Essays on Markets and Information

by

Ahmet Arda Gitmez

M.A. Economics, Koç University (2013)
B.Sc. Electrical and Electronics Engineering, Koç University (2011)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© Ahmet Arda Gitmez, MMXIX. All rights reserved.
The author hereby grants to MIT permission to reproduce and to distribute publicly paper
and electronic copies of this thesis document in whole or in part in any medium now known
or hereafter created.

Signature redacted

Author ...........................................

Signature redacted

Certified by…….

Daron Acemoglu
Elizabeth and James Killian Professor of Economics
Thesis Supervisor

Signature redacted

Certified by…….

Parag A. Pathak
Jane Berkowitz Carlton and Dennis William Carlton Professor of Microeconomics
Thesis Supervisor

Signature redacted

Certified by…….

Alexander G. Wolitzky
Associate Professor of Economics
Thesis Supervisor

Signature redacted

Accepted by……

Ricardo Caballero
Ford International Professor of Economics
Chairman, Departmental Committee on Graduate Studies
DISCLAIMER NOTICE

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available.

Thank you.

The images contained in this document are of the best quality available.
Essays on Markets and Information
by
Ahmet Arda Gitmez

Submitted to the Department of Economics
on May 14, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis contains three chapters on the interplay between markets and information.

In the first chapter, I study how market opportunities influence the information revealed in long-term relationships. To this end, I build a model of employment relationships where the worker has private information about match quality, the firm learns about match quality over time, and the firm makes a match-specific investment. Improved market opportunities for workers promote productive relationships because they let the worker signal her firm-specific productivity by forgoing market opportunities. Signaling allows the firm to bypass the learning stage and encourages investment (signaling effect). Improved market opportunities for firms, however, discourage long-term relationships and undermine investment incentives (layoffs effect). I embed the relationship game in a search market equilibrium where market opportunities for both parties depend on search frictions and market thickness. With intermediate values of market thickness, relationship productivity and worker welfare are u-shaped in search frictions: when search frictions decrease from high to intermediate levels, the layoffs effect dominates; when search frictions are sufficiently low, the signaling effect dominates.

In the second chapter, joint with Umut Dur, Parag Pathak and Tayfun Sönmez, I study the effects of enlarging the message space to allow for further information revelation in centralized markets. School districts with choice plans struggle to expand access to schools across neighborhoods while keeping busing costs down. Existing assignment mechanisms allow students to rank a school, but do not elicit preferences about transportation. Typically, if a student is assigned far from home, the district provides transportation. We propose enlarging the message space in the mechanism by allowing students to apply to a school both with and without transportation. Under our proposal, a non-neighborhood applicant who is willing to forgo transportation services obtains a greater chance of being assigned to a school. In decentralized admissions systems, we show that this option reduces transportation but not access for non-neighborhood applicants. We then generalize these results to a centralized assignment mechanism under special conditions. Expanding the message space provides a new tool for distributional objectives that operates in a different fashion than more traditional levers like changing priorities or choice sets.

In the third chapter, joint with Pooya Molavi, I study an information design problem. We present a model of media capture, a politician having control over the editorial policies of media. At the heart of the model is the trade-off faced by a politician who wants to persuade the citizens: she wants to capture the media and produce news in her favor, but capture leads the citizens to not follow the
media as they find them uninformative. The model is a Bayesian persuasion model (à la Kamenica and Gentzkow (2011)) with an audience of heterogeneous priors. We identify conditions on the distribution of priors that guarantee full information revelation and no information revelation by the captured media. The model also has several testable predictions: (i) the information content of the news provided by the captured media decreases as the politician becomes more popular, (ii) in societies with more extremists than moderates, the media are more likely to produce “negative” news than “positive” ones, and (iii) in societies where the media are less accessible to citizens, they are more informative.

Thesis Supervisor: Daron Acemoglu
Title: Elizabeth and James Killian Professor of Economics

Thesis Supervisor: Parag A. Pathak
Title: Jane Berkowitz Carlton and Dennis William Carlton Professor of Microeconomics

Thesis Supervisor: Alexander G. Wolitzky
Title: Associate Professor of Economics
Acknowledgments

I am, and will always be, incredibly grateful to be a part of the MIT Economics community for my late twenties. The last six years of my life did not only help me define who I am as a researcher, but also as a human being. I was fortunate to be surrounded by an abundance of extraordinary role models throughout this transformative process. Daron Acemoglu, Parag Pathak, Alex Wolitzky and Robert Gibbons were always there when I needed support. I was always impressed by Daron’s accessibility for his advisees, Parag’s commitment to the well-being of his colleagues, Alex’s brilliance and rigor in conducting research, and Bob’s ability to see things from a broader perspective. For the rest of my life, I will aspire to achieve the admirable level they have as researchers, advisors and teachers.

Juuso Toikka has always been an admiration for me through my experience in MIT Economics: his genuine interest to hear about my research and his willingness to help me made me feel valued. Every interaction I had with Alp Simsek consistently ranked among the highlights of the day, and I feel blessed to have known him. Abhijit Banerjee was always the epitome of an ideal researcher for me. Kim Scantlebury was always generous with her time and affection. Being coauthors with Umut Dur is always an invaluable experience: his dedication and ingenuity is fascinating. Tayfun Sönmez was influential in guiding me in making a decision to start my PhD at MIT, and was a constant source of support throughout my journey. I remember promising him to pay it forward to future generations of economists, and I will strive to honor that promise for the rest of my career.

Several mentors guided and inspired me to be an economist and gave me the opportunity to pursue a PhD at MIT. Alp Atakan has always been an idol for me too look up to. Özgür Yılmaz and Levent Koçkesen always shared their wisdom at times I most needed them.

Is it possible that maybe, the real PhD is the friends we made along the way? If so, I had the best of all PhD’s: being around Pooya Molavi and Olivier Wang was always entertaining, enlightening and challenging (for the better). I was fortunate to plan movie nights with Kosti Takala, grab a coffee and chat about research with Chishio Furukawa, and hang out with Selman Erol, Alan Olivi, Román Andrés Zárate, Sydnee Caldwell, Ladin Bayurgil, Zeynep Balcioglu, Aytug Sasmaz, Onur Altındag, Tugba Bozcaga and Özgür Bozcaga. Denizcan Vanlı, Seyhmus Güler and Mathilde Poulain were the best roommates ever. My wonderful friends in Turkey were always a reliable source of relief. I also would like to take this chance to mention Ozan Mert Öndeş’s name, in the hope of clearing the debt I owe him since 2011. I would not be able to complete this journey without the inspiration and support I received from Ömer Karaduman and Mert Demirer. I am blessed to have them as my friends, and I will be eternally grateful to them.

I would like to thank my parents and my brother for their unconditional love and support throughout my life. Defne Kirmizi is a terrific human being, she made me a better person, and she made my life awesome by being a part of it.

My coursework at MIT was generously supported by the UniCredit Foundation. I am particularly grateful to Giannantonio De Roni and Annalisa Aleati for the effort they put in keeping in touch with the UniCredit alumni.

This thesis is dedicated to the memory of Cara Nickolaus, a soul we lost too soon. I am blessed to have crossed paths with Cara, and she will continue to be a source of inspiration for me.
# Contents

1 Search, Matching, and Signaling: Do Markets Help Relationships?  
1.1 Introduction ............................................................................. 13  
1.2 The Employment Relationship  
1.2.1 Timing and Notation ......................................................... 17  
1.2.2 Production Function ......................................................... 18  
1.2.3 Strategies and Solution Concept ....................................... 19  
1.3 Analysis of the Employment Relationship  
1.3.1 The Firm's Problem ........................................................ 21  
1.3.2 The Worker's Problem ..................................................... 23  
1.3.3 The Characterization Result ............................................. 26  
1.3.4 Effects of Outside Options ................................................ 29  
1.4 The Search Market ............................................................... 30  
1.4.1 The Economy ................................................................. 31  
1.4.2 Steady-State .................................................................... 32  
1.4.3 Consistency Requirements .............................................. 34  
1.5 Effect of Search Frictions ...................................................... 34  
1.5.1 Implications on Productivity ............................................. 37  
1.5.2 Implications on Worker Welfare ....................................... 38  
1.6 Conclusion ............................................................................ 41

2 Busing with Contracts: Message Spaces as a Design Tool in School Choice  
2.1 Introduction ........................................................................... 43  
2.2 Model .................................................................................... 45  
2.2.1 Primitives ........................................................................ 45  
2.2.2 Priorities and Choice Functions ......................................... 47  
2.2.3 Relationship with Matching with Contracts ......................... 49  
2.2.4 Stability and Strategy-Proofness ......................................... 50  
2.3 Non-Busing Slots .................................................................. 51  
2.3.1 Decentralized Admissions: The One School Case ..................... 52  
2.3.2 Centralized Admissions: Multiple Schools ............................ 53  
2.3.3 Extensions ....................................................................... 55  
2.4 Conclusion ............................................................................ 57

