \emptyset$, where $\mu(i) = j \iff \mu^{-1}(j) = i$. We allow workers to be unemployed ($\mu(i) = \emptyset$) and firms to hire no worker ($\mu^{-1}(j) = \emptyset$).

When a worker of type $w \in W$ and a firm of type $f \in F$ are matched, the worker (firm) receives a match value $v_{wf} \in \mathbb{R}$ ($\phi_{wf} \in \mathbb{R}$). These values summarize the net utility from the match (after the potential payment of a fixed wage from the firm to the worker). Because the market is without transfers, the payoffs and the values perfectly coincide.

If a worker of type $w \in W$ (a firm of type $f \in F$) is unmatched, he (she) receives a payoff that we normalize to zero $v_{w(\emptyset)} = 0$ ($\phi_{w(\emptyset)f} = 0$).

This defines a matching game. An outcome (μ, w, f) specifies realized types (w, f) and matching μ .

3.2 The information

Firm's indices and the firm type function f are common knowledge. The current matching μ in indices is also common knowledge. Workers' indices are common knowledge, but the worker type function w is unknown, with two exceptions: each worker $i \in I$ knows his own type $w(i)$, and each firm j observes the type of her own partner $w(\mu^{-1}(j))$. This models situations where the workers observe firms' types while the firms do not observe workers' types except for the worker who currently works in the firm.

The worker type function w is drawn from some distribution with support $\Omega \subset W^I$. The support will play a role in the analysis, while the distribution will not - we leave the distribution unspecified.

The value functions v and ϕ are common knowledge. The consequence is that a matched agent observes his or her own realized payoff, but does not observe the payoffs of other matched agents (because the

agent misses the worker's type argument in the value function). An agent always knows his or her payoff from the outside option $v_{wf(\emptyset)} = \phi_{w(\emptyset)f} = 0$.

The next figure summarizes the information structure from the point of view of a firm j at the local level of two pairs (j and her current partner i , and a potential targeted worker i' and his partner j'):¹²

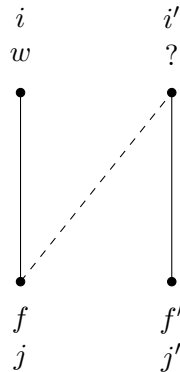


FIGURE 3.1: Local interim information from the point of view of firm j

We add a last element to the firm's knowledge: common knowledge of stability. By definition, a stable matching is a steady state of the market: once the matching is established, it persists unchanged over time. Thus, if the current matching is stable, agents repeatedly observe the same matching. After many time periods, they reach common knowledge that the observed matching is indeed stable.

3.3 An example

Before defining robust incomplete-information stability, we first work on an example to illustrate the simple mechanics behind the concept.

We set a market with $|I| = |J| = 3$ workers and firms, and possible types $W = F = \{1, 2, 3\}$. The match values are additive and symmetric $\forall w, f \in \{1, 2, 3\} : v_{wf} = v_{fw} = w + f$. All types of firms are realized $f(J) = F = \{1, 2, 3\}$. The prior belief support Ω on the worker type function is made of any bijection from J to F . Suppose that the current matching μ is as on the next figure:

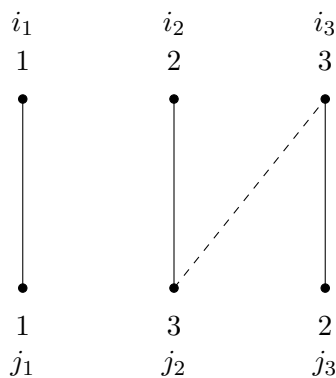


FIGURE 3.2: Current matching and candidate blocking pair

The question is: should we consider this outcome as stable? In particular, is the pair (j_2, i_3) (complete-information blocking pair) blocking under incomplete information?

Worker i_3 has perfect information. He knows that rematching with firm j_2 would increase his payoff by 1, so he wants to block.

Firm j_2 does not observe worker i_3 's type. She just observes that her current partner is of type 2, and

¹²All figures in the paper share the same design: we write all workers (firms) indices and types below (above), we draw plain lines for the current matching and dashed lines for candidate blocking pairs.

because the support Ω includes only bijections, she therefore knows that worker i_3 is of type either 3 (the true type $w(i_3) = 3$) or 1 (a wrong type $w'(i_3) = 1$). If worker i_3 is of type 1 (3), firm j_2 would increase (decrease) her payoff by blocking. It is unclear whether firm j_2 effectively blocks or not. We can go deeper into the wrong scenario. Under belief w' , the outcome would be as follows:

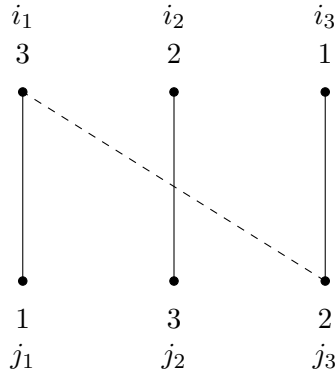


FIGURE 3.3: Counter-factual outcome under belief $w'(i_3) = 1$.

In this outcome, (i_1, j_3) would be a blocking pair. Worker i_1 would observe that firm j_3 is of type 2, higher than 1. Firm j_3 would believe that worker i_1 is of type 2 or 3 and would strictly benefit from blocking under any of those beliefs.

The existence of a blocking pair (i_1, j_3) under belief w' would contradict firm j_2 's knowledge of stability. Firm j_2 should be able to dismiss this belief w' and to focus on the true belief $w(i_3) = 1$, under which she blocks with i_3 . In total, (i_3, j_2) should be considered as a blocking pair and the outcome in FIGURE 2 is unstable.

3.4 Complete information stability

We first recall the definition of pairwise stability with complete information:

Definition 1. [Gale and Shapley (1962) [7]]

An outcome (μ, w, f) is complete-information stable if:

(IR) Individual rationality:

$$\begin{aligned} \forall i \in I, v_{w(i)f(\mu(i))} &\geq 0 \\ \forall j \in J, \phi_{w(\mu^{-1}(j))f(j)} &\geq 0 \end{aligned}$$

(NBP) No blocking pairs:

$$\nexists (i, j) \in I \times J \text{ s.t.: } \begin{cases} v_{w(i)f(j)} > v_{w(i)f(\mu(i))} \\ \text{and} \\ \phi_{w(i)f(j)} > \phi_{w(\mu^{-1}(j))f(j)} \end{cases}$$

An outcome is individually rational if any matched agent receives a weakly positive value, hence higher than the outside option of remaining unmatched. Because matched agents always observe their payoffs and know about the zero payoffs when unmatched, the individual rationality condition is the same for complete and incomplete information. An outcome has no blocking pairs if we cannot find a pair of a worker and a firm who would both prefer to be matched together rather than to stay with their current partners.

Gale and Shapley (1962) [7] have shown that a complete-information stable outcome always exists. They have also provided a method to reach a stable outcome with the famous Deferred Acceptance mechanism.

3.5 Robust incomplete information stability

We define pairwise stability when the worker type function is unknown. We recall what a firm knows: the current matching, the types of all firms, the prior distribution of workers' types, the realized type of her current worker, and common knowledge of stability.

We adapt the definition of incomplete-information stability from LMPS to our no transfer setting:

Definition 2. [Liu, Mailath, Postlewaite, Samuelson (2014) [11]]

(i) Fix a non-empty set of individually rational matching outcomes Σ . An outcome $(\mu, \mathbf{w}, \mathbf{f}) \in \Sigma$ is Σ -blocked if:

$$\begin{aligned} & \exists (i, j) \in I \times J \text{ s.t. :} \\ & (1) v_{\mathbf{w}(i)\mathbf{f}(j)} > v_{\mathbf{w}(i)\mathbf{f}(\mu(i))} \\ & (2) \phi_{\mathbf{w}'(i)\mathbf{f}(j)} > \phi_{\mathbf{w}'(\mu^{-1}(j))\mathbf{f}(j)}, \forall \mathbf{w}' \in \Omega \text{ s.t. :} \\ & \quad (a) (\mu, \mathbf{w}, \mathbf{f}) \in \Sigma \\ & \quad (b) \mathbf{w}'(\mu^{-1}(j)) = \mathbf{w}(\mu^{-1}(j)) \\ & \quad (c) v_{\mathbf{w}'(i)\mathbf{f}(j)} > v_{\mathbf{w}'(i)\mathbf{f}(\mu(i))} \end{aligned}$$

(ii) An outcome $(\mu, \mathbf{w}, \mathbf{f}) \in \Sigma$ is Σ -stable if it is not Σ -blocked.

(iii) Let Σ^0 be the set of all individually rational outcomes.

For $k \geq 1$, set: $\Sigma^k := \{(\mu, \mathbf{w}, \mathbf{f}) \in \Sigma^{k-1} : (\mu, \mathbf{w}, \mathbf{f}) \text{ is } \Sigma^{k-1}\text{-stable}\}$.

(iv) The set of incomplete-information stable outcomes is given by: $\Sigma^\infty := \bigcap_{k=1}^{\infty} \Sigma^k$.

Inequality (1) on line (i) requires that worker i strictly benefits from blocking with firm j . Inequality (2) requires that firm j expects to strictly benefit from blocking with worker i under any reasonable belief \mathbf{w}' that she might have about the workers' type function. To be reasonable, a belief must be in the support Ω and must satisfy conditions (a), (b) and (c). Condition (b) demands that the belief $\mathbf{w}'(\mu^{-1}(j))$ about the type of firm j 's partner is correct. This is a consistency condition saying that the firm uses her observation of her partner's type. Condition (c) is condition (1) rewritten for any reasonable belief about the targeted worker's type $\mathbf{w}'(i)$. This is a consistency condition asking that worker i blocks only if he has an incentive to do so. Through condition (c), the firm can draw inference about the type of the worker she faces just because this worker is willing to block with her, and not all types would be willing to do so. Condition (a) just says that the belief must be consistent with outcomes already in Σ . That is to say, the Σ -blocking test can only get rid of outcomes in Σ . This sets the foundations for the iterative definitions in the next lines.

Line (iii) defines a weakly decreasing sequence of sets of outcomes. Line (iv) defines the limit of the sequence.¹³

To illustrate the iterative argument, we make a formal treatment of the example from FIGURE 2. Belief \mathbf{w}' is a reasonable belief at the first order (it does not contradict the premise that μ is individually rational), but it is not reasonable at second order (it contradicts the premise that there are no blocking pairs in μ). So (i_3, j_2) is a Σ^1 blocking pair and $(\mu, \mathbf{w}, \mathbf{f}) \notin \Sigma^2, \notin \Sigma^\infty$.

Proposition 1.

(i) The set of incomplete-information stable outcomes is a superset of the set of complete-information stable outcomes.

(ii) For each type assignment (\mathbf{w}, \mathbf{f}) , there exists an incomplete-information stable matching μ .

The proof for (i) shows that the true belief \mathbf{w} will always be kept as reasonable in the iterative inference. Statement (ii) uses the existence of a stable matching in complete information (Gale and Shapley (1962) [7]).

¹³The sets $\Sigma^k, k \in \mathbb{N}, \Sigma^\infty$ depend on the support of beliefs Ω , so that the explicit notation should be $\Sigma^k(\Omega), k \in \mathbb{N}, \Sigma^\infty(\Omega)$.

4 Results: Robust stable matchings

4.1 Common preferences

The notion of incomplete-information stability from definition 2. uses a conservative notion of blocking pairs, likely to result in few blocks. The definition is quite permissive, and in the general case, it is expected to produce a large set of incomplete-information stable outcomes.

Yet, under natural assumptions on the match values and the support of the prior distribution, the prediction becomes surprisingly precise.

Assumptions

We list three assumptions on the market structure. When we give the main results, we will make clear why those assumptions are needed.

Assumption 1.

(M) $v(\phi)$ is strictly increasing in $f(w)$.

(SM) The match surplus $v + \phi$ is super-modular in w and f .

With assumption (M), match values are strictly increasing in the partner's type.¹⁴ This defines common ordinal preferences: all agents from a given side agree on the ranking of agents on the other side. All workers (firms) prefer to be matched with firms (workers) of higher cardinal type. Becker (1973) [1] has shown that under this assumption, there is a unique stable matching with complete information: the positive assortative matching, where the k^{th} most preferred worker is matched to the k^{th} most preferred firm, for any k .

Assumption (SM) sets super-modular match surpluses: the increase in the match surplus of a pair from changing the worker to a higher type is larger the higher the type of the firm (and conversely).

We note that with monotonic preferences, condition (c) from definition 2. never binds. Whether a blocking is incentivized for a worker only depends on the firms' types so that a blocking proposal does not reveal anything about the worker's type. In FIGURE 1, i' wants to block with j if and only if $f > f'$. Whether $f > f'$ is observed by j , so that there is nothing that f can learn.¹⁵

Assumption 2.

(I) Ω is the set of injections from I to W .

(S) Ω is the set of surjections from I to W .

(B) Ω is the set of bijections from I to W .

Assumption 2. restricts possible worker types functions in 3 possible ways. (I) imposes that each worker has a different type. (S) imposes that all the possible types have at least one representative worker with this type. (B) is equivalent to (I) and (S): firms perfectly know the set of (non-replicable) worker types they face. They just do not know which worker has which type, and consider that any permutation can happen.

Assumption 3.

(N') $|f(J)| = |J|$

(N) $|I| \leq |J + 1|$.

(N') says that different firms have different types. (N) is an assumption on the congestion on the market: there is at most one more worker than firms. We denote $\Delta := |J| - |I| \geq -1$ the difference in the number of firms vs workers.

¹⁴This is just a convention. With strictly decreasing payoffs, we get the same results.

¹⁵As a consequence, cardinality does not play any role in the results. We could produce the same model and results with ordinal preferences and types. We stick to the cardinal framework, though, to ease the comparison with LMPS.

Main results

To state the main results and to write the proof, we borrow the definition of assortativity at various orders from LMPS. To this end, we order the set of possible types: $W = \{w^1, \dots, w^K\}$, $F = \{f^1, \dots, f^L\}$, with types increasing in the exponents $w^1 < \dots < w^K$, $f^1 < \dots < f^L$. To deal with unmatched agents, we also introduce the notation $w(\emptyset) = w^0$, $f(\emptyset) = f^0$.

Definition 3.

(i) [Liu, Mailath, Postlewaite, Samuelson (2014) [11]]

A matching outcome $(\mu, \mathbf{w}, \mathbf{f})$ is k^{th} -order worker assortative ($k \geq 1$) if:

$$\forall w > w^k, \forall i, i' \in I | \mathbf{w}(i), \mathbf{w}(i') \in \{w^1, \dots, w^k, w\} : \mathbf{w}(i) > \mathbf{w}(i') \implies \mathbf{f}(\mu(i)) \geq \mathbf{f}(\mu(i'))$$

(ii) A matching outcome $(\mu, \mathbf{w}, \mathbf{f})$ is positive assortative if:

$$\begin{aligned} \forall i, i' \in I : \mathbf{w}(i) > \mathbf{w}(i') &\implies \mathbf{f}(\mu(i)) \geq \mathbf{f}(\mu(i')) \\ \forall j, j' \in J : \mathbf{f}(j) > \mathbf{f}(j') &\implies \mathbf{w}(\mu^{-1}(j)) \geq \mathbf{w}(\mu^{-1}(j')) \end{aligned}$$

(iii) A matching outcome $(\mu, \mathbf{w}, \mathbf{f})$ is maximal if only $|\Delta|$ agents are unmatched.

k^{th} -order worker-assortativity requires both that the k lowest worker types w^1, \dots, w^k are matched with firms assortatively, and that any worker with a higher type $w > w^k$ is matched with a firm of (weakly) higher type. k^{th} -order worker assortativity is defined in terms of the grand set of all ex-ante possible worker types W , not with respect to the ex-post realized types. For example, if no worker is effectively of type w^1 , 1st-order worker assortativity is trivially satisfied. 1st-order worker assortativity just means that workers of types w^1 , if any, are matched to the lowest types of firms (or unmatched). Assortativity as in (ii) is the usual positive assortativity as in Becker (1973) [1].

Proposition 2.

Set an incomplete information matching market with assumptions (M), (N) and (N').

Then:

- (1) Under assumption (B), incomplete-information stable outcomes are the assortative outcomes.
- (2) Under assumption (I), incomplete-information stable outcomes are the k^{th} -order worker assortative outcomes, where $k + 1$ is the first omitted worker type in the realized types.
- (3) Under assumption (S), the incomplete-information stable outcomes are the maximal outcomes.

This proposition shows that under monotonic values, the robust approach can lead to maximally small / large stable set. The size of the set depends finely on the assumption about the worker type function.

The proof of statement (1) is by contradiction and induction.¹⁶

In the initial step, we show that if an outcome is not assortative for the lowest worker type w^1 , the usual blocking pair from complete information (between the worker (firm) matched with the lowest type of (firm) worker) is also blocking with incomplete information. Indeed, the firm in this pair, observes her current partner's type w^1 . Due to the bijection assumption, she then knows that the target partner she faces is of strictly higher type $w > w^1$. Although this firm holds many reasonable beliefs, she wants to block under any of those beliefs. So only 1st-order worker assortative outcomes can survive this kind of block. We illustrate this argument below:

¹⁶With a preliminary step dealing with unmatched agents: showing that stable outcomes are maximal, and that the unmatched agents are the lowest types.

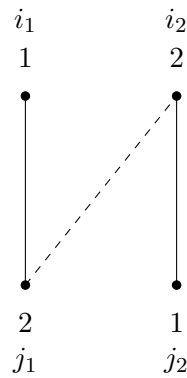


FIGURE 3.4: (Local) non 1st-order worker assortative outcome
 If $W = \{1, 2, 3, 4\}$ and (B) hold, j_1 's reasonable beliefs w' are s.t.: $w'(i_2) \in \{2, 3, 4\}$
 (i_2, j_1) is a Σ^0 blocking pair

At the inductive step k , only matchings that are worker assortative at k^{th} -order remain as candidate stable outcomes. In practice, this means that agents have common knowledge that the observed matching must satisfy k^{th} -order worker assortativity. So whenever a firm observes a firm type above k , she knows the partner of this firm is also a type above k . This enables to build blocking pairs in any outcome that would not be $(k + 1)^{\text{th}}$ -order worker assortative. We illustrate the argument below, using again FIGURE 2:

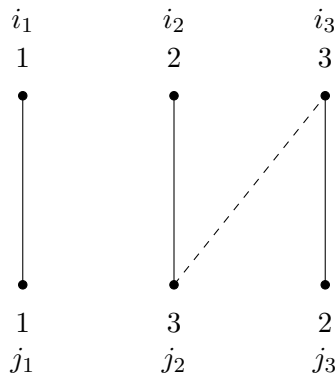


FIGURE 3.5: (Local) 1st-order but not 2nd-order worker assortative outcome
 If $W = \{1, 2, 3, 4\}$ and (B) hold, j_2 's reasonable beliefs w' are s.t.: $w'(i_3) \in \{3, 4\} > 2$
 (i_3, j_2) is a Σ^1 blocking pair.

Statement (1) can be interpreted as an ex-post validation of the mainline matching literature. The complete information assumption can be seen as a technical shortcut to the environment with incomplete information, leading to the same market outcomes. But whether the (B) assumption holds in practice remains unclear.

For statement (2), we assume only (I) and not the stronger (B). Some possible workers' types may not be realized, and firms know this. Starting from the lowest possible types, as long as all possible types are realized (so that the injection behaves locally as a bijection), the induction from the (B) case works the same. However, as soon as one possible type w is not realized, the possible belief of candidate blocking firms includes w , under which there is no incentive to block. The induction stops at this point. We illustrate this argument below:

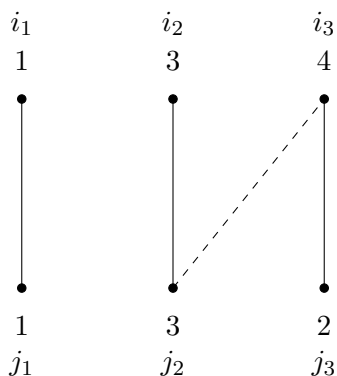


FIGURE 3.6: (Local) 1st-order and 2nd-order worker assortative outcome
 If $W = \{1, 2, 3, 4\}$ and (I) hold, j_2 's reasonable beliefs w' are s.t.: $w'(i_3) \in \{2, 4\}$
 (i_3, j_2) is not a Σ^1 blocking pair.

The prediction from statement (2) is that assortativity applies on the lowest types first. This may match stylized facts about real-life markets, where we observe that low-skilled workers are matched with low-status job positions, but where the success of the intermediate to high-skilled workers seems more variable. This paper proposes an explanation: firms fear that targeted workers could be of some unrealized low types.

For statement (3), we assume only (S) and not the stronger (B). It can be that two different workers have the same realized type. This prevents the very first step of the induction. The possible beliefs of a firm matched to a worker with the lowest realized type w^1 include the belief that a targeted worker is also of type w^1 , hence no strict incentive to block.¹⁷
 The next figure illustrates this point:

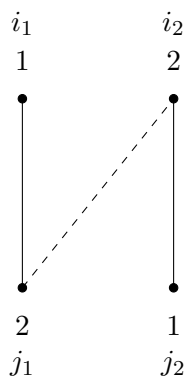


FIGURE 3.7: (Local) non 1st-order worker assortative outcome
 If $W = \{1, 2, 3, 4\}$ and (S) hold, j_1 's reasonable beliefs w' are s.t.: $w'(i_2) \in \{1, 2, 3, 4\}$
 (i_2, j_1) is not a Σ^0 blocking pair.

Statement (3) is a negative result. Incomplete-information stability only restricts the matching of the lowest types, who are necessarily the unmatched agents. The issue is that firms may believe the whole market is made of the same very low type. The theoretical implication is that the complete-information framework cannot be seen as a reduced-form model of the incomplete-information framework. The practical implication is that incomplete information disrupts the market outcome a lot.

Statements (2) and (3) bridge the gap with Bikhchandani (2017) [3], working with continuous workers' types. The interval of workers's types $W = [\underline{w}, \bar{w}]$ in Bikhchandani (2017) can be seen as the

¹⁷A slight modification of the definition of a blocking pair could solve this issue and make the initial step operational: a firm would be considered as willing to block if it strictly wants to block under all reasonable beliefs except one, under which she is indifferent. However, this definition would not restore the inductive step in the proof, as w^1 would still survive as a reasonable belief of any firm about a targeted worker.

limit case of (I) where an infinity of possible types are not realized, or the limit case of (S) where any type may be replicated an infinite number of times. In this sense, Bikhchandani (2017) locates in a large market environment.

Necessity of assumptions on payoffs and market structure

We explain in details why we need assumptions (M) , (N) and (N') .

Assumption (M) implies that all agents want to match with the highest types. Due to the asymmetric information structure, where workers perfectly observe firms' types, only firms with high types will succeed in matching with workers of high types. Firms are aware of this, so that they reach common knowledge that the matching must be assortative.

Assumption (N) says that different firms have different types. If this were not the case, the assortativity condition would leave some flexibility on the exact assignment of workers within groups of firms with similar types. So, a firm considering rematching with a targeted worker matched to a firm whose type has multiple replicas would be confused. She would hold multiple reasonable beliefs about the target's type. The worst of those beliefs (still not contradicting assortativity) would give an incentive to block. We illustrate this point below:

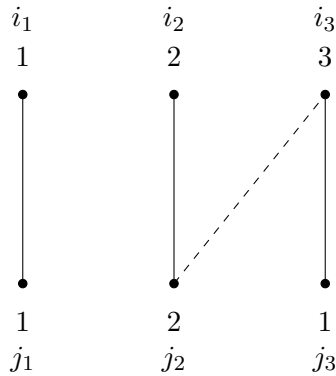


FIGURE 3.8: (Local) 1st-order and not 2nd-order worker assortative outcome - (N) does not hold
If $W = \{1, 2, 3, 4\}$ and (B) holds, j reasonable beliefs w' are s.t.: $w'(i_3) \in \{1, 3, 4\}$
 (i_3, j_2) is not a Σ^1 blocking pair.

Assumption (N') demands that there are at most one more worker than firms. This is a technical condition to ensure that the preliminary part of the proof (dealing with unmatched agents) works. If there are two more workers than firms, at least two workers remain unmatched. Then we face a similar issue to the one in FIGURE 8 (where we can replace the types 1 for firms by \emptyset). The candidate blocking firm j_2 pools the two unmatched workers. Reasonable beliefs about worker i_3 's type include the worst reasonable belief $w'(i_3) = 1$.

If there are significantly more firms than workers, many firms remain unmatched, but this is not an issue. Because workers have perfect information, they do not need to make inference about firms' types based on the firms' current partners' types.

Efficiency

With yet a little more structure on match values, we are able to comment on the efficiency of the stable outcomes, where an outcome is efficient if it maximizes the sum of all match values.

Corollary 1.

Set an incomplete-information matching market with assumptions (M) , (SM) , (N) , (N') , and (B) .
The incomplete-information stable outcomes coincide with the efficient outcomes.

We add assumption (SM) to the set of assumptions from proposition 2: match surpluses are super-modular. Having as few unmatched agents as possible and assortativity for the matched agents (as predicted by proposition 2) is what maximizes the sum of the surpluses.

Under assumptions (I) or (S) , incomplete-information matchings are not necessarily efficient. The market designer could consider taking action to restore efficiency. The straightforward way would be to give more information to agents. In the (I) case, revealing only which types are not realized on the market would solve the issue on FIGURE 6. Under assumption (S) though, giving information about which types are realized several times would not solve the issue in FIGURE 7 - only full information would.

4.2 Comparison with LMPS

To illustrate the differences between our work and robust stability in its original model, we take an example from LMPS.

In the TU case, an outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ also includes a *payment scheme* \mathbf{p} associated with the matching μ . The payment scheme specifies a payment from a firm to the worker $p_{i\mu(i)} \in \mathbb{R}$ for each $i \in I$, $p_{\mu^{-1}(j)j} \in \mathbb{R}$ for each $j \in J$.

Suppose that the current outcome is as follows:

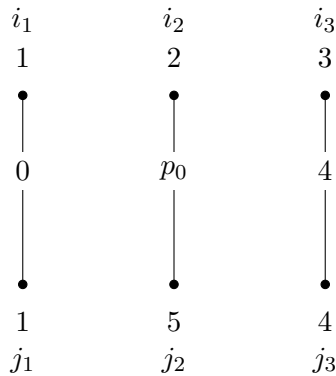


FIGURE 3.9: Candidate stable outcome in the TU case

Match values are given by: $v_{wf} = \phi_{wf} = w \times f$, hence symmetric, strictly increasing in the agent's and the partner's types. The values and the match surpluses are super-modular in types. Because of the transfers, payoffs aggregate the values and the payments. For example, worker i_3 (firm j_3) currently gets as a payoff: $4 \times 3 + 4 = 16$ ($4 \times 3 - 4 = 8$).