3 Media Capture: A Bayesian Persuasion Approach  
3.1 Introduction ........................................................................... 59  
3.1.1 Relation to the Literature .................................................. 62  
3.2 The Model ............................................................................ 64
3.2.1 Notation and Assumptions ........................................... 64
3.2.2 Discussion of The Model ............................................ 66
3.2.3 The Optimization Problem .......................................... 68
3.3 The Solution .............................................................. 72
  3.3.1 Simplifying the Problem ........................................... 72
  3.3.2 Discarding Interior Solutions .................................... 73
3.4 Characterization Results .............................................. 74
3.5 Comparative Statics ..................................................... 77
  3.5.1 Effect of Popularity ............................................... 77
  3.5.2 Effect of Polarization ............................................ 78
3.6 Cost of Listening to Media ........................................... 80
3.7 Conclusion ............................................................... 82

A Appendix to Chapter 1 .................................................. 83
  A.1 Formal Definition of Equilibrium Concept ....................... 83
  A.2 Proofs in Section 1.3.1 .............................................. 84
    A.2.1 Continuation Decision ......................................... 84
    A.2.2 The Firm’s Continuation Payoffs ............................. 90
    A.2.3 Capital Choice .................................................. 91
    A.2.4 Role of Assumption 1.2 ....................................... 102
    A.2.5 Proof of Proposition 1.1 ..................................... 103
  A.3 Proofs in Section 1.3.2 ............................................. 103
  A.4 Proof of Theorem 1.1 ............................................... 108
  A.5 Proof of Proposition 1.3 ............................................ 111
  A.6 Proof of Lemma 1.2 ................................................ 112
  A.7 Proofs in Section 1.5 ............................................... 113

B Appendix to Chapter 2 .................................................. 121
  B.1 Proof of Proposition 2.3 ............................................ 121
  B.2 Proof of Proposition 2.4 ............................................ 123
  B.3 Irregular Precedence Orders ....................................... 124
  B.4 Omitted Example ..................................................... 125
  B.5 Proof of Theorem 2.1 ............................................... 126
  B.6 Proof of Proposition 2.5 ............................................ 129

C Appendix to Chapter 3 .................................................. 131
  C.1 Proof of Lemma 3.1 ................................................ 131
  C.2 Simplification of Politician’s Problem ......................... 135
  C.3 Proof of Lemma 3.3 ................................................ 135
  C.4 Checking the Boundaries of $\Delta$ ................................ 136
  C.5 Proof of Proposition 3.1 ........................................... 138
  C.6 Proof of Proposition 3.2 ........................................... 139
  C.7 Proof of Proposition 3.3 ........................................... 142
  C.8 Proof of Proposition 3.4 ........................................... 142
  C.9 Proof of Proposition 3.5 ........................................... 143
  C.10 Proof of Proposition 3.6 .......................................... 145
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Illustration of Proposition 1.1</td>
<td>22</td>
</tr>
<tr>
<td>1-2</td>
<td>Illustration of Proposition 1.2</td>
<td>25</td>
</tr>
<tr>
<td>1-3</td>
<td>Illustration of Theorem 1.1</td>
<td>28</td>
</tr>
<tr>
<td>1-4</td>
<td>Illustration of Theorem 1.2</td>
<td>36</td>
</tr>
<tr>
<td>1-5</td>
<td>Market tightness effect</td>
<td>39</td>
</tr>
<tr>
<td>1-6</td>
<td>Match payoff effect</td>
<td>39</td>
</tr>
<tr>
<td>3-1</td>
<td>The intensive versus extensive margin trade-off</td>
<td>71</td>
</tr>
<tr>
<td>3-2</td>
<td>The set of feasible ((\sigma_0, \sigma_1)) pairs</td>
<td>73</td>
</tr>
<tr>
<td>3-3</td>
<td>Optimal policy in a single-dipped distribution</td>
<td>77</td>
</tr>
<tr>
<td>3-4</td>
<td>Making the distribution of priors more polarized</td>
<td>79</td>
</tr>
<tr>
<td>A-1</td>
<td>The value function in Lemma A.1</td>
<td>86</td>
</tr>
<tr>
<td>A-2</td>
<td>Employment dynamics</td>
<td>88</td>
</tr>
<tr>
<td>A-3</td>
<td>Continuation payoffs for the firm</td>
<td>91</td>
</tr>
<tr>
<td>A-4</td>
<td>The objective function in Equation (A.9)</td>
<td>93</td>
</tr>
<tr>
<td>A-5</td>
<td>The maximization problem in Equation (A.14)</td>
<td>95</td>
</tr>
<tr>
<td>A-6</td>
<td>Semi-separating equilibria in Theorem 1.1</td>
<td>109</td>
</tr>
<tr>
<td>A-7</td>
<td>Proof of Theorem 1.2 for intermediate markets</td>
<td>114</td>
</tr>
<tr>
<td>C-1</td>
<td>Illustration of Lemma C.2</td>
<td>133</td>
</tr>
<tr>
<td>C-2</td>
<td>Optimality of a piecewise linear function</td>
<td>134</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Slot Types ................................................................. 48
Chapter 1

Search, Matching, and Signaling: Do Markets Help Relationships?

1.1 Introduction

Building a productive employment relationship is a challenging task. For a relationship to be productive, it needs to have high match quality: the worker’s productivity in the firm depends on her skills in the job and her ability to work with the firm’s technology. Productivity also depends on match-specific capital built in the relationship, which complements match quality: the firm can enhance the worker’s skills by offering her training and installing a more productive technology. Firms’ incentives for such investments depend on the expected length of a relationship. If the firm expects the relationship to be short-lived, it has weaker incentives to build match-specific capital. This reasoning implies that labor market opportunities impair productive relationships. A well-functioning labor market constitutes an attractive outside option for the parties in an ongoing relationship, which may induce quits and layoffs. It thus reduces the length of relationships and curtails investments. The ensuing tension between markets and relationships has led to arguments that markets in the form of outside options can undermine, and sometimes destroy, relationships (Kranton, 1996a; MacLeod and Malcolmson, 1989; Baker et al., 1994; McAdams, 2011; Ramey and Watson, 2001).

This paper argues that attractive market opportunities help to create value in an employment relationship. The employment relationship considered in this paper has asymmetric information about match quality: at the beginning of the relationship, the worker knows more about her compatibility with a job than the firm knows. A well-functioning labor market allows the worker to signal the match quality to the firm by forgoing attractive market opportunities (i.e. other jobs). Because match-specific capital is complementary to match

---

I am indebted to Daron Acemoglu, Robert Gibbons, Parag Pathak and Alex Wolitzky for their guidance and support. I thank Pooya Molavi, Juuso Toikka and Olivier Wang for insightful discussions, and Alp Atakan, Abhijit Banerjee, Alessandro Bonatti, Oğuzhan Çelebi, Mert Demirer, Mehmet Ekmekci, Glenn Ellison, Drew Fudenberg, Chishio Furukawa, Daniel Garrett, Ömer Karaduman and Alp Simsek for helpful comments.

The idea that markets undermine relationships has been discussed extensively outside the field of economics. Among many others, Polanyi (1944), Titmuss (1970) and Sundel (2012) contain arguments that markets destroy relationships by undermining social norms.
quality, such a signal induces the firm to undertake investments and increases productivity. A well-functioning labor market thus acts as a double-edged sword: it both helps and hurts relationships. Indeed, I demonstrate that the effect of market opportunities on relationship productivity may be u-shaped.

To explore this idea, I present a model of a search market where workers and firms look for potential matches and form employment relationships with each other. The employment relationship is a repeated interaction between the firm and the worker; it has three major elements. First, there is an idiosyncratic match quality associated with each employment relationship. A high match quality means that the worker is more productive in the employment relationship. The match quality is idiosyncratic because each job requires a particular combination of skills (Neal, 1995; Gibbons and Waldman, 2004; Lazear, 2009). Second, the firm has the ability to learn about match quality over time. As long as the firm does not lay off the worker, it observes output over time and collects information about the worker’s productivity (Jovanovic, 1979). Third, at the beginning of the relationship, the firm chooses a match-specific capital level. It corresponds to firm-specific training (Becker, 1964) and other activities that enhance the worker’s skills in the job. It is complementary to the match quality: the worker with higher skills benefits more from the training.