We do not specify the payment p_0 because different blocking pairs arise depending on p_0 . For the first case, we consider (as in LMPS) a payment reasonably close to 0: $p_0 = -2$ and the candidate blocking pair (i_3, j_2) at price p :

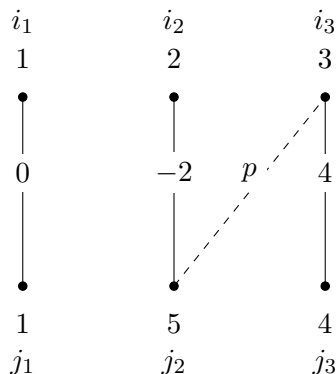


FIGURE 3.10: Outcome from FIGURE 9 and candidate blocking pair with $p_0 = -2$

Worker i_3 by rematching with firm j_2 would get $15 + p$. So he wants to block iff: $15 + p > 16 \iff p > 1$. What firm j_2 gets by rematching with worker i_3 depends on the type of worker i_3 and the price p . For example, if i_3 is of type 2, firm j_2 wants to block iff: $10 - p > 12 \iff p < -2$. Whether different types of workers (other than type 3) want to block at each possible price matters for firm j_2 's inference. We summarize all willingnesses to block of the pair (i_3, j_2) on a line for possible blocking prices below:

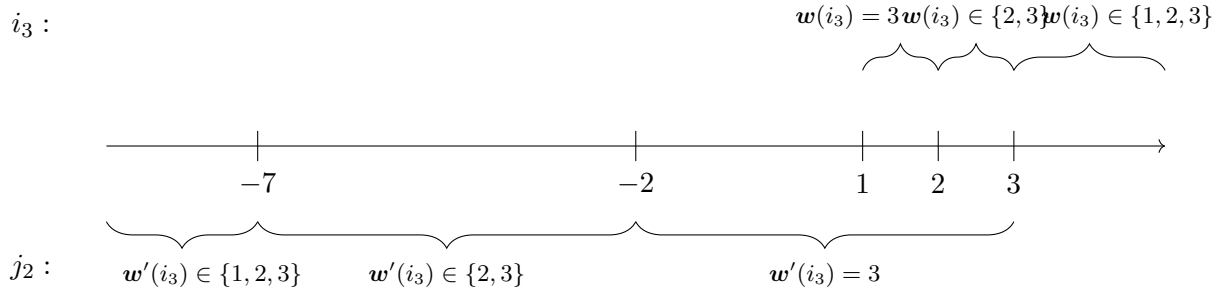


FIGURE 3.11: Willingnesses to block in pair (i_3, j_2) as a function of blocking price p

The blocking price $p = 1 + \epsilon$, $\epsilon > 0$ small, will make (1) Only a worker i_3 of type 3 willing to block (2) The firm j_2 willing to block with a worker i_3 of type 3. We have built a blocking pair. In practice, firm j_2 makes inference from a blocking proposal at a given price. If a worker i_3 makes a proposal at price $p = 1 + \epsilon$, this signals that he is of the best type 3.

For the second case, consider a lower payment $p_0 = -4$ and a candidate blocking pair (i_2, j_3) at price p :

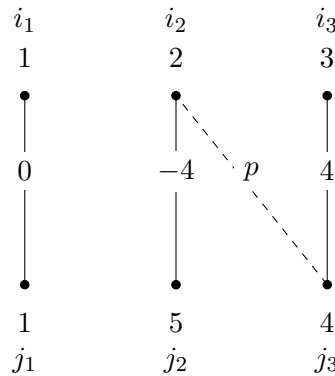


FIGURE 3.12: Outcome from FIGURE 9 and candidate blocking pair with $p_0 = -4$

The graph of willingnesses to block is the following:

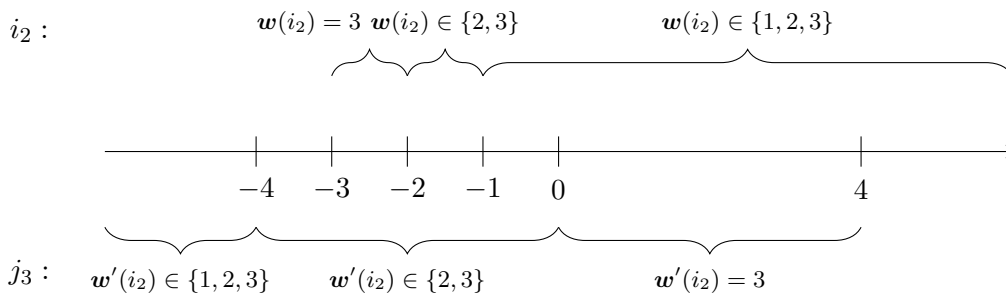


FIGURE 3.13: Willingnesses to block in pair (i_2, j_3) a function of blocking price p

The blocking price $p = -2 + \epsilon$, $\epsilon > 0$ small, will make (1) Worker i_2 of type 2 willing to block with firm j_3 (2) Firm j_3 willing to block with a worker i_3 of type at least 2. We have built a blocking pair.

In both cases, we find that the outcome from FIGURE 9 is unstable. LMPS show that this is general: in any matching that is not assortative, they can construct at least one blocking pair. Just as in complete information, there are two kinds of blocking pairs: if current payments are close to being neutral, the blocking pair gathers high types. If current payments are very much in favor of the high types, the blocking pair gathers low types. Supermodularity in values ensures that we always fall in one of the two cases.

The nature of the inference made by firms differs significantly between the TU and NTU cases. In LMPS, firms make inference from blocking proposals. A proposal at a given blocking price signals that a type is at least as good as a reference type. This restricts the set of reasonable beliefs by dismissing the worst beliefs, making firms willing to block. There always exists a blocking price that enables this kind of inference. This price (characterized by LEMMA 2 in LMPS) is such that it gives an ϵ benefit to the highest workers' types in the block. Super-modularity says that the benefit for a worker from rematching with a higher type is higher the higher the worker type. It therefore implies that the blocking price from LEMMA 2 in LMPS will be unattractive to any lower types. This suggests that cardinality of types and super-modularity of values is key in the NTU case, and that the analysis cannot be extended in a non super-modular world.

By contrast, the inference made by firms in the TU case has simpler, ordinal, and more general mechanics. It does not rely on any assumption about complementarity in the match values. From observing that a targeted worker is matched with a quite good type of firm, a firm infers that this worker is quite a good type as well, at least above her own partner's type. Otherwise, the current partner of the targeted worker would already have formed a blocking pair with another worker matched with a very low type of firm.

The inductive aspect of the proof is similar in both the TU and NTU cases.

4.3 Other preference structures

Another central difference with the TU case is that the robust approach in the NTU has bite even when preferences are not common (match values are not monotonic in the partner's type). We explain the general principle and propose examples.

When preferences are not common (but they are still common knowledge), blocking proposals disclose information about workers' types. In FIGURE 1, the fact that worker i' proposes to block with j reveals that i' is one of the workers who value type f more than type f' , hence a restriction on the set of reasonable beliefs about i' 's type.

Suppose, for instance, that preferences are cyclical and symmetric as follows:¹⁸

	w^1	w^2	w^3		f^1	f^2	f^3
1 st choice	f^1	f^2	f^3	1 st choice	w^1	w^2	w^3
2 nd choice	f^2	f^3	f^1	2 nd choice	w^2	w^3	w^1
3 rd choice	f^3	f^1	f^2	3 rd choice	w^3	w^1	w^2

We also assume that the belief support for workers types function is made of bijections from I to $W = \{w^1, w^2, w^3\}$.

Proposition 3.

There is a unique incomplete-information stable outcome, where the worker of type w^k is matched with the firm of type f^k , $k \in \{1, 2, 3\}$, so that every agent is matched with his or her favorite partner.

We propose a sketched proof in two steps. Suppose first that a local matching is such that:

¹⁸Without loss of generality, we work directly with ordinal preferences instead of cardinal utilities and with non-cardinal types.

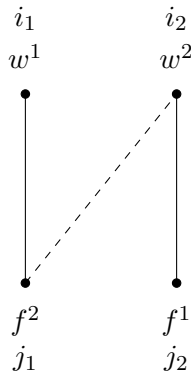


FIGURE 3.14: (Local) non 1st-order worker assortative outcome

Worker i_2 is matched with his worst choice, so he is willing to block with firm f^2 . Firm j_1 does not observe worker i_2 's type, but observes her current partner's type w^1 . So reasonable beliefs about i_2 's type only include $w'(i_2) \in \{w^2, w^3\}$. Under any of those beliefs, firm j_1 is willing to block with i_2 . So (j_1, f_2) is a Σ^0 blocking pair. Thus, in an incomplete-information outcome, we cannot have two pairs of "adjacent" types in the preference tables m^k, m^l, f^k, f^l that would be crossed matched ($\mu(m^k) = f^l, \mu(m^l) = f^k$).

Consider now an outcome as follows:

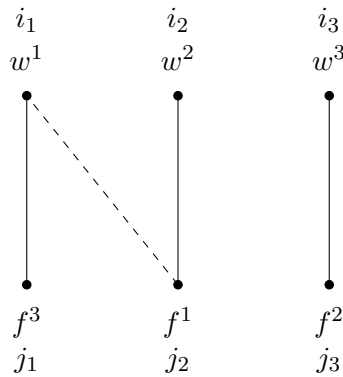


FIGURE 3.15: 1st-order but non 2nd-order worker assortative outcome

Worker i_1 of type w^1 is matched with his worst choice. He prefers to rematch with firm j_2 of type f^1 . Firm j_2 has reasonable beliefs $w'(i_1) \in \{w^1, w^3\}$. Firm j_2 is of type f^1 is not willing to block under belief $w'(i_1) = w^3$. So (i_1, j_2) is not a Σ^0 blocking pair. But belief $w'(i_1) = w^3$ implies $w'(i_3) = w^1$. Then the local outcome on the sub-market $\{i_2, i_3\} \times \{j_2, j_3\}$ would be exactly as in FIGURE 14. There would be a blocking pair, contradicting the premise of stability. So $w'(i_1) = w^3$ is not a reasonable belief. (i_1, j_2) is a Σ^1 blocking pair.

The conclusion is that at a stable outcome, there cannot be a cycle in the types matched as in FIGURE 15.

In total, this forbids any matching different from the one described in proposition 3..

5 Concluding comments

In this paper, we apply the robust stability approach to a market with non-transferable utilities. Compared to TU markets, NTU markets stand as an environment that is equally and increasingly relevant for economics. The present paper shows it is also richer in terms of predictions about stable outcomes.

LMPS had envisioned robust stability in the NTU case in these terms: "A marriage model is a special case of our model in which there are no transfers. (1) The absence of transfers would make it more difficult for an agent to convey information through a proposed block, precluding results analogous to LEMMA 2.¹⁹ Hence, we should expect (2) the set of incomplete-information outcomes to be relatively large, but (3) the structure of the analysis remains unchanged."

Our paper answers in the following way: statement (1) is true under monotonic match values only. Statement (2) is wrong in general, especially when the uniformed side's belief support includes only bijections. Statement (3) is wrong as, even when we find results that look similar to LMPS (in particular when the complete and incomplete-information stable sets coincide), the mechanics of the inference behind differ.

From an applied perspective, our results mean that strong inference by the uniformed side of the market about productive types on the other side may happen. The extent of this inference depends finely on the patterns of productive types. Targeted information provision may be enough to close the gap and reach the complete-information stable outcomes.

Deepening the analysis under non-common preferences is the primary direction for future research. Secondary directions include a generalization to many-to-one matching and a study of stability when deviations by coalitions larger than a pair are allowed.

Finally, beyond the "what?" question (what is stable?), we are interested in the "how?" question (how does the market get to what is stable?). The convergence question is a challenging one in an incomplete information environment where preferences are endogenous to the current matching.

¹⁹This is the lemma saying that there is always a blocking price that allows a high-type firm to disentangle between different workers' types.

References

- [1] G. S. Becker. A theory of marriage: Part i. *Journal of Political economy*, 81(4):813–846, 1973.
- [2] B. D. Bernheim. Rationalizable strategic behavior. *Econometrica: Journal of the Econometric Society*, pages 1007–1028, 1984.
- [3] S. Bikhchandani. Stability with one-sided incomplete information. *Journal of Economic Theory*, 168:372–399, 2017.
- [4] A. Chakraborty, A. Citanna, and M. Ostrovsky. Two-sided matching with interdependent values. *Journal of Economic Theory*, 145(1):85–105, 2010.
- [5] Y.-C. Chen and G. Hu. Learning by matching. *Theoretical Economics*, 15(1):29–56, 2020.
- [6] F. Forges, E. Minelli, and R. Vohra. Incentives and the core of an exchange economy: a survey. *Journal of Mathematical Economics*, 38(1-2):1–41, 2002.
- [7] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- [8] G. Laure. How can i know how much i like you? a heuristic approach to stability in matching. *mimeo*, 2023.
- [9] E. A. Lazarova and D. Dimitrov. Paths to stability in two-sided matching with uncertainty. *FEEM Working Paper*, 2013.
- [10] Q. Liu. Stability and bayesian consistency in two-sided markets. *American Economic Review*, 110(8):2625–2666, 2020.
- [11] Q. Liu, G. J. Mailath, A. Postlewaite, and L. Samuelson. Stable matching with incomplete information. *Econometrica*, 82(2):541–587, 2014.
- [12] D. G. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica: Journal of the Econometric Society*, pages 1029–1050, 1984.
- [13] L. Pomatto. Stable matching under forward-induction reasoning. *Theoretical Economics*, 17(4):1619–1649, 2022.
- [14] A. E. Roth. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of political Economy*, 92(6):991–1016, 1984.
- [15] A. E. Roth and M. A. O. Sotomayor. Two-sided matching, volume 18 of econometric society monographs, 1990.
- [16] A. E. Roth and J. H. Vande Vate. Random paths to stability in two-sided matching. *Econometrica: Journal of the Econometric Society*, pages 1475–1480, 1990.
- [17] R. Wilson. Information, efficiency, and the core of an economy. *Econometrica: Journal of the Econometric Society*, pages 807–816, 1978.

Appendix A Proofs

Proof of proposition 1.

Set a type assignment (w, f) . By Gale and Sapley (1962) [7], there exists a stable matching in complete information.

We denote the set complete information stable outcomes when the true workers types function is $w \in \Sigma(\{w\})$, and one element in this set $\mu \in \Sigma(\{w\})$.

Because μ is stable, it is individually rational, so: $(\mu, w, f) \in \Sigma^0$.

Set $k \in \mathbb{N}$, $(\mu, w, f) \in \Sigma^k$. By definition, w is a reasonable belief in Σ^k . Because μ is stable, it has no blocking pair under belief w . So for each pair, either the worker does not want to block ((1) is not satisfied) or the firm is not willing to block under at least one reasonable belief $w' = w$ ((2) is not satisfied). So $(\mu, w, f) \in \Sigma^{k+1}$.

By induction: $\forall k \in \mathbb{N}$, $(\mu, w, f) \in \Sigma^k$.

So $(\mu, w, f) \in \Sigma^\infty$. This proves (i)

μ is an incomplete-information stable matching for type assignment (w, f) .

We conclude, that, whatever the type assignment, there always exists a stable matching ((ii)).

Proof of proposition 2., statement (1)

We first prove how incomplete information stability restricts the outcome for (potentially) unmatched agents.

- CLAIM A - $(\mu, w, f) \in \Sigma^1 \implies$ Only $|\Delta|$ agents are unmatched.

PROOF - By contradiction, suppose that more than Δ agents are unmatched.

So there exists an unmatched worker and an unmatched firm. Denote them i and j respectively.

$v_{w(i)f(\emptyset)} = 0 < v_{w(i)f(j)}$, i wants to block with j .

By condition (b) in definition 2., at any reasonable belief w' : $w'(\emptyset) = w(\emptyset) = \emptyset$.

So $\forall w'$ reasonable: $\phi_{w'(\emptyset)f(j)} = 0 < \phi_{w'(i)f(j)}$, j wants to block with i .

(i, j) is a Σ^0 blocking pair.

(μ, w, f) is Σ^0 -blocked, $\notin \Sigma^1$.

More than Δ agents are unmatched $\implies (\mu, w, f) \notin \Sigma^1$.

BY contraposition, this proves CLAIM A.

- CLAIM B - $(\mu, w, f) \in \Sigma^1 \implies$ Unmatched agents are the lowest types.

PROOF

- Case $\Delta = -1$ (one more worker than firms).

Set an outcome $(\mu, w, f) \in \Sigma^1$.

By CLAIM A, there is exactly one unmatched worker. Denote this worker i .

Suppose he is not the lowest type $w(i) > w^1$.

By B, there is another (matched) worker who is the lowest type. Denote this worker i' : $w(i') = w^1$.

$v_{w(i)f(\emptyset)} < v_{w(i)f(\mu(i'))}$, i wants to block with $\mu(i')$.

By condition (b) in definition 2., at any reasonable belief w' : $w'(i') = w(i') = w^1$.

By (B), $w'(i) > w^1$.

So $\forall w'$ reasonable: $\phi_{w'(i')f(\mu(i'))} = 0 < \phi_{w'(i)f(j)}$, $\mu(i')$ wants to block with i .

$(i, \mu(i'))$ is a Σ^0 blocking pair.

(μ, w, f) is Σ^0 -blocked, $\notin \Sigma^1$.

Unmatched worker is not the lowest type $\implies (\mu, w, f) \notin \Sigma^1$.

By contraposition, this proves CLAIM B.

- Case $\Delta \geq 1$ (more workers than firms).

Set an outcome $(\mu, w, f) \in \Sigma^1$.

By CLAIM A, there are exactly Δ unmatched firms.

Suppose they are not the lowest types of firms.

$\exists(j, j') | \mu^{-1}(j) = \emptyset, \mu^{-1}(j') \neq \emptyset, f(j) > f(j')$.

By (M): $v_{w(\mu^{-1}(j'))f(\emptyset)} < v_{w(\mu^{-1}(j'))f(j)}$, $\mu^{-1}(j')$ wants to block with j .

By condition (b) in definition 2., at any reasonable belief w' : $w'(\mu(j)) = w(\emptyset) = \emptyset$.

So $\forall w'$ reasonable: $\phi_{w'(\emptyset)f(j)} = 0 < \phi_{w'(\mu^{-1}(j'))f(j)}$, j wants to block with $\mu^{-1}(j')$.

$(j, \mu^{-1}(j'))$ is a Σ^0 blocking pair.

(μ, w, f) is Σ^0 -blocked, $\notin \Sigma^1$.

Unmatched firms are not the lowest types $\implies (\mu, w, f) \notin \Sigma^1$.

By contraposition, this proves CLAIM B.

We now characterize how incomplete information stability restricts the outcome for matched agents.

- We first set $\Delta = 0$.

INDUCTION PROOF

Induction hypothesis: $H(k) : \Sigma^k(\Omega) \subseteq \{(\mu, \mathbf{w}, \mathbf{f}) \text{ } k^{\text{th}}\text{-order worker assortative}\}$.

- Initial step $k = 1$.

Suppose $(\mu, \mathbf{w}, \mathbf{f}) \in \Sigma^1$ is not 1st-order worker assortative.

Then, there exists $i, i' \in I$ s.t. $\begin{cases} \mathbf{w}(i) = w^1 < \mathbf{w}(i') \\ \mathbf{f}(\mu(i)) > \mathbf{f}(\mu(i')) \end{cases}$.

But then, by (M):

$v_{\mathbf{w}(i')\mathbf{f}(\mu(i))} < v_{\mathbf{w}(i')\mathbf{f}(\mu(i'))}$, so i' wants to block with $\mu(i)$.

By definition 2., condition (b), reasonable beliefs \mathbf{w}' of firm $\mu(i)$ are s.t.: $\mathbf{w}'(i) = \mathbf{w}(i) = w^1$.

By (B), $\mathbf{w}'(i) > w^1$.

So: $\forall \mathbf{w}'$ reasonable: $\phi_{\mathbf{w}'(i)\mathbf{f}(\mu(i))} < \phi_{\mathbf{w}'(i')\mathbf{f}(\mu(i))}$, so $\mu(i)$ wants to block with i' .

So $(i', \mu(i))$ is a Σ^0 blocking pair.

So $(\mu, \mathbf{w}, \mathbf{f}) \notin \Sigma^1$.

By contraposition: $(\mu, \mathbf{w}, \mathbf{f}) \in \Sigma^1 \implies$ 1st-order worker assortative.

So: $\Sigma^1 \subseteq \{(\mu, \mathbf{w}, \mathbf{f}) \text{ } 1^{\text{st}}\text{-order worker assortative}\} = H(1)$

- Inductive step

Set $k \geq 1$ and $H(k)$.

Set $(\mu, \mathbf{w}, \mathbf{f}) \in \Sigma^k$.

By $H(k)$, $(\mu, \mathbf{w}, \mathbf{f})$ is k^{th} -order worker assortative.

Suppose it is not $(k+1)^{\text{th}}$ -order worker assortative.

Then, there exists $i, i' \in I$ s.t. $\begin{cases} \mathbf{w}(i) = w^{k+1} < \mathbf{w}(i') \\ \mathbf{f}(\mu(i)) > \mathbf{f}(\mu(i')) \end{cases}$.

But then, by (M):

$v_{\mathbf{w}(i')\mathbf{f}(\mu(i))} < v_{\mathbf{w}(i')\mathbf{f}(\mu(i'))}$, i' wants to block with $\mu(i)$.

By definition 2., condition (a), reasonable beliefs are in Σ^k , so by $H(k)$, reasonable beliefs \mathbf{w}' include only outcomes that are k^{th} -order worker assortative.

In addition, by condition (b): $\mathbf{w}'(i) = \mathbf{w}(i) = w^{k+1}$.

By (B), $\mathbf{w}'(i) > w^{k+1}$.

So: $\forall \mathbf{w}'$ reasonable $\phi_{\mathbf{w}'(i)\mathbf{f}(\mu(i))} < \phi_{\mathbf{w}'(i')\mathbf{f}(\mu(i))}$, $\mu(i)$ wants to block with i' .

So $(i', \mu(i))$ is a Σ^k blocking pair.

So $(\mu, \mathbf{w}, \mathbf{f}) \notin \Sigma^{k+1}$.

By contraposition: $(\mu, \mathbf{w}, \mathbf{f}) \in \Sigma^{k+1} \implies (k+1)^{\text{th}}$ -order worker assortative.

So: $\Sigma^{k+1}(\Omega) \subseteq \{(\mu, \mathbf{w}, \mathbf{f}) \text{ } (k+1)^{\text{th}}\text{-order worker assortative}\}$.

We have proven $H(k+1)$.

- We conclude: $\forall k \in \{1, \dots, |I|\}$: $H(k)$ is true.

In particular: $H(|I|-1) : \Sigma^{|I|-1}(\Omega) \subseteq \{(\mu, \mathbf{w}, \mathbf{f}) \text{ } (|I|-1)^{\text{th}}\text{-order worker assortative}\}$.

Due to (B), $(|I|-1)^{\text{th}}$ -order worker assortativity is equivalent to assortativity.

Going to the limit we get that $\Sigma^\infty \subseteq \{(\mu, \mathbf{w}, \mathbf{f}) \text{ assortative}\}$.

- In the proof of proposition 1., we have proven that $\Sigma(\{\mathbf{w}\}) \subseteq \Sigma(\Omega)$, $\mathbf{w} \in \Omega$.

Under assumption (M), by Becker (1973) [1], complete-information stable outcomes are the assortative outcomes.

So assortative outcomes are also incomplete-information stable.

We conclude: $\Sigma^\infty = \{(\mu, \mathbf{w}, \mathbf{f}) \text{ assortative}\}$.

- If $\Delta = -1$, CLAIM B already sets the matching for type w^1 .

We do the same proof as above for the case $\Delta = 0$ on the submarket where we remove w^1 .

We get $|I| - 2^{\text{th}}$ -order worker assortativity.

Both results combined give assortativity.

- If $\Delta \geq 1$, CLAIM B fixes that there are Δ unmatched firms, and that they are the lowest types.

We do the same proof as above for the case $\Delta = 0$ on the submarket where we remove the Δ lowest type of firms.

We get $|I| - \Delta - 1^{\text{th}}$ -order worker assortativity.

Both results combined give assortativity.

Proof of proposition 2., statement (2)

We do a similar proof as for statement (1).

If all lowest possible types are realized, the injection behaves locally as a bijection.

The induction works until one of the possible worker types is not realized.

FIGURE 6 in the main shows on an example how the inductive step fails at the $k + 1$ stage, where k is the first unrealized worker type.

Proof of proposition 2., statement (3)

The kind of proof that we did for statement (1) does not work.

FIGURE 7 in the main shows how the initial step fails.

Proof of corollary 1.

By proposition 2., under (M) , (N) , (N') , and (B) , the set of incomplete information stable outcomes coincide with the set of positive assortative outcomes.

In addition, under supermodular values $((SM))$, the set of efficient outcomes is the set of positive assortative outcomes - The proof is omitted.

So the set of incomplete information stable outcomes coincide with the set of efficient outcomes.

Proof of proposition 3.

In the main.

Chapter 4

CAMPUS VISITS

Or Pre-Matching Information Acquisition

Francis BLOCH* and Laure GOURSAT† – December 2023 – [*Preliminary version*]

We thank Olivier COMPTE, Philippe JEHIÉL, Olivier TERCIEUX for helpful comments and conversations.

Laure thanks Ecole Nationale des Ponts et Chaussées (doctoral funding contract n°20/092) for financial support.

Abstract: This paper studies a college admission problem gathering heterogeneous students and colleges where students can endogenously acquire information on their own preferences. Students' preferences over colleges include a common component, which is common knowledge, and a private component, which is unknown ex-ante. Students can learn about the private components, before matching occurs through a standard Deferred Acceptance mechanism with common priorities. The question is: What information do students acquire, as a function of their priority rank? With unit constraint on learning and unit capacities at colleges, we find that the best student learns about one of the best colleges. Students with lower-priority learn about the best college among the ones where they are admitted for sure. The proof uncovers a novel additive property of the values of information. We discuss matching and welfare implications and ongoing generalizations.

Keywords: Matching markets, information acquisition, capacity constraint, value of information.

JEL codes: C78, D91.

*Chaired Professor at Paris School of Economics (PSE) and Professor at Paris 1 Panthéon-Sorbonne University - francis.bloch@psemail.eu.

†Ph.D. candidate at Paris School of Economics (PSE) - 48 boulevard Jourdan, 75014 Paris, France (Office R6-41) - laure.goursat@psemail.eu

1 Introduction

Motivations

Matching refers to the formation of productive partnerships. In most matching problems, by default, agents do not know about their own preferences over potential partners. Or rather, the amount of available information depends on the nature of the preferences. If the preferences include a common component, reflecting that some characteristics in a targeted partner are valued by any agent, objective information about this common component may be available on the market. In all likelihood, the market operator has gained knowledge about the common component in the course of operations and is willing to reveal the information to participants, in a concern for efficiency. Yet in most real-life matching markets, preferences also include a private component, reflecting that different agents may value the same targeted partners differently so that there is a notion of “fit” on the market. Then, it is very unlikely that information about the private components is freely available. Even a market operator with the best intentions would be unable to gather information about the private components because it involves individual tastes. Only each participant individually could collect information on the private aspect of her preferences. Even for the participant, collecting this information on private preferences could prove to be a hard task.

This is especially true in college admissions, where students start with bad or imprecise guesses on how they would be satisfied studying in each college. Our favorite interpretation of the bad guess is that it is hard to assess the outcome of a match before effectively experiencing it. Students usually attend college only once in their lives, so they cannot accumulate any experience. Fortunately, students often have the possibility to acquire information on their preferences before participating in the centralized allocation mechanism. Due to the centralization, applications are coordinated in time so that the application deadline is generally known much in advance. Students can use their time and resources to invest in information acquisition before the deadline. Indeed, we observe that they (jointly with their parents) do go to visit campuses, seek advice from former students, and make tailored searches online. The issue is that, in general, information acquisition is either costly (because it involves some effort) or capacity-constrained (because of the time constraint). In this context, we expect information acquisition to remain partial as well as heterogeneous among different students and colleges.

The empirical literature supports that intuition. Hoxby and Avery (2012) [12] report that most high-achieving, low-income high school students do not apply to selective colleges. They argue that these students’ low ambitions result from a lack of information in connection with a high cost of information acquisition; the students in question live in small districts where it is unlikely to encounter a teacher or schoolmate from an older cohort who attended a selective college. In the same line, Hastings and Weinstein (2008) [10] and Hoxby and Turner (2013) [11] in a respectively lab and field experiments show that providing students with information about colleges’ fees and outcomes raises the number of high-achieving low-income students’ applications to top colleges. Dustan, De Janvry, and Sadoulet (2017) [6] show that information on colleges’ values is by nature limited because of high uncertainty on the outcome at college through possible dropout.

The theoretical matching literature mainly¹ assumes exogenous information on preferences, most often perfect information (at least on ordinal preferences).² It, therefore, completely abstracts from the information acquisition problem. This is mainly for tractability reasons but at the expense of realism. Exogenous information models are unfit to describe real-life college admissions. They are bound to produce incorrect predictions in terms of matching and welfare, and biased design recommendations.