Because match quality is idiosyncratic, it is not persistent across relationships. As a result, the worker in a relationship with high match quality prefers to stay in her relationship. Behavior of the worker with low match quality depends on the opportunities in the search market. When market opportunities are attractive, she quits her job and goes back to the search market to find a better match. In such a case, the worker’s willingness to stay in the relationship serves as a signal of the match quality. This signal is valuable for the firm because it resolves the information asymmetry at the beginning of the relationship and allows the firm to bypass the learning stage. Consequently, the firm has incentives to choose a higher level of match-specific capital for the worker who is willing to stay. This intuition is central to the conclusion that improved market opportunities for workers promote productive employment relationships. I denote the positive effect of workers’ market opportunities as the “signaling effect”.

Improved market opportunities for firms have the opposite effect. When information asymmetries are not resolved at the beginning of the relationship, the firm learns about the match quality over time. As its market opportunities improve, the firm becomes more selective upon learning: it terminates workers with low match quality. The relationship thus becomes shorter in expectation: the worker is more likely to be laid off. The firm, then, is less willing to build match-specific capital at the beginning of the relationship. Low levels of match-specific capital render the relationship less productive and induce termination, resulting in a vicious cycle of unproductive and short employment relationships. I denote the negative effect of firms’ market opportunities as the “layoffs effect”.

To demonstrate the contrasting effects of improved market opportunities for informed ver-

---

2 See Topel (1991) and Jacobson et al. (1993) for empirical evidence suggesting that match-specific skills exist.
sus uninformed parties, I analyze the employment relationship between a firm and a worker (Section 1.3). In an employment relationship, outside options are exogenous and common knowledge. I characterize the equilibrium of the employment relationship (Theorem 1.1) and illustrate the fundamental trade-off: productivity in an employment relationship is increasing in the worker’s outside option and decreasing in the firm’s outside option (Proposition 1.3). I then endogenize the outside options in a search market with a continuum of unmatched workers and firms (Section 1.4). The main parameter of interest is the search friction, which determines how quickly a party returns to the search market once its relationship breaks down.

A decrease in search frictions reduces the time one must wait before finding another match and improves the market opportunities for both parties. Which effect dominates depends on market tightness. When the market is loose (i.e. when there are few vacancies compared to unemployed workers in the search market), an unmatched firm finds a match faster than an unmatched worker. In such cases, a decrease in search frictions improves outside options for firms faster than for workers, and the layoffs effect dominates. As a result, a decrease in search frictions impairs the productivity of employment relationships and reduces worker welfare. In contrast, when the market is tight, the signaling effect dominates. In this case, a decrease in search frictions promotes productive employment relationships and increases worker welfare. For markets with intermediate values of tightness, a decrease in search frictions has a nonmonotonic effect on productivity and worker welfare. When the economy moves from high search frictions to intermediate search frictions, the layoffs effect dominates. When search frictions are low, the signaling effect dominates. As a result, productivity and worker welfare are u-shaped in search frictions.

My model shows how outside options generated by market opportunities influence the formation and development of relationships, complementing and extending earlier theories. It is cast in terms of employment relationships, but the insights apply to a wide range of settings with asymmetric information in long-term interactions. In many developing economies, weak law enforcement mechanisms do not provide contractual assurance. This obliges businesses to rely on long-term bilateral relationships (Jagannathan, 1987; McMillan and Woodruff, 2000). Similarly, relationships between buyers from developed countries and sellers from developing countries tend to take the form of long-term partnerships (Egan and Mody, 1992; Macchiavello and Morjaria, 2015). Many business-to-business transactions rely on the threat of termination in repeated interactions to prevent opportunism (Williamson, 1985; Baker et al., 2002). In all of these examples, information asymmetry is an important feature of the relationship. Tangible or intangible investment by the uninformed party is also a prevalent feature of these relationships. My model provides a framework for studying the effects of markets on such long-term relationships.

Related Literature The most important feature of the employment relationship considered in this paper is the idiosyncratic match quality. Jovanovic (1979) is the classic theoretical treatment on learning about match quality. Moscarini (2005) is a model of learning about match quality in a frictional search market, and Felli and Harris (1996) is a model of learning in a search market with perfect recall. Learning in these models is two-sided: initially,
both the firm and the worker have imperfect information about match quality. At the other extreme, there is a strand of literature considering models with perfectly observable idiosyncratic match quality, which do not contain any learning (Pissarides, 1984; Mortensen and Pissarides, 1994). The symmetric information in all these models rules out the possibility of signaling, which is at the heart of my analysis.

The other important feature of the employment relationship in this paper is match-specific investments made by the firm, complementary to match quality. There is extensive literature on incentives for match-specific investments (Grout, 1984; Williamson, 1985; MacLeod and Malcomson, 1993; Hart, 1995). My focus here is the interaction of investment incentives with asymmetric information about match quality. Ramey and Watson (1997) analyzes a model with match-specific investments by the firm and stochastic match quality, which both parties can observe in a given period. Acemoglu (1999) and Acemoglu and Pischke (1998) also consider models where firm investment is complementary to worker productivity. The worker’s productivity in these models takes the form of general human capital without a match-specific component.

There is a growing literature discussing long-term employment relationships with persistent private information. In Chassang (2010), there is private information about the productivity of certain actions. In Halac (2012), there is private information about one party’s outside option. In MacLeod and Malcomson (1988) and Malcomson (2016), the worker has private information about her cost of effort. Sobel (1985), Watson (1999) and Watson (2002) consider models of long-term interactions (not necessarily employment relationships) with one-sided or two-sided private information, where the partners go through an initial trust-building phase in a relationship. These models do not consider investment decisions, and they treat outside options as exogenous.

The influence of outside options on behavior within a partnership has been investigated extensively. Shaked and Sutton (1984) and Binmore et al. (1989) consider this idea in the context of bargaining, Board and Pycia (2014) analyzes the effects of outside options in the context of a durable goods monopolist, Shapiro and Stiglitz (1984) discusses this as the foundation of the efficiency wage theory, and Beaudry and DiNardo (1991) considers and tests a dynamic model of wage adjustments based on outside options. My model derives these outside options from a search market, not unlike those in Kranton (1996b), Ghosh and Ray (1996) and Rob and Yang (2010), which focus on pairwise relationships with private information. McAdams (2011), Fujiwara-Greve and Okuno-Fujiwara (2009) and Datta (1996) are other models of repeated games where the possibility of rematching with another partner serves as an outside option. None of these models considers investment decisions.

The model of a search market in Section 1.4 is similar to the one analyzed in Atakan and Ekmecci (2013), both in terms of the steady-state and the notion of the search friction. Shimer and Smith (2000), Nöldeke and Tröger (2009) and Lauermann (2013) also contain similar notions of steady-state in environments with ex ante heterogeneous agents.
The learning dynamics in my model are similar to those in the models of strategic experimentation. Bolton and Harris (1999), Keller et al. (2005) and Keller and Rady (2015) are important examples and Keller and Rady (2015) is the closest model to mine. Bergemann and Valimaki (2006) provides a review of the literature.

1.2 The Employment Relationship

This section describes the repeated interaction between two matched parties: the firm and the worker.

When the two parties match, the worker privately observes her type, which is a measure of the idiosyncratic match quality and the worker's productivity in the firm. It corresponds to the worker's compatibility with the firm: for instance, when the worker's set of skills are consistent with the particular set of skills required by the firm, match quality is higher. After observing her type, the worker decides whether to join the firm. If the worker decides to join, the firm chooses a level of match-specific capital complementary to worker's type and pays the cost of capital. The capital choice is made only once and is irreversible: it corresponds to the firm-specific training offered by the firm, and the type of job the firm has designed. The worker's type and the capital choice being match-specific is important for the analysis: worker's skills and firm's investment are not carried over to their future relationships with alternative matches, so each relationship can be analyzed in isolation. Capital choice being early and irreversible is also important for the analysis. In reality, the firm has the ability to delay investments and adapt over time, but these are costly alternatives in themselves and I am abstracting away from such possibilities.3

Once the capital is chosen, the worker stochastically produces an output each period, which depends on the worker's type and the capital level. Because the output depends on the worker's type, the firm learns about the type over time by observing the output. At the end of each period, the worker and the firm share the surplus. Then, the firm makes a decision to quit the relationship (i.e. lay off the worker) or continue.

Throughout the relationship, both the worker and the firm have access to an outside option. The outside options accrue to the parties if the worker does not join the firm, or the firm quits. They correspond to payoffs from going to the search market to look for alternative partners, which will be introduced in Section 1.4, but in a relationship they are taken as given. The values of outside options are common knowledge: both the worker and the firm are aware of the market conditions surrounding them. Moreover, since the worker's type denotes idiosyncratic match quality, worker's outside option is independent of her type.

3In the case of training, this assumption captures the fact that it is physically impossible to initiate production without making the investment first.
1.2.1 Timing and Notation

Time is discrete, runs forever, and periods are indexed by $t \in \{0, 1, 2, \ldots \}$. The parties share a common discount factor $\delta \in (0, 1)$.