Research question

In this paper, we jointly model the information acquisition problem with the usual matching stage. The question is: How does the possibility for endogenous information acquisition shape the learning,

¹ Exceptions (mostly very recent papers) are discussed in section §2.

² This contrasts with Roth (1989) [18] early on identifying modeling of information as the priority next step in the analysis of matching markets.

the applications, the admissions, and the welfare?

We are particularly interested in two elements: the strategic interactions within students and the heterogeneity of students and colleges.

Incentives to acquire information for different students are intertwined. Indeed, the fact that a student acquires information modifies her expected application, expected match, and admission chances of all lower-priority agents. At equilibrium, students best-respond to (private) learning of higher-priority students.

In our model, both the students and the colleges are heterogeneous: students are ordered by their rank in priorities, and colleges are ex-ante ordered by their common component. Our main contribution to the literature on information acquisition is that we fully address the two-sided heterogeneity. Computing the learning equilibrium strategies amounts to answering the question: “Who learns about what?”. One nested question is whether, at equilibrium, there is some monotonicity in learning with higher-priority students applying to colleges with higher common attributes. The final aim should be to characterize how the matching and welfare are distorted with respect to the full information benchmark. In this perspective, our model is especially relevant to study (in)equity in college admissions.

Preview of model and results

The baseline model sets a stylized college admission problem with n students and colleges and unit capacities at colleges. Preferences are hybrid and additive: they include a common component and a private component. The common components are common knowledge. The private components are unknown ex-ante, but students can freely and perfectly observe the private component at one college. Priorities are common and common knowledge. The matching occurs through a standard Deferred Acceptance (DA) mechanism.

At Bayes-Nash equilibrium, we find that student 1 learns about college 1 or 2. Student $i \in \{1, \dots, n\}$ learns about college $i + 1$, and the value of this information equals the sum of the values from learning about all ex-ante better colleges. Student n may learn about any college and has zero value of information. There is thus monotonicity in learning in the sense that higher-priority agents explore ex-ante better colleges. Characterizing how this affects matching and welfare involves heavy ongoing combinatorial computations.

The more general model should include non-unit learning constraints and college capacities and compare learning under different mechanisms (general mechanism).

Outline of paper

The rest of the paper is structured as follows. Section §2 reviews the related literature. Section §3 sets and motivates the baseline model. Section §4 formally characterizes equilibrium learning, ranking strategies, and values of information. Proofs for the most immediate (respectively advanced) results are in the main body of the article (resp. in the appendix §A). Section §5 discusses implications for matching and welfare and the more general models.³ Section §6 concludes.

2 Literature review

Pre-Matching Information Acquisition

An interesting literature has recently emerged on pre-matching information acquisition. Artemov (2021) [1] is probably the closest related paper. It sets a continuum matching model with a continuum of students of mass 1 and 3 schools, A , B , and C , with respective capacities q , q , and 1, with $q \leq \frac{1}{3}$. Utilities at schools A and B are fixed to 1 and 0, but utility at school B is random $\frac{1}{2} + \epsilon^i$, $\mathbb{E}[\epsilon^i] = 0$. The default ranking over schools is therefore ABC . Students can privately observe ϵ_i at a cost $c(i)$, so that the optimal ranking at the interim stage may become BAC or ACB . The allocation is made through DA. Artemov (2021)’s interest is in the externality that the learning from one student

³Results in section Section §5 are currently only partial or stated in intuitive terms.

exerts on other students: when students are non-informed (informed), they rely on a common ex-ante criteria, competition is strong (weak), and the fit between matched students and colleges is bad (good). Because of the externalities, at equilibrium, too few students acquire information, resulting in a significant welfare loss. The paper proposes policy interventions in the spirit of affirmative action. It also compares incentives for learning under Deferred Acceptance (DA) vs Immediate Acceptance (IA). It runs simulations for the case with more than three schools.

The externalities described in Artemov (2021) are also at play in our model. Yet, there are several key differences between this paper and our paper. In Artemov (2021), students are ex-ante identical, with random priorities, whereas in our model, students differ by their commonly-known rank in priority. Therefore, our analysis can characterize learning as a function of the priority. We also consider more sophisticated preferences, where all colleges' values are uncertain, and all colleges' orders are possible at the ex-post stage. We solve the model for any number of colleges and students.

Chen and Hu (2022) [5] are interested in the incentives to learn cardinal preferences beyond ordinal ones. They define a game in three stages. First, each student decides to invest $\alpha \in [0, \bar{\alpha}]$ to acquire information on their ordinal preferences. The signal is perfect with probability α (and with probability $1 - \alpha$, the student does not observe anything). After the signal is realized, students decide to invest $\beta \in [0, \bar{\beta}]$ to acquire information on their cardinal preferences. The cost of information is a function of the sum $\alpha + \beta$. Then, allocation occurs through the DA or IA mechanisms. In the most interesting development of the original model, students can also acquire information ($\delta \in [0, \bar{\delta}]$) on other students' preferences after learning about their own preferences. The paper finds that under both DA and IA, a Bayes-Nash equilibrium exists and $\alpha^* > 0$. Under DA, $\beta^* = 0$, $\delta^* = 0$, for any preference and for any α^* , whereas under IA potentially $\beta^* > 0$, $\delta^* > 0$. This directly comes from DA (IA) being strategy-proof (non strategy-proof). In the paper, they also evaluate the welfare effects of endogenous information acquisition and of some information provision policies.

Just as in Artemov (2021), in Chen and Hu (2022), students are ex-ante identical and do not know priorities when entering the mechanism. This is very unrealistic for college admissions, where students have privileged access to their grades. This is especially inconsistent with the modeling of a pre-matching stage where students engage in information acquisition on preferences. Students could take advantage of this time period to also investigate the priorities, with positive benefits expected under non-strategy-proof mechanisms such as IA.

At the other extreme, Bucher and Caplin (2021) [4] set a model with 3 students and 3 colleges with unit capacities. Priorities are common and common knowledge. Preferences are described by a vector of colleges' values and are unknown ex-ante. The prior over the vector of values is exchangeable, meaning that all permutations of values across colleges are equally likely. In a first stage, students can choose any arbitrary information structure on the vector of values and pay a cost that is linear in the entropy reduction. Equivalently, students solve a rational inattention problem (Matejka and McKay (2015) [16], Mackowiak, Matejka, and Wiederholt (2023) [15]). In the second stage, the allocation occurs with DA. The analysis solves for students' posteriors after the first stage, the probabilistic matching, and the welfare. Technically speaking, the proof applies the already known solution of the general symmetric rational inattention problem from Matejka and McKay (2015) [16] inductively. It finds that student 2 attains a lower fraction of the full information surplus than student 1. The reason is that student 2 suffers from diluted incentive to acquire information. The conclusion is that endogenous information acquisition further distorts the allocation in favor of high-priority students.

Many forces driving the incentives to learn as a function of the priority rank in Bucher and Caplin (2021) also apply in our model. However, we find the assumption of exchangeable prior highly unrealistic. It implicitly means that preferences are fully idiosyncratic and completely unknown ex-ante. Before exerting costly effort to acquire information, students have no idea that, in general, Harvard is preferred to the district college. This notably implies that low-ranked students expect higher-ranked students to apply uniformly to all colleges. The assumption is key in the proof techniques as it makes the rational inattention problem symmetric at every stage of the induction. All in all, our view on Bucher and Caplin (2021) is that the paper is mostly an elegant application of the general rational inattention model rather than a credible description of college admission markets.

Similarly to us, most papers are motivated by the college admission application. Overall, the most consistent finding in this literature regards the complexity of the problem. All papers explicitly report this complexity as a major challenge. To maintain tractability, they are forced to simplify the problem in various ways: looking at a small dimensional problem with only a few students and colleges, posing a symmetric framework where colleges play the same role *ex-ante*, setting restrictive preferences, fixing small capacities at colleges. Setting a continuum of students is also a way to avoid dealing with numerous cases in the analysis.

With the present paper, our main contribution to the literature is a study of information acquisition with heterogeneous students and heterogeneous colleges. We wonder: “What student learns about what college?”. By contrast, Artemov (2021) wonder: “How much one student learns?” Bucher and Caplin (2021) wonder: “How much each student learns?” Chen and Hu (2022) wonder: “About what aspect of preferences one student learns?”.

General mechanism design with endogenous information acquisition

A few papers explore mechanism design with endogenous information acquisition. Bade (2015) [2] considers a housing problem where agents have unknown preferences. In the first learning stage, each agent chooses a partition of the set of possible vectors of the values of houses at some cost. She is revealed in what subset of this partition her true vector of values lies. In the second stage, a direct mechanism is used to allocate the houses. The equilibrium concept is Perfect Bayesian Equilibrium with truthful revelation of interim beliefs in the second stage. The main theorem says that (unlike what happens in the full information benchmark) there is a unique *ex-ante* Pareto-optimal, strategy-proof, and non-bossy allocation mechanism: serial dictatorship.

In a school choice context, Immorlica, Leshno, Lo, and Lucier (2020) [13] define “regret-free stability” as optimal information acquisition from students and no blocking pair at the interim stage. They show that regret-free stable outcomes always exist and that the stable outcomes are fully characterized by market-clearing cutoffs, where a college cutoff is defined by the lowest priority of students admitted to the college. They also show an impossibility theorem for mechanism design: no mechanism is regret-free stable for general economies (equivalently, no mechanism leads to price discovery). This is because students are trapped in “information deadlocks” where students need information to know which information they should efficiently gather. This suggests that only policy interventions aimed at providing information about (historical) cutoffs could help achieve regret-free outcomes.

Pre-Auction information acquisition

The present paper is also related to information acquisition in auctions. In particular, in Bobkova (2021) [3], bidders are uncertain about their valuation for the object. Their valuation consists of a common component and a private component. A bidder can choose which component to learn. The paper identifies conditions under which bidders only learn about their private component in the second-price auction, so that we stay in an independent value framework, ensuring efficiency.

Cardinal preferences, imperfect information, and ordinal mechanism

This paper is also part of a wider questioning that we have develop in several recent papers Goursat (2023) [8] and Goursat (2023) [14]. All these papers look at situations where information on preferences or priorities is imperfect, and the centralized allocation mechanism or decentralized blocking pair process is fully ordinal. In spite of this ordinality, the cardinality in the preferences and priorities always matters for the final allocation (through the belief formation on possible orders).

3 Model

3.1 The college admission problem

We model a college admission problem. There are $n \in \mathbb{N}$ students and colleges. A student (college) is denoted $i \in \{1, \dots, n\}$ ($j \in \{1, \dots, n\}$).

Preferences are defined as follows. The value of college j to student i , denoted v_j^i , is made of two additive⁴ components: $v_j^i := x_j + y_j^i$.

x_j is a common component, hence only indexed by the college j . It models the part of the college value that is the same among all students, for instance, the value of the degree in terms of employability and expected wage after college. We fix the order between the common components: $x_1 > \dots > x_n$ - college 1 (n) is the best (worst) college when we take into account only the common component. We also assume that $x_n > \frac{1}{2}$ and $\forall j \in \{1, \dots, n-1\} : x_j - x_{j+1} < \frac{1}{2}$.⁵

y_j^i is a private component, hence depending both on the college j and on the student i . It models the part of the college value that different students can value differently: the resources put in the different fields, the teaching styles, the living conditions on campus. We assume all y_j^i are i.i.d. across students and colleges. The prior is given by a continuous distribution with cdf F , pdf f on support $[-\frac{1}{2}, \frac{1}{2}]$. We assume f is symmetric around 0. Symmetry means that the patterns of good and bad surprises are similar in terms of frequencies and magnitudes.

When student i is admitted to college j , she gets as payoff the value v_j^i . We normalize the payoff from the outside option to 0.

Priorities are common, meaning that there is an objective ranking over students: $\forall j \in \{1, \dots, n\} : 1 \succ_j \dots \succ_j n$ - student 1 (n) is the best (worst) student. An underlying cardinal score, such as SAT score, may determine this ranking.

3.2 Information and learning

We assume the vector of common components $(x_j)_{j \in \{1, \dots, n\}}$ is common knowledge. The idea is that students can freely access information about employability and wages after college from objective data available in the press. Priorities are also common knowledge.

The private components $(y_j^i)_{i, j \in \{1, \dots, n\}}$ are unknown ex-ante. But, before the matching stage, student i can privately acquire information on the private components in her own preferences $(y_j^i)_{j \in \{1, \dots, n\}}$. Learning is modeled as perfect and free: if i decides to learn about j , she directly observes y_j^i . However, learning is capacity-constrained: student i can observe at most $K \in \{0, \dots, n\}$ components y_j^i .⁶ A learning strategy l^i consists in specifying K college indices in $\{1, \dots, n\}$.

Our favorite interpretation is that a student willing to learn about a given college can visit the campus or attend the open days of the college. Talking with teachers and students there will give the student a very good clue about how satisfied she would be studying in the college. The capacity constraint can model a time constraint or a logistical constraint. Indeed, there is limited time for exploration before the allocation stage begins. For instance, if students live far away from the colleges' campuses, traveling there will take a lot of time. In the meantime, students involved in the admission process need to work hard at high school to obtain high grades. The logistical constraint could be that some colleges organize their open days on the same days, forcing students to choose between colleges to visit. K is not indexed by i because students all face the same constraints.

For simplicity, we also denote the possibility that a student remains uninformed about her private components as an additional strategy l_0^i .

⁴The additive functional form implicitly assumes there is no interaction effect between the two components.

⁵We bound the gap between two colleges' common values to ensure that the gap does not always fully dominate the noise introduced by the private components.

⁶There are thus two extreme case: $K = 0$ ($K = n$) corresponds to no (full) information about the private components.

3.3 Matching

After the information acquisition stage, the allocation occurs through a centralized Deferred Acceptance (DA) mechanism (Gale and Shapley (1962) [7]). A ranking strategy r^i consists of submitting a rank-ordered list of the n colleges for each possible learning strategy l^i that student i could have adopted and for each possible realized signal that she can receive.

Because priorities are common, DA is equivalent to a dictatorship mechanism. Student 1 always gets her top choice; student 2 gets her top choice if it is different from student 1's top choice, otherwise, she gets her second choice; and so forth. The outcome of DA for player i is denoted μ^i .

3.4 Bayesian Game

This sets a game of incomplete information. The timing of the game is as follows: Nature draws private components $(y_j^i)_{i,j \in \{1, \dots, n\}}$. (1) Each student i chooses a learning strategy l^i . Then, the perfect signals are realized: each student i privately observes the y_j^i for the $j \in l^i$. She forms a posterior distribution over her preferences. (2) Each student i submits a rank-ordered list $r^i((y_j^i)_{j \in l^i})$. Then, the allocation is made through DA. Each student i receives her payoff, equal to v_j^i , the value of college j if she is assigned to this college ($\mu^i = j$), equal to zero if she is unassigned ($\mu^i = 0$). A strategy profile is denoted (L, R) , and the payoff to student i is $u^i(L, R)$.

What a student learns impacts the posterior belief and the submitted ranking of the student. Hence, it ultimately affects the final allocation. This imposes externalities on other students through the admission chances. We call these externalities “strategic uncertainty”. Due to the dictatorship mechanism in the second stage, strategic uncertainty goes only in one direction from high to low-priority students. The payoff effectively depends only on strategies of higher-priority students and of the student herself. Formally, if we denote $L^i := (l^1, \dots, l^i)$ and $R^i := (r^1, \dots, r^i)$ the truncated strategy profiles, the payoff simplifies as: $u^i(L, R) = u^i(L^i, R^i)$.

4 Results

We compute the Bayes-Nash equilibrium of the game described in the previous section §3.⁷

4.1 Preliminaries

Equilibrium ranking

We begin with a characterization of the rank-ordered lists submitted in stage (2) of the game. Because serial dictatorship is a strategy-proof mechanism, agents submit truthful rank-ordered lists. The order of colleges in the ranking is given by the order of the colleges' expected values in the posterior distribution. The expected value of college j in the posterior distribution of student i is equal to the true value v_j^i if student i has learned on college j ($j \in l_j^i$); otherwise ($j \notin l_j^i$), it is equal to the prior expectation $\mathbb{E}[v_j^i] = x_j$.

The default ranking, defined as the one submitted absent any information about the private component, is given by the ex-ante order over common components: $r_\emptyset := (1, \dots, n)$.

Equilibrium learning: gross analysis

Unlike in the ranking stage, in the learning stage, students are strategic. A student best-responds to what higher-ranked students decide to learn and rank; they fully internalize the strategic uncertainty. The only exception is student 1. Because he is the dictator, he bears no strategic uncertainty and faces a decision problem.

Learning about the private components improves a student's knowledge of her own preferences. The informed student will, therefore, submit a ranking that better reflects the true preferences. This

⁷Equilibrium strategies are denoted with stars *.

should translate in higher probabilities of being assigned to the preferred college, so in the end, into higher payoffs. The values of the information associated with each learning strategy capture this payoff increase.

In the first stage of the game, students acquire cardinal information by learning some colleges' values. But, in the second stage, students submit a ranking in an ordinal mechanism. Information has value in the first stage if and only if it changes the ranking in the second stage. We thus expect that students want to learn about colleges to which they have high chances of being admitted. The equilibrium probabilities that college j is still available when it is student i 's turn in the dictatorship mechanism, denoted p_j^i must matter in the analysis. The p_j^i are equilibrium variables: their values will be characterized, if not computed, in the main proof.⁸

In the meantime, information has a large value if and only if it is information about a college that a student values a lot. We thus expect that students want to learn about colleges with high ex-ante expected values, equivalently with high common components x_j .

There is tension between the two objectives (learning about safe colleges and learning about the best colleges). For low-ranked students, the tension is bound to be maximal.

4.2 The case $K = 1$

In the rest of the section, we focus on the case where $K = 1$. $K = 1$ means that each student can visit only one campus. A learning strategy specifies only one index in $\{1, \dots, n\}$. We denote l_j^i the strategy for student i visiting college j .

$K = 1$ is a natural benchmark case. When $K = 1$, the learning trade-off described in the previous section §4.1 resolves in a very mechanical and intuitive way.

We denote $U_j^i := \mathbb{E}[u^i(L^{(i-1)*}, l_j^i, R^{i*}(L^{(i-1)*}, l_j^i))]$ the ex-ante expected payoff for student i associated with learning strategy l_j^i , given that she will play the optimal ranking strategy, and fixing that higher-priority students play according to the equilibrium strategies. We denote $U_\emptyset^i := \mathbb{E}[u^i(L^{(i-1)*}, l_\emptyset^i, R^{(i-1)*}(L^{(i-1)*}, r_\emptyset^i))]$ the ex-ante expected payoff for student i when she has no information about her private components, so that she plays the default ranking, still fixing that higher-priority students play according to the equilibrium strategies.

Ranking with $K = 1$

Learning about just one college gives partial and limited information on the preferences. We write a series of lemmas stating how this influences ranking strategies. The lemmas are quite immediate; we omit the formal proofs and give the arguments in the main.

Lemma 1.

- (i) Learning about college l or learning about college $k \neq l$ are equally informative on the true order between v_l^i and v_k^i .
- (ii) Learning about college j gives (partial) information on the pairwise orders between college j and any other college.

Statement (i) will prove helpful to determine equilibrium strategies for student 1. It directly relies on the symmetry of the distribution of private components. Statement (ii) says that the posterior distribution over the rest of the preference orders between colleges other than j is unchanged and equal to the prior. The only perturbation regards j 's position in this order.

This directly characterizes possible ranking strategies after learning about college j .

Lemma 2.

After learning strategy l_j^i , the ranking $r^i(l_j^i)$ can only preserve the ex-ante order over colleges other than j and move college j in this order.

⁸Because student 1 is the dictator: $p_j^1 = 1, \forall j \in \{1, \dots, n\}$.

Consequently, there are only n possible rankings after learning strategy l_j^i .

For the next lemma, we locate at an interim stage of the dictatorship mechanism, where some colleges have already been assigned. The lemma characterizes optimal ranking strategies for fixed colleges' capacities and fixed preferences. In particular, it emphasizes that a student may be indifferent between several rankings, depending on the pattern of remaining seats at colleges, when it is her turn to choose in the dictatorship mechanism.

Lemma 3.

- (i) *If there is at most one available seat in a subset of colleges C , then a student is indifferent between the ranking strategies that would differ only in the orders within set C .*
- (ii) *If there is at least one available seat in the subset of her most preferred k colleges, then the optimal rankings rank the k colleges first and differ only in the orders within the set of the $(n - k)$ other colleges.*

In statement (i), the student understands that which college she can be allocated in set C depends only on the remaining capacities. The ranking of colleges within C has no influence on the outcome. In statement (ii), the student understands that she should rank her preferred colleges first and that she will be allocated a college within this set so that the rest of the ranking does not matter. In the induction proof for our main result, it will appear clearly that all students are in these (i) and (ii) situations. Whenever we want to emphasize that a student is indifferent locally in her ranking strategy, we use brackets. For instance, ranking strategy $(\{1, 2\}, 3)$ represent both lists $(1, 2, 3)$ or $(2, 1, 3)$ with indifference. When we know that a student will be allocated a college in the top tier of a list, we only describe the relevant part of the list. For example, $(1, 2, _)$ represents any list where college 1 is ranked first, college 2 is ranked second, and conditional on submitting this list, student i will be assigned college 1 or 2 for sure. For example, suppose that we are in the context of statement (ii) and denote C the set of the k most preferred colleges of the student where there is at least one seat. Statement (ii) says that the student should submit any ranking represented by $(\{C\}, _)$.

For the next lemma, we will need the following vocabulary.

Definition 1.

- (a) *A college j is safe for student i if $p_j^i = 1$.*
- (b) *A set of colleges is collectively safe for student i if there is exactly one seat available in these colleges when it is i 's turn to choose in the dictatorship mechanism.*

We emphasize that collective safety in item (b) means that there is exactly one seat available and not more. If a set of college C is collectively safe for student i , this implies: $\sum_{j \in C} p_j^i = 1$, but collective safety is a stronger condition than just the probabilities summing up to one. It also says something about the dependence between admission events: if student i is admitted to a college $j \in C$, she is rejected at any other college in the set $j' \notin C$.

We will also need the following assumption in the induction.

Definition 2.

(A^i) For student $i \in \{1, \dots, n - 1\}$:

- *Colleges $j \in \{1, \dots, i\}$ are collectively safe.*
- *Each college $j \in \{i + 1, \dots, n\}$ is a safe college.*

We denote $r_k := (1, \dots, k)$, $k \in \{1, \dots, n\}$ ($r_n = r_\emptyset$) the truncated list of colleges in the ex-ante order. By lemma 3., the default ranking submitted by an uninformed student i under assumption (A^i) writes $(\{r_i\}, _)$.

We are finally able to characterize optimal rankings after any learning strategy and signal in the environment of assumption (A^i) .

Proposition 1.

If $i \in \{1, \dots, n-1\}$, assumption (A^i) is verified, and student i learns about college j , then:

(i) If $1 \leq j \leq i$, student i submits:

- $r^i(l_j^i) = (\{r_i\}, _)$ with probability $F(x_j - x_{i+1})$
- $r^i(l_j^i) = (\{r_i \setminus \{j\}\}, i+1, _)$ with probability $1 - F(x_j - x_{i+1})$

(ii) If $i+1 \leq j \leq n$, student i submits:

- $r^i(l_j^i) = (\{r_i\}, _)$ with probability $F(x_i - x_j)$
- $r^i(l_j^i) = (\{r_k\}, j, _)$ with probability $F(x_k - x_j) - F(x_{k+1} - x_j)$, $\forall k \in \{0, \dots, i-1\}$ ⁹

When student i learns about college j , two cases arise.

In the first case, $1 \leq j \leq i$, so that college j is part of the default ranking. If, after learning about college j , student i observes a high or intermediate y_j^i , this modifies the interim preferences. This does not modify the ranking strategy, due to assumption (A^i) combined with lemma 3., statement (i) (moving j within r_i would have no influence on the final allocation). The probability $F(x_j - x_{i+1})$ is the probability that $v_j^i > \mathbb{E}[v_{i+1}^i] = x_{i+1}$. But if student i observes a very low y_j^i , she excludes college j from the relevant part of the list r_i and simply appends the next ex-ante best college, college $i+1$. By assumption (A^i) , this is a safe college, hence (by lemma 3., statement (ii)) no need to specify the ranking further. This sub-case happens with the complement probability $1 - F(x_j - x_{i+1})$, that is, the probability that $v_j^i < \mathbb{E}[v_{i+1}^i] = x_{i+1}$.

In the second case, $i+1 \leq j \leq n$, so that college j is not part of the default ranking. If after learning about college j , student i observes an intermediate or low y_j^i , she sticks to the default ranking. If she observes a high y_j^i , she pushes college j to the top part of the ranking. Where exactly depends on the exact y_j^i : $F(x_k - x_j) - F(x_{k+1} - x_j)$ is the probability that $\mathbb{E}[v_{k+1}^i] = x_{k+1} < v_j^i < \mathbb{E}[v_k^i] = x_k$. By assumption (A^i) , college j is safe. By lemma 3., statement (ii), the ranking can stop at j . It need not include the colleges that were in the default ranking but that are interim less attractive than j .

The fact that assumption (A^i) is indeed verified for any student i at equilibrium will emerge in the induction proof of our main result.

Learning with $K = 1$

Proposition 1. characterizes optimal rankings. From this, we can get interim expected payoff after each potential signal realization. For example, the interim expected payoff for student i after learning about college $1 \leq j \leq i$ (case (i)) and after observing $v_j^i > x_{i+1}$ ($v_j^i < x_{i+1}$) is $p_j^i v_j^i + \sum_{l \in r_i \setminus \{j\}} p_l^i x_l$

$$(p_j^i x_{i+1} + \sum_{l \in r_i \setminus \{j\}} p_l^i x_l).$$

We are, therefore, equipped to compute ex-ante payoffs and the value of information associated with each learning strategy. The value of information for strategy l_j^i of student i given that higher-priority students play the equilibrium strategies is denoted V_j^i . It is defined as the increase in ex-ante payoff from learning about j with respect to not learning $V_j^i := U_j^i - U_\emptyset^i$.

The next proposition gives concise formulas for the values of information associated with any possible learning strategy of student $i \in \{1, \dots, n\}$.

Proposition 2.

If assumption (A^i) is verified:

(i) If $1 \leq j \leq i$:

$$V_j^i = p_j^i \int_{x_j - x_{i+1}}^{\frac{1}{2}} (x_{i+1} - x_j + y) f(y) dy$$

⁹We abuse notations and set: $x_0 := x_j + \frac{1}{2}$.

(ii) If $i + 1 \leq j \leq n$:

$$V_j^i = \sum_{m=0}^{i-1} \int_{x_{m+1}-x_j}^{x_m-x_j} \left[\sum_{l \in r_i \setminus r_m} p_l^i(x_j - x_l + y) \right] f(y) dy$$

The proof (available in appendix §A) integrates the interim expected payoffs for all possible values of the private components y , computes the difference with the benchmark payoff from the default no information case, and simplifies. The symmetry of the distribution of y is key.

The next theorem compares the values of information across all colleges within each student.

Theorem 1.