At time $t = 0$,

- The worker's type $\theta \in \Theta := \{0, 1\}$ is drawn according to:

\[
\theta = \begin{cases} 
1, & \text{w.p. } \phi \\
0, & \text{w.p. } 1 - \phi
\end{cases}
\]

with $\phi \in (0, 1)$. I will refer to the worker with $\theta = 1$ as a "high-type worker", and a worker with $\theta = 0$ as a "low-type worker".

- The worker decides whether to join the firm ($j$) or not ($n$). The decision is denoted by $a^0_w \in A_w := \{j, n\}$.
  - If the worker chooses $a^0_w = n$, the relationship ends. At that point, the worker receives a lifetime discounted payoff of $v > 0$ and the firm receives a lifetime discounted payoff of $\pi > 0$.
  - If the worker chooses $a^0_w = j$, the firm chooses a capital level $k \in \mathbb{R}^+$ and pays the cost of capital $c(k)$. The relationship proceeds to $t = 1$.

At time $t \geq 1$,

- The firm decides whether to quit the relationship ($q$) or continue ($c$). The decision is denoted by $a^t_f \in A_f := \{q, c\}$.
  - If the firm chooses $a^t_f = q$, the relationship ends. At that point, the worker receives $v$ and the firm receives $\pi$.
  - If the firm chooses $a^t_f = c$, the output $y^t$ is drawn i.i.d. conditional on $\theta$ and $k$, and is observed. The firm receives $(1 - \beta)y^t$ and the worker receives $\beta y^t$, where $\beta \in (0, 1)$ is the exogenously given bargaining power of the worker. The relationship proceeds to $t + 1$.

1.2.2 Production Function

Production of output $y^t$ in period $t \geq 1$ requires inputs from both the worker and the firm: labor ($l^t$) and capital ($k$), respectively. Given inputs $(l^t, k)$ in a period, the output is:

\[
y^t := l^t \cdot k
\]

$k \in \mathbb{R}^+$ is chosen at $t = 0$, and $l^t \in \{0, 1\}$ is drawn every period conditional on $\theta$.

- Choosing capital $k \geq 0$ costs $c(k)$ to the firm, and the cost of capital satisfies the usual conditions given below.
Assumption 1.1. \( c(0) = c'(0) = 0, \) \( c(.) \) is strictly increasing and strictly convex, and \( \lim_{k \to \infty} c'(k) = \infty. \)

For the purposes of illustration, I will occasionally be referring to the following cost function which satisfies Assumption 1.1:

\[
c(k) = \frac{k^{1+\Gamma}}{1 + \Gamma} \quad \text{with} \; \Gamma > 0
\]

This is a fairly standard cost function, with \( \Gamma \) parametrizing its convexity. In particular, as \( \Gamma \) gets smaller, the capital choice becomes more responsive to the labor.

- The labor \( l' \) is drawn i.i.d. conditional on \( \theta \), according to:

\[
l' = \begin{cases} 
1 & \text{w.p. } \eta_0 \\
0 & \text{w.p. } 1 - \eta_0
\end{cases}
\]

where \( 1 = \eta_1 > \eta_0 > 0 \). That is, the likelihood of having high labor depends on the worker's type, and in particular, when the worker is high-type, labor input is guaranteed to be high. In the terminology of Keller and Rady (2015), then, low output is conclusive evidence that a worker is low-type. This assumption simplifies the firm's learning problem and is made for its analytical convenience.

Throughout this paper, I will focus on the case where \( \eta_0 \) is "sufficiently small". Intuitively, I will restrict attention to the employment relationships where there is a meaningful difference between a high-type worker and a low-type worker. The alternative case where \( \eta_0 \) is close to one is the less interesting one: in that case, there would be only a minor difference between the two types of workers. The learning dynamics, capital choice, and the participation decision would lose their interesting features.

Assumption 1.2. \( \eta_0 \in (0, \eta^*_0) \) for some \( \eta^*_0 > 0 \).

The calculations in the Appendix explicitly derive the value of \( \eta^*_0 \) required for the results to hold.

1.2.3 Strategies and Solution Concept

Given the worker's type \( \theta \in \Theta \), the worker's strategy specifies a decision to join the firm or nor at \( t = 0 \). Formally, the worker's strategy is:

\[
\sigma_w : \Theta \to \Delta(A_w)
\]

The firm's strategy specifies (i) a capital choice at \( t = 0 \) conditional on the worker joining the firm, and (ii) the decision to quit the relationship or continue at \( t \in \{1, 2, \ldots\} \) conditional
on reaching the relevant history. Due to the exit structure of the game, if the firm makes a
decision at \( t \geq 1 \), the worker joined the firm at \( t = 0 \) and the firm chose to continue at times
\( 1, \ldots, t - 1 \). Therefore, a public history for the firm at time \( t \) is a sequence of the following form:

\[
h_t = \begin{cases} 
(a_w^0 = j) & \text{if } t = 0 \\
(a_w^0 = j, a_f^0 = k) & \text{if } t = 1 \\
(a_w^0 = j, a_f^0 = k; a_f^1 = c, l_1, a_f^2 = c, l_2, \ldots; a_f^{t-1} = c, l^{t-1}) & \text{if } t \geq 2
\end{cases}
\]

Let \( \mathcal{H} \) denote the set of all such sequences. A strategy for the firm is a mapping

\[ \sigma_f : \mathcal{H} \rightarrow \Delta(\mathbb{R}^+) \cup \Delta(A_f) \]

such that \( \sigma_f(h^0) \in \Delta(\mathbb{R}^+) \) and \( \sigma_f(h^t) \in \Delta(A_f) \) for \( t \geq 1 \).

In addition, the firm has a posterior belief about the worker's type following a history \( h^t \in \mathcal{H} \).
It is denoted by:

\[ \mu(h^t) = \Pr\{\theta = 1|h^t\} \in [0, 1] \quad (1.2) \]

Bayesian updating on beliefs imposes that, given a strategy \( \sigma_w \) for the worker and following
a history \( h^0 = (a_w^0 = j) \),

\[ \mu(h^0) = \frac{\phi \cdot \sigma_w(j|\theta = 1)}{\phi \cdot \sigma_w(j|\theta = 1) + (1 - \phi) \cdot \sigma_w(j|\theta = 0)} \quad (1.3) \]

whenever the denominator is nonzero. Moreover, following a history \( h^t = (h^{t-1}; a_f^{t-1} =
c, l^{t-1}) \) with \( t \geq 1 \) and beliefs from the previous period \( \mu(h^{t-1}) \),

\[ \mu(h^t) = \begin{cases} 
\frac{\mu(h^{t-1})}{\mu(h^{t-1}) + (1-\mu(h^{t-1}))\eta_0}, & \text{if } l^{t-1} = 1 \\
0, & \text{if } l^{t-1} = 0
\end{cases} \quad (1.4) \]

The solution concept is **Perfect Bayesian Equilibrium (PBE)**. It is a pair of strategies
\( (\sigma_w^*, \sigma_f^*) \) and beliefs \( \{\mu(h^t)\}_{h^t \in \mathcal{H}} \) such that:

\footnote{Throughout the analysis, I assume that the firm chooses to continue when it is indifferent between \( c \) and \( q \). This restriction simplifies the analysis by ensuring upper hemi-continuity of worker's payoff correspondence.}

\footnote{Equation (1.4) imposes a mild restriction in addition to Bayes' Rule. It requires \( \mu(h^t) = 0 \) following \( l^{t-1} = 0 \), even when \( \mu(h^{t-1}) = 1 \). This contrasts with the "never dissuaded once convinced" assumption made in bargaining games with incomplete information (Osborne and Rubinstein, 1990). Because there is conclusive and verifiable bad news in this model, it is reasonable to assume that the firm would change its mind about the worker type, even though its initial beliefs rules the worker being low type out.}

\footnote{See Appendix A.1 for a formal definition.
Given \((\sigma_0^f), \sigma_w^*\) maximizes the worker’s payoff for all \(\theta \in \Theta\).

- Given \((\sigma_w^*)\) and \(\{\mu(h^t)\}_{h^t \in \mathcal{H}}\), \(\sigma^*_f\) maximizes the firm’s payoff for all \(h^t \in \mathcal{H}\).
- \(\{\mu(h^t)\}_{h^t \in \mathcal{H}}\) satisfies (1.3) and (1.4).

## 1.3 Analysis of the Employment Relationship

In this section, I characterize the PBE of employment relationship and discuss its implications. I start by characterizing the firm’s capital choice and its continuation decisions in the following periods (Section 1.3.1). Then, I characterize the participation decision by the worker (Section 1.3.2). In Section 1.3.3, I combine the two pieces. Section 1.3.4 discusses the comparative statics that follow from the characterization result and discusses the critical features of the model.

### 1.3.1 The Firm’s Problem

This section characterizes the firm’s optimal strategy following a history of participation by the worker \(h^0 = (a^0_w = j)\) and beliefs \(\mu^0 := \mu(h^0)\). Given \(\mu^0\) and the outside option \(\pi > 0\) of the firm, there are two types of decisions the firm faces:

- The capital choice in period \(t = 0\), denoted by \(k^*(\mu^0, \pi) = \sigma^*_f(h^0)\).
- The continuation decision in periods \(t \geq 1\), denoted by \(\sigma^*_f(h^t)\).

The key insight from the firm’s problem is: given \(\mu^0\), an increase in \(\pi\) leads to a higher probability of quitting by the firm and a lower level of capital in equilibrium. To build the intuition, consider the case \(\mu^0 < 1\). Even though the firm’s beliefs are not degenerate, it can learn about the worker’s type by observing the output over time. Indeed, as long as the worker keeps producing a high output, the firm’s belief about the worker’s type increases. If a low output is observed, the firm conclusively learns that the worker is low-type. When \(\pi\) is sufficiently low, the firm does not quit in such a case: its outside option is not attractive, so even if the worker is low-type the firm does not lay off the worker. In this case, regardless of the output, the relationship lasts forever. When \(\pi\) increases above a certain value, the firm quits the relationship with a low-type worker. In this case, any relationship with the low-type worker is terminated in finite time, and the relationship lasts shorter. If \(\pi\) is sufficiently high, the firm’s outside option is too attractive: it immediately quits without observing any output.\(^7\) As a result, an increase in \(\pi\) leads to a shorter relationship.

When choosing the capital level, the firm knows about its own continuation behavior. When \(\pi\) is low, therefore, the firm knows that the relationship will last forever, and chooses the capital level accordingly. For intermediate values of \(\pi\), the firm knows that with probability \(1 - \mu^0\), the relationship will be terminated in finite time and capital will be sunk. Taking this into account, the firm chooses a relatively lower level of capital. For high values of \(\pi\),

\(^7\)Such a high value of outside option is never realized in the model of search market where outside options are endogenized. Nevertheless, to have a complete characterization, I am not ruling this case out throughout this section.
the firm knows that it will quit immediately, and does not invest in the worker at all: it sets \( k = 0 \). As a result, the capital choice decreases with \( \pi \) because the relationship lasts shorter in expectation. The following result formalizes this discussion.

**Proposition 1.1.** Consider a history \( h^0 = (a_w^0 = j) \) with beliefs \( \mu^0 := \mu(h^0) \in [0, 1] \). There are values \( \pi_1(\mu^0) \) and \( \pi_2(\mu^0) \), with \( \pi_2(\mu^0) > \pi_1(\mu^0) > 0 \), such that the firm’s capital choice is given by:

\[
\begin{align*}
 k^*(\mu^0, \pi) &= \begin{cases} 
 k(\mu^0), & \text{if } \pi < \pi_1(\mu^0) \\
 \{k(\mu^0), k(\mu^0)\}, & \text{if } \pi = \pi_1(\mu^0) \\
 k(\mu^0), & \text{if } \pi \in (\pi_1(\mu^0), \pi_2(\mu^0)) \\
 \{k(\mu^0), 0\}, & \text{if } \pi = \pi_2(\mu^0) \\
 0, & \text{if } \pi > \pi_2(\mu^0)
\end{cases}
\end{align*}
\]

Here, \( \overline{k}(\mu^0) > k(\mu^0) > 0 \) for all \( \mu^0 \in [0, 1] \) and \( \overline{k}(1) = k(1) \). Moreover, both \( \overline{k}(\mu^0) \) and \( k(\mu^0) \) are strictly increasing in \( \mu^0 \).

When \( k^*(\mu^0, \pi) = \overline{k}(\mu^0) \), the firm never quits in the continuation game. When \( k^*(\mu^0, \pi) = k(\mu^0) \), the firm does not quit as long as high output is realized, and immediately quits following a low output. When \( k^*(\mu^0, \pi) = 0 \), the firm immediately quits without observing any output.

**Proof.** See Appendix A.2.5.

Figure 1-1 below illustrates Proposition 1.1.

![Figure 1-1: Illustration of Proposition 1.1.](image)
Note that the firm is indifferent between at most two capital levels given $\mu^0$. Even though the
firm has a static optimization problem with convex cost of capital, having multiple optimal
capital levels is possible. This is because there is a complementarity between the capital level
and the firm’s continuation incentives: a high capital level increases the firm’s continuation
probability in the following periods. For $\pi \in \{\pi_1(\mu^0), \pi_2(\mu^0)\}$, this complementarity leads to
different levels of optimal capital choice and continuation behavior.

The cutoff structure in Proposition 1.1 makes it clear that a higher outside option for the
firm reduces relationship length. Here, $\pi_1(\mu^0)$ corresponds to the value of $\pi$ which induces
the firm to quit after learning that the worker is low-type, when capital is $k(\mu^0)$. Below
this value, there are no layoffs. $\pi_2(\mu^0)$ is the outside option which induces the firm to quit
without observing any output, when capital is $k(\mu^0)$. Above this value, the firm quits im-
mediately. In between, the firm does not immediately quit, but there are layoffs following
low output in equilibrium.\footnote{Assumption 1.2 plays a crucial role in this characterization. Without a meaningful difference between low-
type and high-type workers (i.e. with high $\eta_0$), the region with layoffs disappears. In this case, one would have
$\pi_1(\mu^0) = \pi_2(\mu^0)$ for all $\mu^0$. Appendix A.2.4 discusses the role played by Assumption 1.2 in more detail.}

$\bar{k}(\mu^0) \geq k(\mu^0) > 0$ implies that an increase in the firm’s outside option reduces match-
specific options. Proposition 1.1 therefore has a pessimistic implication on the effect of
outside options: given a selection of workers, a high outside option for the firm leads to
shorter and less productive relationships. Note, however, that the discussion so far has been
about the firm’s outside option. The worker’s outside option changes the selection of workers
and induces different levels of capital. The next subsection considers the worker’s problem
and characterizes the selection given the firm’s optimal strategy.

1.3.2 The Worker’s Problem

This section characterizes the worker’s participation decision at $t = 0$. At this stage of the
game, worker knows her type, and given Proposition 1.1, knows her continuation payoffs
from joining the firm in a PBE. The continuation payoffs depend on the firm’s beliefs, which
are in turn determined based on the participation decisions. In this subsection, I will ignore
the dependence of the beliefs on the participation decision and analyze the participation de-
cision given a fixed level of firm’s beliefs. In Section 1.3.3, I will explicitly derive the beliefs
based on the participation decision.

Fix $\mu^0 = \mu(h^0)$. Given $\mu^0$, outside options $(\pi, v)$ and her type $\theta$, the worker decides whether
to join the firm or not. This decision is given by the worker’s optimal strategy:

$$\sigma_w^*(j|\theta) \in [0, 1] \quad \text{for } \theta \in \{0, 1\}$$

I will begin by demonstrating that the high-type worker has higher incentives to join than the
low-type worker in equilibrium. This is due to two underlying forces of the model. First, the
multiplicative production function implies that given a positive capital level, the high-type
worker has higher payoff in any period where the firm continues. Second, the firm’s optimal
strategy in Proposition 1.1 implies that the firm continues with higher probability when the
worker is high-type. Combination of these two forces imposes a version of single-crossing
condition for the workers, formalized below.

**Lemma 1.1. (Single-Crossing.)** When the firm uses its optimal strategy given \( \mu^0 \),
\[
\sigma_w^*(j|\theta = 0) > 0 \implies \sigma_w^*(j|\theta = 1) = 1
\]

*Proof. See Appendix A.3.*

The next result demonstrates that, when outside option for the worker is sufficiently high,
the two types have different optimal strategies. That is, the workers *separate* in joining
the firm. The reasoning is simple: Given Lemma 1.1, for separation to occur, the low-type
worker must prefer not join the firm. This is the case only when the outside option is suffi-
ciently attractive compared to the low-type worker’s payoff from joining the firm.