(i) If assumption (A^i) is verified: $\forall i \in \{1, \dots, n-1\}$: $\begin{cases} V_{i+1}^i = \sum_{j=1}^i V_j^i & (1) \\ V_{i+1}^i > V_k^i, \forall k \in \{i+2, \dots, n\} & (2) \end{cases}$

(ii) If assumption (A^n) is verified: $\forall j \in \{1, \dots, n\}$: $V_j^n = 0$

The proof (available in appendix §A) just plays with the formulas from proposition 2.. We give a short intuition. Recall that under assumption (A^i) , colleges $C_i := \{1, \dots, i\}$ are collectively safe, each college $j \in S_i := \{i+1, \dots, n\}$ is safe for student i . The student cannot influence what college in the set C_i she could get. So what matters for her is to be able to rank correctly each college in C_i with respect to each college in S_i , or colleges within S_i . Because of the capacity constraint on learning, the student can learn about the order between one college and the rest of the colleges. The only benefit from learning about a college $j \in C_i$ is that it (partially) reveals the order between j and $i+1$. The information revealed on the orders between j and any $j' \in \{i+2, \dots, n\}$ is irrelevant because after learning about j , the student's posterior belief is still that $i+1$ is better than $i+2$, and $i+1$ is safe. Conversely, there are various benefits from learning about college $i+1$: it (partially) reveals all pairwise orders $i+1$ vs $j \in \{1, \dots, i\}$. Hence the summation result (i)(1). Learning about a college $k \in \{i+2, \dots, n\}$ (partially) reveals all pairwise orders k vs $j \in \{1, \dots, i\}$. But the ex-ante probability that college k is better than some colleges j is smaller than with college $i+1$, leading to less frequent pairwise preference orders reversals and lower value of information. Hence the strict inequality (i)(2). Finally, when it is student n 's turn in the dictatorship mechanism, there is exactly one seat left in the whole set of n colleges. Student n endures all higher-priority students' choices and cannot influence her final outcome at all. She is fully indifferent between any ranking that ranks all colleges ex-ante, and also interim, after any learning strategy. Learning has no stakes, hence no value ((ii)). Statement (i)(1) could be interpreted as follows: for the generic student, the strict unit capacity constraint on learning is not too tragic. By learning about the right college, they can still capture much of the relevant information.

At equilibrium, given what higher-ranked students have learned and ranked, each student chooses the learning strategy with the highest value. So equilibrium strategies are immediate from theorem 1..

Theorem 2.

At Bayes-Nash equilibrium of the college admission problem with information acquisition and $K = 1$.

(i) Student 1:

(l^1) learns about colleges 1 or 2

(r^1) submits rankings:

- ($1, _$) with probability $F(x_1 - x_2)$

- ($2, _$) with probability $1 - F(x_1 - x_2)$

(μ^1) is assigned to college 1 or college 2

(ii) Student $i \in \{2, \dots, n-1\}$:

(l^i) learns about college $i+1$

(r^i) submits rankings:

- ($\{r_i\}, _$) with probability $F(x_i - x_{i+1})$

- $(\{r_k\}, i+1, _)$, $k \in \{0, \dots, i-1\}$ with probability $F(x_k - x_{i+1}) - F(x_{k+1} - x_{i+1})$
 (μ^i) is assigned to a college in $\{1, \dots, i+1\}$

(iii) Student n :

(l^n) learns about any college $j \in \{1, \dots, n\}$
 (r^n) submits as ranking any permutation of $(\{r_n\})$
 (μ^n) is assigned to a college $j \in \{1, \dots, n\}$

Statement (ii) is a direct consequence of statement (i) in 1. for the learning part (and proposition 1. for the application part). We have described the general intuition above. Statement (i) also comes from statement (i) in theorem 1.. Taken with $i = 1$, we get $V_1^1 = V_2^1 > V_k^i$, $\forall k \in \{3, \dots, n\}$, hence indifference between l_1^1 and l_2^1 . Intuitively, student 1 as a dictator will always get the object she submits first in her ranking. For her, the problem simplifies to learning what is her most preferred college and applying to this college. The most likely preferred colleges are colleges 1 and college 2, so student 1 wants to learn about the pairwise order between the two. To this end, and by lemma 1., statement (i), learning strategies l_1^1 or l_2^1 are equivalent. Both strategies lead to applying either college 1 or to college 2 depending on the signal. Statement (iii) directly comes from statement (ii) in theorem 1.. In the proof, we construct an induction argument, showing that any student i knows that she will be in the situation described by assumption (A^i) when it is her turn to choose in the dictator mechanism.

This result predicts a strong monotonicity in learning: higher-priority students learn about the ex-ante best colleges.

5 Work in Progress

5.1 Matching and welfare in the baseline model

Theorem 2. fully characterizes the learning strategies and the ranking strategies. With respect to the allocation, the theorem only specifies the set of possible allocations at equilibrium. Equivalently, in the proof of the theorem, we only characterize whether the probabilities of admission p_j^i are positive or zero. However, the theorem does not specify the probability distribution over the possible allocations. Equivalently, in the proof, we do not characterize the values of the probabilities of admission p_j^i . This step of the analysis is ongoing. It is again an induction proof, but not immediate. In the inductive step, it is not enough to keep track of the probabilities of admission of all higher-ranked students in the college. We need the full description of rankings submitted by all higher-ranked students. The combinatorial computations involved in the proofs are intricate.

Characterizing the matching is a necessary step towards a welfare analysis. The questions are: (1) What is the absolute value of information? To answer this one, we need to explicitly compute the probabilities of acceptance p_j^i and to input them into the formulas from proposition 2.. (2) What is the relative value of information - equivalently, what share of the payoff difference between no and full information does a student capture by learning about just one college? Under no (full) information, students submit rankings consistent with the common components (rankings that perfectly reflect their true preferences). (3) How do the absolute and relative values of information vary with the student priority rank?

In the full information benchmark, we know there is some monotonicity. Higher-priority students capture higher values of information, because they do not suffer from strategic uncertainty. Lower-ranked students suffer from strategic uncertainty: there is always the possibility that ranking colleges in a more appropriate way leads to no payoff improvement because the colleges in question are already full with higher-priority students.

Under capacity-constrained learning, the answer is not as obvious, even considering the absolute value of information. The equilibrium strategy of a low-priority (high-priority) student enables him to learn partially about many (few) pairwise orders. At the extreme, student 1 only learns about the pairwise order between colleges 1 and 2. This effect could offset the strategic uncertainty effect.

We emphasize the stakes of the welfare analysis. If the possibility for information acquisition improves more the situation of lower-ranked (higher-ranked) agents, learning is a force towards equity (inequity) on the market. It alleviates (adds up to) the existing discrimination performed by the priority order. A market designer keen to keep control over the amount of discrimination on the market could consider information regulation interventions.

5.2 Larger capacities at colleges

In the baseline model, each college offers one seat on the market. In practice, though, the same college offers multiple replica seats. Ultimately, we should be able to switch to a model of many-to-one matching if we really want to accommodate the college admissions application.

We give an overview of the model and equilibrium computation when we relax the capacities at colleges. We set that each college $j \in \{1, \dots, n\}$ offers c_j seats. There are $\sum_{j=1}^n c_j$ students on the market.

All students $i \in \{\min(c_1, c_2)\}$ iteratively face the same problem as student 1 in the baseline model. They know that higher-priority students learn about college 1 or 2 and apply to colleges 1 or 2. They also know that capacities cannot be exhausted at any college yet so that all colleges remain individually safe. These students $i \in \{\min(c_1, c_2)\}$ all keep on learning about college 1. When it is student $i = \min(c_1, c_2)$'s turn, some capacities could be exhausted. Student i knows that (1) capacities could be exhausted at one of the colleges 1 or 2, in which case she would be in the same situation as student 2 in the baseline model and should learn about college 3. But potentially, (2) there are seats remaining at both colleges 1 and 2, in which case student i would be again in the situation of student 1 in the baseline model and should learn about college 1 or 2. Whether student i learns about colleges 1, 2 or 3 depends on the probabilities of case (1) vs case (2). These probabilities directly depend on capacities, and indirectly on other parameters of the game through the equilibrium probabilities of application of higher-priority students.

The (provisional) conclusion is that with larger capacities in colleges, the equilibrium structure is little robust. Comparative static analysis may become interesting.

5.3 Larger learning capacities

The obvious next step in the analysis consists of relaxing the capacity constraint on learning and allowing students to visit several colleges.

Baseline model with relaxed learning capacities

In a first extension, we can set $K = 2$, with students allowed to visit two colleges k, l . This enables them to learn partially about all orders involving k and l and to reveal fully the order between colleges k and l .

The case $K = 2$ is already very computationally intense, even for player 1, for mainly two reasons: the large size of the action space (made of any set of two college indices in $\{1, \dots, n\}$, hence cardinal $\frac{n(n-2)}{2}$), and the comparison between multiple stochastic variables. Restricting the number of colleges to $n = 3$, as common in the literature, solves the first issue, but not the second one. For example, after learning strategy $l_1^1 = \{1, 2\}$, student 1 applies to college 1/2/3 if and only if $v_1 > v_2, x_3 / v_2 > v_1, x_3 / x_3 > v_1, v_2$. The interim expected payoffs write with double integrals, and the ex-ante payoff formulas are extremely big.

Yet, some intuitions from the baseline model with $K = 1$ may help. With $K = 2$, a student can learn about multiple pairwise orders. Student 1 is mostly interested in the pairwise orders involving good ex-ante colleges, with low indices. $l_1^1 = \{1, 2\}$ would perfectly (partially) reveal the order 1 vs 2 (1 vs 3 and 2 vs 3). l_1^1 therefore appears as student 1's optimal strategy. Student 1 applies to and is assigned to a college in 1, 2 or 3. As a consequence, when it is student 2's turn to choose in the mechanism, there are exactly two seats left in colleges 1, 2 and 3. All colleges $j \geq 4$ are individually safe. learning about some colleges in $\{1, 2, 3\}$ should be considered because there are not only one seat available in this set, but two seats. It remains possible that ranking colleges correctly in this set can matter.

In the general model with relaxed capacities, the trade-off between the expected values of the colleges and the equilibrium acceptance chances appears. It cannot be simplified with any equivalent to assumption (A^i) from the baseline model. The equilibrium will likely differ for different parameters (common components' values, private components' distributions).

Model with two tiers of colleges

One way to simplify the model to be able to accommodate the relaxed capacities would be to introduce tiers of colleges. Colleges belonging to the same tier would share the same common component. For instance, we can consider two tiers, a top tier T , made of colleges T_1 and T_2 with common component x_T and a bottom tier made of colleges B_1 and B_2 with common component x_B . The values of colleges to students write: $v_{T_j}^i = x_t + y_{T_j}^i$ or $v_{B_j}^i = x_t + y_{B_j}^i$. Private components $y_{T_j}^i, y_{B_j}^i$ are i.i.d. distributed according to some continuous distribution on $[\frac{1}{2}, \frac{1}{2}]$, symmetric around 0. We denote $\Delta := x_T - x_B$ the ex-ante gap in value between a college from the top tier and a college from the bottom tier. We bound $\Delta < \frac{1}{2}$, to make sure that learning about one college could flip some pairwise orders. When $\Delta \rightarrow 0$, the 4 colleges have the same role. When $\Delta \rightarrow \frac{1}{2}$, top colleges remain better than bottom colleges even at the interim stage if the student learns about just one college, or about colleges belonging to the same tier;

In the case $K = 1$, the model is just a limit case of the baseline model, setting $n = 4$ and $x_2 \rightarrow x_1, x_4 \rightarrow x_3$. The action space simplifies to specifying a letter T (B)¹⁰ if the student wants to learn about a top (bottom) college.¹¹

At equilibrium, student 1 learns about a top college and is assigned with uniform probability to college T_1 or T_2 . Student 2 learns about a bottom college, and the value of information from this strategy equals twice the value of information from learning about a top college. Indeed, learning about one B_j gives a good clue about the pairwise orders between this B_j and any of the T_i s. learning about one T_j gives a good clue about the pairwise orders between this T_j and any of the B_i s. But there is always a one-half chance that knowing these orders is useless because T_j would have accepted student 1 already and would not be available anymore. For student 2, ranking colleges within tiers is never helpful because the allocation within T is fully set by student 1's choice, and because she will never rank the two bottom colleges, she will always put the unexplored top colleges first. Student 2 can be assigned to any college. Student 3 also learns about a bottom college. Indeed, learning about one T_j enables one to rank T_j against the bottom colleges, but this is helpful only if T_j is still available, which is quite unlikely. Learning about one B_j enables one to rank B_j within B , which is helpful if both bottom colleges are still available, and to rank B_j against top colleges, which is helpful if B_j is still available. Student 3 can be allocated to any college. Finally, student 4 is indifferent between any learning and ranking strategies.

When we set $K = 2$, the action space is made of only 3 actions: $T := \{T_1, T_2\}$, $B = \{B_1, B_2\}$ and $H := \{T_j, B_l\}$ (the "hybrid" action, where the student learns about one top and one bottom college).

5.4 Other mechanisms

The baseline model fixes the allocation mechanism to be the famous DA mechanism. Although most college admission markets use some version of DA, it would be interesting to know what happens with other widely-used mechanisms.

In the Immediate Acceptance (IA) mechanism, each college's acceptance decision is final. The strength of DA is in its incentive property: DA is strategy-proof for both sides, whereas the strength of IA is in the efficiency: IA is Pareto-efficient (conditional on truthful reporting by the students). IA, though, is not strategy-proof: A student may put an intermediate college in her top choice, dreading that this intermediate college may accept a student in the first rounds.

¹⁰Because colleges within the same tier are ex-ante the same, a student is indifferent between which college precisely he learns about, and this is random.

¹¹This model is different from the baseline model with relaxed capacities at colleges from section 5.2, because the private components differ across different colleges within the same tier, so that students separately learn about each.

Because our model uses a capacity constraint on learning rather than a cost function, the results from Artemov (2021) [1] and Chen and Hu (2022) [5] (Bucher and Caplin (2021) [4]) that IA induces more (more equal) information acquisition is irrelevant.¹² Yet, it could change the equilibrium ranking and learning strategies. Student i is no longer indifferent between rankings that differ only in the order within set $\{1, \dots, i\}$. And IA basically increases the stakes of the first rounds of the mechanism. There is a stronger incentive to learn about the ex-ante best colleges.

On real-life college admission markets, mechanisms are almost always truncated (as reported by Pathak (2016) [17]), meaning that students are allowed to rank only a limited number of colleges. One reason could be that there are operational costs associated with reviewing many applications. Unfortunately, the truncated version of DA is no longer strategy-proof (Haeringer and Klijn (2009) [9]). The preference manipulation consists of including safe colleges in the list. In our setting, a truncation of order t should have no impact on the highest priority students $i \leq t - 1$, because they submit rankings shorter than t anyway. For the lowest-priority students, the learning strategy should also remain unchanged. But the ranking strategy could change. Indeed, if student $i \geq t$ sticks to the ranking strategy that is just the truncation of the equilibrium strategy from the baseline model, she has a positive probability of remaining unassigned. It may be profitable for her to rather include the best safe college $i + 1$ at the t rank in the list.