**Proposition 1.2.** Suppose the firm uses its optimal strategy given \( \mu^0 \). There are functions
\( v_1(\mu^0, \pi) \) and \( v_2(\mu^0, \pi) \), with \( v_2(\mu^0, \pi) > v_1(\mu^0, \pi) \) for all \( \pi \leq \pi_2(\mu^0) \), such that:

- \( \sigma_w^*(j|\theta = 0) = \sigma_w^*(j|\theta = 0) = 1 \) if and only if
  \[
  \pi \leq \pi_2(\mu^0) \\
  v \leq v_1(\mu^0, \pi)
  \]

- \( \sigma_w^*(j|\theta = 0) \in (0, 1) \) and \( \sigma_w^*(j|\theta = 0) = 1 \) if and only if
  \[
  \pi \leq \pi_2(\mu^0) \\
  v = v_1(\mu^0, \pi) \quad \text{and} \quad \pi = \pi_1(\mu^0) \\
  v \in (v_1(\mu^0, \pi), v_2(\mu^0, \pi)) \quad \text{and} \quad \pi = \pi_2(\mu^0) \\
  v \in (0, v_1(\mu^0, \pi))
  \]

- \( \sigma_w^*(j|\theta = 0) = 0 \) and \( \sigma_w^*(j|\theta = 0) = 1 \) if and only if
  \[
  \pi \leq \pi_2(\mu^0) \\
  v \in [v_1(\mu^0, \pi), v_2(\mu^0, \pi)]
  \]

*Proof. See Appendix A.3.*

Below is an illustration of Proposition 1.2.
Neither Type Joins

Only High Type Joins

Both Types Join

Figure 1-2: Illustration of Proposition 1.2.

The $v_1(\mu^0, \pi)$ curve corresponds to the values of $v$ that leave the low-type worker indifferent between joining the firm or not. Below this curve, the outside option is not attractive enough and both types join the firm. The $v_2(\mu^0, \pi)$ curve is the values of $v$ that leave the high-type worker indifferent. Above this curve, the outside option is too attractive and neither type joins. In between, only the high-type worker joins. $v_2(\mu^0, \pi) > v_1(\mu^0, \pi)$ for any $\pi < \pi_2(\mu^0)$ because the high-type worker receives a higher payoff from joining when the firm chooses a positive level of capital. Both curves has a discontinuity at $\pi = \pi_1(\mu^0)$ because the firm’s strategy changes discontinuously at $\pi = \pi_1(\mu^0)$. When $\pi > \pi_1(\mu^0)$, the firm chooses a lower level of capital, which reduces the incentives to join the firm. As a result, the discontinuity takes the form of a downward jump.

One remark about Proposition 1.2 is: for a given $\mu^0$, the low-type worker is indifferent on a very limited set of $(\pi, v)$ pairs. This corresponds to the blue curve plotted in Figure 1-2. In the horizontal parts of the curve, the firm sets the optimal level of capital as given in Proposition 1.1. In the vertical parts of the curve, the firm mixes between two levels of capital, between which it is indifferent by Proposition 1.1. The proof in Appendix A.3 explicitly constructs the strategies which leave the low-type worker indifferent.

Of course, Proposition 1.2 offers an incomplete picture, because it takes $\mu^0$ as given, whereas $\mu^0$ depends on the participation decision. The next subsection completes the analysis by taking into account that $\mu^0$ is pinned down by the Bayes’ rule.

---

9 Such a value of outside option is never realized in the search market when outside options are endogenized.
1.3.3 The Characterization Result

This section brings Propositions 1.1 and 1.2 together. The final piece not used so far is the Bayes’ Rule in Equation (1.3). Given $\sigma_w^* = \{\sigma_w^*(\theta)\}_{\theta \in \Theta}$ it gives the firm’s beliefs:

$$
\mu^0 := \mu(h^0) = \frac{\phi \cdot \sigma_w^*(j|\theta = 1)}{\phi \cdot \sigma_w^*(j|\theta = 1) + (1 - \phi) \cdot \sigma_w^*(j|\theta = 0)}
$$

(1.5)

The key insight of the analysis is: high $v$ creates a signaling opportunity for the high-type worker and induces a high capital choice by the firm. By the analysis of the worker’s problem, when $v$ is higher, the low-type worker joins the firm with lower probability. By Equation (1.5), this leads to a higher $\mu^0$, i.e. a better selection of workers for the firm. A better selection induces a higher level of match-specific capital. Intuitively, because capital is complementary to labor, the firm chooses a higher level of capital in response to a better selection of workers. This can be observed by the firm’s problem, and the fact that both $k(\mu^0)$ and $\mu^0$ are increasing in $\mu^0$.

The early stages of the employment relationship resembles a signaling game. Here, the worker acts as the sender, the firm acts as the receiver, and joining the firm acts as the message that the worker’s type is high. To make the analogy more apparent, I will adopt the well-known taxonomy of the PBE for signaling games: there may be a pooling equilibrium, a separating equilibrium or a semi-separating equilibrium. Lemma 1.1 implies that, when at least one type joins with positive probability, the equilibrium must be one of the following:10

1. A **pooling equilibrium** is the one where $\sigma_w^*(j|\theta = 0) = \sigma_w^*(j|\theta = 1) = 1$.

2. A **semi-separating equilibrium** is the one where $\sigma_w^*(j|\theta = 0) \in (0, 1)$ and $\sigma_w^*(j|\theta = 1) = 1$.

3. A **separating equilibrium** is the one where $\sigma_w^*(j|\theta = 0) = 0$ and $\sigma_w^*(j|\theta = 1) = 1$.

One observation that needs to be made at this point is: for sufficiently high values of $\pi$,11 there is a PBE where neither type joins the firm. Such a PBE is supported with off-path beliefs $\mu^0 = 0$ and (off-path) capital choice $k = 0$. The workers, believing that the firm will be pessimistic about their types, do not join the firm, and since there is no participation on the equilibrium path, the Bayes’ rule does not have a bite and the firm is allowed to have pessimistic beliefs off the equilibrium path. Consequently, there is no production in this PBE. To rule out this case, I will focus on a **PBE with positive participation** whenever it exists, defined as a PBE with:

$$
\phi \cdot \sigma_w^*(j|\theta = 1) + (1 - \phi) \sigma_w^*(j|\theta = 0) > 0
$$

When it exists, a PBE with positive participation Pareto dominates the PBE where the workers do not join. An immediate implication of Lemma 1.1 is: in any PBE with positive

10 There is, potentially, a fourth type of semi-separating equilibrium with $\sigma_w^*(j|\theta = 1) \in (0, 1)$ and $\sigma_w^*(j|\theta = 0) = 0$. Such an equilibrium occurs only in a knife-edge case of parameters when $v = \frac{1}{1-\beta} k(1)$, and whenever it occurs, a Pareto-dominating PBE where the high-type worker participates also exists.

11 Formally, when $\pi > \pi_2(0)$. 

26
participation, the firm’s beliefs upon participation are:

\[
\mu^0 = \frac{\phi \cdot \sigma_w^*(j|\theta = 1)}{\phi \cdot \sigma_w^*(j|\theta = 1) + (1 - \phi) \cdot \sigma_w^*(j|\theta = 0)} = \frac{\phi}{\phi + (1 - \phi) \sigma_w^*(j|\theta = 0)} \geq \phi
\]

The PBE that invokes \( \mu^0 = 1 \) is the separating equilibrium, the PBE that invokes \( \mu^0 = \phi \) is the pooling equilibrium, and the PBE that invokes \( \mu^0 \in (\phi, 1) \) is a semi-separating equilibrium. The characterization of PBE follows from repeated applications of Proposition 1.2 twice, for \( \mu^0 = \phi \) and \( \mu^0 = 1 \).

- Let \( v(\pi) := v_1(\phi, \pi) \). By Proposition 1.2, these are the values of \( v \) which leave the low-type worker indifferent between joining or not when the firm uses its optimal strategy under pooling. Consequently, there is a pooling equilibrium below this curve.
- Let \( \overline{v}(\pi) := v_1(1, \pi) \). By Proposition 1.2, these are the values of \( v \) which leave the low-type worker indifferent between joining or not when the firm uses its optimal strategy under separation. Consequently, there is a separating equilibrium above this curve.

Below is the main result of this section.

**Theorem 1.1.** A PBE with positive participation exists if and only if

\[
\pi < \pi^* := \pi_2(1)
\]

\[
v < v^* := v_2(1, \pi_2(1))
\]

When \( \pi \leq \pi^* \):

- If \( v \in [\overline{v}(\pi), v^*] \), there is a separating equilibrium. In the separating equilibrium,
  - The firm chooses \( \overline{k}(1) \) as the capital level. On the equilibrium path, the worker always produces a high output and the firm never quits the relationship.

- If \( v \leq v(\pi) \) and \( \pi \leq \pi_2(\phi) \), there is a pooling equilibrium. In this case:
  - If \( \pi \leq \pi_1(\phi) \), the firm chooses \( \overline{k}(\phi) \) as the capital level. The firm never quits the relationship regardless of the output produced.
  - If \( \pi \in [\pi_1(\phi), \pi_2(\phi)] \), the firm chooses \( k(\phi) \) as the capital level. The firm continues the relationship as long as a high output is produced, and immediately quits following a low output.

- Otherwise, the PBE with positive participation is a semi-separating equilibrium.

**Proof.** See Appendix A.4. \( \square \)

The following is an illustration and description of Theorem 1.1.
In Figure 1-3:

- **Region A** is the region where $v$ is high compared to the low-type worker's payoff from joining the firm. In this region, the low-type worker takes her outside option and the workers separate ($\mu^0 = 1$). The firm sets $k(1) = k(1)$ in this equilibrium and continues as long as a high output is produced. Since this is a separating equilibrium, a low output is never produced: on the equilibrium path, the worker always produces a high output and the firm never quits.

- **Regions B and C** are the regions where $v$ is low compared to the low-type worker’s payoff from joining the firm. As a result, both types join and there is a pooling equilibrium ($\mu^0 = \phi$).
  - In Region B, the firm’s outside option is also low. Therefore, the firm never quits even after learning about the worker’s type: there is a *pooling equilibrium without layoffs* in this range. The firm sets $\tilde{k}(\mu^0)$, which is strictly lower than $\tilde{k}(1)$ because the selection of workers is not achieved.
  - In Region C, the firm’s outside option is relatively high. Thus, the firm quits if it learns that the worker is low-type: there is a *pooling equilibrium with layoffs* in this range. The firm sets $k(\mu^0)$, which is strictly lower than $k(1)$ because the relationship is shorter.

- **Region D** is the region where $v$ is not high enough to achieve separation and not low enough to induce pooling. In this region, the low-type worker joins with some
probability so that \( \mu^0 \in (\phi, 1) \) and \( \pi \in \{\pi_1(\mu^0), \pi_2(\mu^0)\} \). The capital choice and quitting decision following a low output is characterized in Appendix A.4, which is omitted here for brevity. In this region, the participation probabilities and the capital choice change continuously. Qualitatively, the equilibria in this region “connects” the equilibria in Regions A, B and C.

It should also be noted that whenever a PBE with positive participation exists, it is generically unique.\(^{12}\) For instance, the interior of regions marked in Figure 1-3 have a unique PBE with positive participation, whereas in the boundaries that separate these regions, there are two PBE with positive participation.

### 1.3.4 Effects of Outside Options

Theorem 1.1 yields an important insight: when the worker’s outside option is sufficiently high, there is a separating equilibrium regardless of the firm’s outside option. It is only when the worker’s outside option is low that the firm’s outside option matters.

Having a high outside option for the worker is crucial in inducing a high capital choice: it does not only achieve positive selection, but also increases the match-specific capital. When separation occurs \( (\mu^0 = 1) \), the firm chooses the highest level of capital among all PBE with possible participation. If the worker’s outside option is low so that separation is not achieved \( (\mu^0 < 1) \), an increase in the firm’s outside option has the opposite implication: it reduces the capital level chosen by the firm because it shortens the relationship.

The comparative statics regarding the total production in a relationship exhibit these contrasting effects.

**Proposition 1.3.** Given a pair of outside options \((\pi, v)\), take the behavior in the PBE \((\sigma^*_w, \sigma^*_f)\). Define the lifetime discounted output in an employment relationship as:

\[
Y(\pi, v) := \mathbb{E}_{\theta, h^t} \left[ \sigma^*_w(j|\theta) \sum_{t=1}^{\infty} \delta^t \sigma^*_f(c|h^t)y^t \right]
\]

When \( \pi \leq \pi_1(\phi) \), for \( v \in (0, v^*] \),

\[
\frac{\partial Y(\pi, v)}{\partial v} \geq 0
\]

When \( v \leq v(\pi_2) \), for \( \pi \in (0, \pi^*] \),

\[
\frac{\partial Y(\pi, v)}{\partial \pi} \leq 0
\]

\(^{12}\)Formally, this means that the set of \((\pi, v)\) pairs where there is a unique PBE with positive participation is a countable collection of dense open sets, whereas the set of \((\pi, v)\) pairs where there are multiple PBE with positive participation is a nowhere dense set.
Proof. See Appendix A.5.

In words, starting in Region B in Figure 1-3 and moving upwards increases the total production in a relationship by inducing signaling opportunities for the high-type worker (the signaling effect).\(^{13}\) In contrast, starting in Region B and moving rightwards decreases the total production by reducing relationship length (the layoffs effect).

Asymmetric information about the match quality is the main driving force behind these contrasting effects. To illustrate, consider an alternative version of the model where the worker also has incomplete information about her type. As it is the case with the firm, the worker believes that she is high-type with probability \(q\), and she learns about her type by joining the firm and observing the output. This would be the Jovanovic (1979) model, with the additional element that the firm makes a capital choice. Under such a model, Proposition 1.1 continues to hold. Nevertheless, because the worker has incomplete information, any PBE in this model is restricted to be a pooling equilibrium. In such a model, an increase in both the firm's and the worker's outside option has a negative effect.

The main idea of the analysis is so far is: an attractive outside option for the worker ensures that participating to the firm is a signal of productivity within the firm. To emphasize this mechanism, I assumed away other possible forms of signaling, such as backloading payments, bonding, or money burning by workers (Carmichael and MacLeod, 1997; Camerer, 1988). One reason for such an assumption is that limited liability and budget constraints may not always allow for such signaling opportunities. Yet, there may be a richer interaction between the existence of such opportunities and the setup presented here. Such an analysis warrants further investigation.

1.4 The Search Market

The analysis in Section 1.3 characterizes the equilibrium in an employment relationship, taking outside options as given. It is natural to expect that these outside options are generated from setting up alternative employment relationships with new partners that workers and firms find in a search market. This section formulates the model of such a search market and identifies the steady-state equilibrium by imposing two requirements:

- The flow rate of matches depend on the market tightness, which is determined by the turnover in the prevailing equilibrium.
- Conditional on finding a match, the payoff in a relationship is the ex ante payoff in the prevailing PBE.

The outside options \((\pi, \upsilon)\) depend on the flow rate of new matches in the search market and the payoffs obtained in these matches.

\(^{13}\)There is the negative effect of lower participation by the low-type worker, which reduces total production. Under Assumption 1.2, the positive effect of higher capital choice outweighs this force.
1.4.1 The Economy

Consider a dynamic economy populated with a continuum of risk neutral firms and workers. Time is discrete and agents discount the future with the common discount factor $\delta \in (0, 1)$.

At the beginning of each period, a unit measure of firms looking for workers and a measure $L > 1$ of workers seeking jobs enter a search market. As in Section 1.3, the matching is one-to-one: a firm can employ only one worker and a worker can work for only one firm at a given time.

In any given period, there is a measure $\nu \geq 1$ of vacancies and a measure $u \geq L$ of unemployed workers in the search market. A random matching function generates the maximum potential number of matches:

$$x(u, \nu) = \min\{u, \nu\}$$  \hspace{1cm} (1.6)

This is an example of a linear matching technology as considered in Shimer and Smith (2001) and Nödeke and Tröger (2009), and is assumed here due to its analytical convenience.\(^{14}\)

Once a firm and worker are matched, they play the game discussed in the previous section. In equilibrium, an employment relationship does not necessarily last forever: the worker may choose not to join the firm, or the firm may choose to quit the relationship in a later period, which results in the relationship breaking down. If a relationship breaks down, both the firm and the worker return to the pool of unmatched parties after $K > 1$ periods. $K$ is the crucial parameter of interest in this section: it imposes a cost of breaking down a relationship\(^{15}\) and proxies for the magnitude of search frictions, as in Atakan and Ekmekci (2013). $K = 1$ corresponds to the case where breaking down a relationship does not impose any cost: the parties can return to the pool immediately in the next period. A large $K$ implies that it is costly to break up relationships and look for alternative partners.

At the end of a period, if an agent is unmatched, she dies with exogenous probability $s \in (0, 1)$. This assumption is made to ensure the existence of a steady-state in the search market, as without it the market size may grow unboundedly.\(^{16}\)

\(^{14}\) One could modify the matching function so that it is parametrized by an efficiency parameter $\lambda \in [0, 1]$, and assume $x(u, \nu) = \lambda \min\{u, \nu\}$. In this case, the results would continue to hold with slight modifications of expressions. In particular, I am assuming $\lambda = 1$ to abstract away from questions regarding the efficiency of the matching function, as my main parameter of interest $K$ yields a notion of search friction already. A more general matching technology is analytically less convenient to work with in this setup, but I conjecture that as long it satisfies Constant Returns to Scale (Diamond, 1982; Mortensen, 1982; Pissarides, 1990; Mortensen and Pissarides, 1994), the findings should go through.

\(^{15}\) An alternative modeling strategy is considering an additive linear cost of breaking down a relationship. The form adopted here is chosen because it is useful in obtaining closed-form solutions, but the general conclusions would continue to hold under additive cost.