6 Provisional conclusion

In this note, we answer a simple but empirically relevant question: If each student has time to visit one college to learn about how she values the college before participating in a centralized matching mechanism, what college does she visit? The analysis shows that learning strategies are monotonic, with higher-priority students learning about the ex-ante better colleges. By optimally learning about her best safe college, a generic student manages to learn partially about all relevant pairwise orders. Consequences in terms of matching and welfare are still to be explored.

The long-run project is far more ambitious. The plan is to enrich the model with relaxed capacities at colleges, relaxed capacity constraint on learning, and mechanism design, while keeping the heterogeneity in students and colleges. The analysis should solve the student trade-off between learning about high-value colleges or learning about safe colleges.

The general framework stands as the most appropriate model to explore the effects of endogenous information acquisition of (in)equity in matching markets.

¹²Except for student n , who now captures a positive value of information.

References

- [1] G. Artemov. Assignment mechanisms: common preferences and information acquisition. *Journal of Economic Theory*, 198:105370, 2021.
- [2] S. Bade. Serial dictatorship: The unique optimal allocation rule when information is endogenous. *Theoretical Economics*, 10(2):385–410, 2015.
- [3] N. Bobkova. Information choice in auctions. *CEPR Discussion Paper No. DP16832*, 2021.
- [4] S. F. Bucher and A. Caplin. Inattention and inequity in school matching. Technical report, National Bureau of Economic Research, 2021.
- [5] Y. Chen and Y. He. Information acquisition and provision in school choice: a theoretical investigation. *Economic Theory*, 74(1):293–327, 2022.
- [6] A. Dustan, A. De Janvry, and E. Sadoulet. Flourish or fail?: The risky reward of elite high school admission in Mexico City. *Journal of Human Resources*, 52(3):756–799, 2017.
- [7] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- [8] L. Goursat. Whether and where to apply: Information and discrimination on frictional matching markets. *mimeo*, 2023.
- [9] G. Haeringer and F. Klijn. Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947, 2009.
- [10] J. S. Hastings and J. M. Weinstein. Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics*, 123(4):1373–1414, 2008.
- [11] C. Hoxby, S. Turner, et al. Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, 12(014):7, 2013.
- [12] C. M. Hoxby and C. Avery. The missing "one-offs": The hidden supply of high-achieving, low income students. Technical report, National Bureau of Economic Research, 2012.
- [13] N. Immorlica, J. Leshno, I. Lo, and B. Lucier. Information acquisition in matching markets: The role of price discovery. Available at SSRN 3705049, 2020.
- [14] G. Laure. How can I know how much I like you? a heuristic approach to stability in matching. *mimeo*, 2023.
- [15] B. Maćkowiak, F. Matějka, and M. Wiederholt. Rational inattention: A review. *Journal of Economic Literature*, 61(1):226–273, 2023.
- [16] F. Matějka and A. McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298, 2015.
- [17] P. A. Pathak. What really matters in designing school choice mechanisms. *Advances in Economics and Econometrics*, 1:176–214, 2017.
- [18] A. E. Roth. Two-sided matching with incomplete information about others' preferences. *Games and Economic Behavior*, 1(2):191–209, 1989.

Appendix A Proofs

Proof of proposition 2.

Consider a student i who learns about college j .

(i) $1 \leq j \leq i$

We summarize below, the possible signal realizations, and the associated optimal rankings (from proposition 1.) and interim expected payoffs:

Signal	Ranking	Payoff
$v_j^i > x_{i+1} \iff y_j^i > x_{i+1} - x_j$	$(\{r_i\}, _)$	$p_j^i v_j^i + \sum_{l \in r_i \setminus \{j\}} p_l^i x_l$
$v_j^i < x_{i+1} \iff y_j^i < x_{i+1} - x_j$	$(\{r_i \setminus \{j\}\}, i+1, _)$	$p_j^i x_{i+1} + \sum_{l \in r_i \setminus \{j\}} p_l^i x_l$

We compute the ex-ante payoffs integrating the interim expected payoff with respect to the signal y on the whole support $[-\frac{1}{2}, \frac{1}{2}]$:

$$U_j^i = \int_{x_{i+1}-x_j}^{\frac{1}{2}} (p_j^i(x_j + y) + \sum_{l \in r_i \setminus \{j\}} p_l^i x_l) f(y) dy + \int_{-\frac{1}{2}}^{x_{i+1}-x_j} (p_j^i x_{i+1} + \sum_{l \in r_i \setminus \{j\}} p_l^i x_l) f(y) dy$$

$$U_\emptyset^i = \int_{-\frac{1}{2}}^{\frac{1}{2}} (p_j^i(x_j + y) + \sum_{l \in r_i \setminus \{j\}} p_l^i x_l) f(y) dy$$

The value of information is just the difference. The terms corresponding to the default ranking cancel out:

$$\begin{aligned} V_j^i &= \int_{-\frac{1}{2}}^{x_{i+1}-x_j} \left[(p_j^i x_{i+1} + \sum_{l \in r_i \setminus \{j\}} p_l^i x_l) - (p_j^i(x_j + y) + \sum_{l \in r_i \setminus \{j\}} p_l^i x_l) \right] f(y) dy \\ &= \int_{-\frac{1}{2}}^{x_{i+1}-x_j} [p_j^i(x_{i+1} - x_j - y)] f(y) dy \\ &= p_j^i \left[(x_{i+1} - x_j) \int_{-\frac{1}{2}}^{x_{i+1}-x_j} f(y) dy - \int_{-\frac{1}{2}}^{x_{i+1}-x_j} y f(y) dy \right] \\ &= p_j^i \left[(x_{i+1} - x_j) \int_{x_{i+1}-x_j}^{\frac{1}{2}} f(y) dy + \int_{x_{i+1}-x_j}^{\frac{1}{2}} y f(y) dy \right] \text{ (by symmetry of } f) \\ &= p_j^i \int_{x_{i+1}-x_j}^{\frac{1}{2}} (x_{i+1} - x_j + y) f(y) dy \end{aligned}$$

(ii) $i+1 \leq j \leq n$

We follow the same steps:

Signal	Ranking	Payoff
$v_j^i < x_i \iff y_j^i < x_i - x_j$	$(\{r_i\}, _)$	$\sum_{l \in r_i} p_l^i x_l$
$k \in \{1, \dots, i-1\}: x_{k+1} < v_j^i < x_k \iff x_{k+1} - x_j < y_j^i < x_k - x_j$	$(\{r_k\}, j, _)$	$(\sum_{l \in r_i \setminus r_k} p_l^i) v_j^i + \sum_{l \in r_k} p_l^i x_l$

$$U_j^i = \int_{-\frac{1}{2}}^{x_i-x_j} (\sum_{l \in r_i} p_l^i x_l) f(y) dy + \sum_{k=0}^{i-1} \int_{x_{k+1}-x_j}^{x_k-x_j} \left[(\sum_{l \in r_i \setminus r_k} p_l^i) (x_j + y) + \sum_{l \in r_k} p_l^i x_l \right] f(y) dy$$

$$U_\emptyset^i = \int_{-\frac{1}{2}}^{\frac{1}{2}} (\sum_{l \in r_i} p_l^i x_l) f(y) dy$$

$$V_j^i = \sum_{k=0}^{i-1} \int_{x_{k+1}-x_j}^{x_k-x_j} \left[(\sum_{l \in r_i \setminus r_k} p_l^i) (x_j + y) + \sum_{l \in r_k} p_l^i x_l - \sum_{l \in r_i} p_l^i x_l \right] f(y) dy$$

$$V_j^i = \sum_{k=0}^{i-1} \int_{x_{k+1}-x_j}^{x_k-x_j} \left[\sum_{l \in r_i \setminus r_k} p_l^i (x_j - x_l + y) \right] f(y) dy$$

Proof of theorem 1.(i) $1 \leq i \leq n - 1$

– Preliminary

By proposition 2., statement (ii), with $i + 1 \leq j \leq n$:

$$\begin{aligned}
V_j^i &= \sum_{k=0}^{i-1} \int_{x_{k+1}-x_j}^{x_k-x_j} \left[\sum_{l \in r_j \setminus r_k} p_l^i (x_j - x_l + y) \right] f(y) dy \\
&= \sum_{k=0}^{i-1} \sum_{l \in r_j \setminus r_k} p_l^i \int_{x_{k+1}-x_j}^{x_k-x_j} (x_j - x_l + y) f(y) dy \\
&= \sum_{r=1}^i p_r^i \sum_{q=0}^{r-1} \int_{x_{q+1}-x_j}^{x_q-x_j} (x_j - x_r + y) f(y) dy \\
&= \sum_{r=1}^i p_r^i \int_{x_r-x_j}^{\frac{1}{2}} (x_j - x_r + y) f(y) dy
\end{aligned}$$

– Taking $j = i + 1$ above gives:

$$\begin{aligned}
V_{i+1}^i &= \sum_{r=1}^i p_r^i \int_{x_r-x_{i+1}}^{\frac{1}{2}} (x_{i+1} - x_r + y) f(y) dy \\
&= \sum_{j=1}^i V_j^i \quad (1)
\end{aligned}$$

– In addition for $i + 1 \leq j \leq n$:

$$\begin{aligned}
V_{i+1}^i - V_j^i &= \sum_{r=1}^i p_r^i \left[\int_{x_r-x_{i+1}}^{\frac{1}{2}} (x_{i+1} - x_r + y) f(y) dy - \int_{x_r-x_j}^{\frac{1}{2}} (x_j - x_r + y) f(y) dy \right] \\
&= \sum_{r=1}^i p_r^i \left[\int_{x_r-x_j}^{\frac{1}{2}} (x_{i+1} - x_j) f(y) dy + \int_{x_r-x_{i+1}}^{x_r-x_j} (x_{i+1} - x_r + y) f(y) dy \right] \\
&= \sum_{r=1}^i p_r^i \left[\left(\frac{1}{2} + x_{i+1} - x_r \right) (x_{i+1} - x_r) + \int_{x_r-x_{i+1}}^{\frac{1}{2}} y f(y) dy - \left(\frac{1}{2} + x_j - x_r \right) (x_j - x_r) - \int_{x_r-x_j}^{\frac{1}{2}} y f(y) dy \right] \\
&= \sum_{r=1}^i p_r^i \left[\left(\frac{1}{2} + x_{i+1} - x_r \right) (x_{i+1} - x_r) - \left(\frac{1}{2} + x_j - x_r \right) (x_j - x_r) + \int_{x_r-x_{i+1}}^{x_r-x_j} y f(y) dy \right] \\
x_{i+1} > x_j &\implies \left(\frac{1}{2} + x_{i+1} - x_r \right) (x_{i+1} - x_r) > \left(\frac{1}{2} + x_j - x_r \right) (x_j - x_r) \\
x_r - x_j > x_r - x_{i+1} > 0 &\implies \int_{x_r-x_{i+1}}^{\frac{1}{2}} y f(y) dy > 0
\end{aligned}$$

$$V_{i+1}^i - V_j^i > 0 \quad (2)$$

(ii) $i = n$ When it is student n 's turn in the dictatorship mechanism, there is exactly one seat left in the whole set of n colleges.From lemma 3., statement (i), student n is indifferent between any ranking. So whatever the learning strategy, student n cannot increase her payoff by adapting her ranking strategy to the signal.So $\forall j \in \{1, \dots, n\} : V_j^n = 0$ **Proof of theorem 2.**• $1 \leq i \leq n - 1$ (Strong) induction

- Induction hypothesis
 $H(i) : (A^i)$ is verified and l^i, r^i, μ^i are as in theorem 2..
- Initial step
 (A^1) means that all colleges are safe for student 1. Because student 1 is the dictator, it is trivially verified.
 l^1 follows from theorem 1. (i), and the maximization of the value of information.
 r^1 follows from proposition 1., (i) and (ii). It says that student 1 applies to college 1 or college 2. Because student 1 is the dictator, whenever he applies to a college, he gets it. So μ^1 is also characterized.
 So $H(1)$ is verified.
- Inductive step
 Set $i \in \{1, n - 2\}$. Suppose that $\forall l \in \{1, \dots, i\}, H(l)$ is verified.
 The ranking strategies from $H(l), l \in \{1, \dots, i\}$ say that each student l 's rankings include colleges $\{1, \dots, l + 1\}$.
 Because rankings always include either the whole set of collectively safe colleges $\{1, \dots, l\}$ or include the safe college $l + 1$, any student l will be assigned to a college.
 In total only the $(l + 1)$ colleges in $\{1, \dots, i + 1\}$ receive applications from students in $\{1, \dots, i\}$. l out of the $(l + 1)$ colleges are allocated a student.
 Given the ranking strategies from $H(l), l \in \{1, \dots, i\}$, the remaining seat can be at any college. Suppose, for instance, that student 1 applies to college 2 and all students $l \in \{2, \dots, i\}$ apply only to their safe colleges. Then college 1 remains available after all students in $\{2, \dots, i\}$ have been allocated.
 This gives that colleges $\{1, \dots, i\}$ are collectively safe, college $i + 2$ is safe for student $i + 1 - (A^{i+1})$.
 l^{i+1} follows from theorem 1. (i), and the maximization of the value of information.
 r^{i+1} follows from proposition 1., (ii).
 Combining r^{i+1} with (A^{i+1}) , we get that student $i + 1$ can be assigned to any college in $\{1, \dots, i + 2\}$.
 So μ^{i+1} is also characterized.
 In total, this proves $H(i + 1)$.
- Conclusion: $\forall i \in \{1, \dots, n - 1\}, H(i)$ is verified.
- $i = n$
 By the same argument than above, $\forall l \in \{1, \dots, n - 1\}, H(l)$ implies (A^n) .
 Theorem 1. (ii), implies both l^n and r^n .
 Student n will be allocated to the residual college: by (A^n) , it can be any college. So μ^n is also characterized.