\(^{16}\) As in Atakan and Ekmekci (2013), an agent does not die when she is in a relationship or when she is waiting to go back to the market after breaking down a relationship. This has the advantage of simplifying the steady-state calculations in Section 1.4.2.
1.4.2 Steady-State

Throughout the analysis, I will focus on the steady-state of this economy. As usual, it is defined as the measure of unemployed workers and the vacancies which are constant over time:

\[(u, \nu) \in [L, \infty) \times [1, \infty)\]

Let \( \tau := \xi \) denote the \textit{market tightness} in the steady-state. The market tightness acts as a sufficient statistic for the flow rates of matches. That is, letting

\[q(\tau) := \frac{x(u, \nu)}{\nu} = \min\{u, 1\}\]

a vacancy is matched with an unemployed worker with probability \( q(\tau) \) each period in the steady-state. Similarly, an unemployed worker is matched with probability \( \tau q(\tau) = \min\{\xi, 1\} \).

The pool of unemployed workers in the steady-state consists of three groups:

1. The newborn workers (measure \( L \)),
2. The workers who were unemployed in the previous period, who did not find a match and who did not die (measure \( u(1 - \tau q(\tau))(1 - s) \)),
3. The workers whose matches broke down \( K \) periods earlier and who returned to the market.

The last quantity depends on the PBE that prevails in each partnership, \( (\sigma^*_w, \sigma^*_f) \). For the market size in the steady-state, what matters is the parts of the strategies that determine the turnover in a relationship:

- The participation probability by the low-type worker: \( \sigma^*_w(j|\theta = 0) \in [0, 1] \), and,
- The probability that the firm continues the relationship after learning that the worker is low-type: \( \zeta^* := \sigma^*_f(c|h') \in [0, 1] \) where \( h' \) is a history with at least one \( l' = 0 \) for some \( t < t \).

Given these variables, the probability that a match between a firm and a worker lasts forever, i.e. the probability that a match is \textit{consummated} is:

\[\rho^* := \phi + (1 - \phi) \cdot \sigma^*_w(j|\theta = 0) \cdot \zeta^* \in [\phi, 1] \tag{1.7}\]

If a low-type worker does not participate, the relationship breaks down immediately. Otherwise, because \( \eta_0 < 1 \), the low-type worker produces a low output in finite time with probability one. If if \( \zeta^* = 0 \), then, the relationship will break down with probability one in finite time.

Given \( \rho^* \), a measure \( u \tau q(\tau)(1 - \rho^*) \) of workers matched in a given period eventually break up

\[\text{In any PBE with positive participation, high type always participates and the firm never quits following a high output, so these variables are pinned down.}\]
and return to the market, and a measure \(u r q(\tau) \rho^*\) of matched workers leave the search market forever. In the steady-state, the flow into unemployment and flow out of unemployment must be equal. Therefore, \(u\) must satisfy:\(^{19}\)

\[
\frac{L}{\text{flow into employment}} = \frac{u \cdot \tau q(\tau) \cdot \rho^* + u \cdot (1 - \tau q(\tau)) \cdot s}{\text{consummated matches deaths}}
\]  

(1.8)

A similar reasoning applies to the vacancies as well, so \(\nu\) must satisfy:

\[
\frac{1}{\text{flow into vacancies}} = \frac{\nu \cdot q(\tau) \cdot \rho^* + \nu \cdot (1 - q(\tau)) \cdot s}{\text{consummated matches deaths}}
\]  

(1.9)

The following Lemma proves the existence of a steady-state \((u, \nu)\) and shows that it can be calculated directly as a function of \(\rho^*\).

**Lemma 1.2.** Given \(\rho^* \in [0, 1]\), in the steady-state,

\[
u = \frac{1}{\rho^*}
\]

\[
u = \frac{L - 1}{s} + \frac{1}{\rho^*}
\]

Consequently, \(q(\tau) = 1\) and \(\tau q(\tau) = \frac{s}{u} = \frac{s}{(L - 1)\rho^* + s}\).

**Proof.** See Appendix A.6. \(\Box\)

Lemma 1.2 demonstrates a crucial feature of the search market: in any steady state, the firms are in the short side of the market and any vacancy immediately finds a match. It is an implication of the assumption that \(L \geq 1\): in each period more workers than firms are born into the market. As a result, there are more unemployed workers than vacancies in the steady-state.

Another implication of Lemma 1.2 is: a worker's flow rate of matches is decreasing in \(L\) and \(\rho^*\). As \(L\) increases, there are more unemployed workers in the steady-state and the competition among them intensifies. Thus, each unemployed worker finds a vacancy at a slower rate. As \(\rho^*\) increases, there is less turnover in a given relationship. Thus, there are fewer matched firms whose matches break down. This leads to fewer firms returning to the search market, fewer vacancies, and a lower rate of finding a vacancy.
1.4.3 Consistency Requirements

In an economy in the steady-state \((u, v)\), there is a pair of outside options \((\pi, v)\) generated by the market conditions. These outside options induce the unique PBE characterized in Theorem 1.1. Let the ex ante expected payoff of a worker from the PBE be denoted by:

\[ \Sigma_w(\pi, v) \]

and the ex ante expected payoff of a firm is:

\[ \Sigma_f(\pi, v) \]

Let \(J^U\) denote the net present value of being unemployed for a worker in the steady-state. By definition, then, \(J^U\) must satisfy:

\[ J^U = \tau q(\tau) \Sigma_w(\pi, v) + (1 - \tau q(\tau))\delta(1 - s)J^U \]

where \(\tau = \frac{e}{u}\) is the market tightness in the steady-state. Similarly, the net present value of being a vacancy is:

\[ J^V = q(\tau) \Sigma_f(\pi, v) + (1 - q(\tau))\delta(1 - s)J^V \]

The consistency requirement is that \(v\) and \(\pi\) are the payoffs from breaking down a relationship and going back to the search market. In other words, the outside option is the market opportunity. In this setup, breaking down a relationship implies waiting for \(K\) periods for returning to the market. Therefore:

\[ v = \delta^K \cdot J^U \quad \pi = \delta^K \cdot J^V \]

The equations presented in this section describe a procedure towards investigating whether a particular PBE exists given the level of search frictions \(K\). Given the behavior in the PBE, Lemma 1.2 and Equations (1.10)-(1.12) can be solved to derive \((\pi, v)\). Then, it remains to verify that the \((\pi, v)\) invokes the PBE under consideration, which can be done by appealing to Theorem 1.1.

1.5 Effect of Search Frictions

This section considers the model of the search market outlined in Section 1.4 and derives the type of PBE that prevails under a particular value of \(K\) in the steady-state. For expositional simplicity, I will only focus on the pure strategy PBE defined in Theorem 1.1 throughout this section.\(^{20}\) These are:

\(^{20}\)Consequently, the characterization result will contain the regions that may not collectively add up to the whole range of \(K\). It should be kept in mind that in the remaining regions, there are the semi-separating equilibria described in Appendix A.4. These equilibria change continuously with \(K\) such that the indifference condition for the workers and the firm is satisfied. Quantitatively, they "connect" the pure strategy equilibria listed here.
The pooling equilibrium without layoffs,

- The pooling equilibrium with layoffs, and,

- The separating equilibrium.

By Lemma 1.2, \( L \) plays a crucial role in determining the flow rate for workers. This translates into a crucial role in determining the worker's outside options. As a result, \( L \) plays an important role in determining the PBE as well. The result provided below characterizes the interplay between search frictions and the prevailing PBE for different values of \( L \).

**Theorem 1.2.** There are \( \bar{L} > L > 1 \) such that the following holds.

(Thick markets.) When \( L \geq \bar{L} \), there exists \( K_1 \) such that:

- When \( K > K_1 \), the PBE in the steady-state is a pooling equilibrium without layoffs.
- When \( K \leq K_1 \), the PBE in the steady-state is a pooling equilibrium with layoffs.
- There is never a separating equilibrium in the steady-state.

(Thin markets.) When \( L < L \), there exists \( K_2 \geq K_3 \) such that:

- When \( K > K_2 \), the PBE in the steady-state is a pooling equilibrium without layoffs.
- When \( K \leq K_3 \), the PBE in the steady-state is a separating equilibrium.
- There is never a pooling equilibrium with layoffs in the steady-state.

(Markets with intermediate thickness.) When \( L \in (L, \bar{L}) \), there exists \( K_1 > K_2 > K_3 \) such that:

- When \( K > K_1 \), the PBE in the steady-state is a pooling equilibrium without layoffs.
- When \( K_1 \geq K > K_2 \), the PBE in the steady-state is a pooling equilibrium with layoffs.
- When \( K \leq K_3 \), the PBE in the steady-state is a separating equilibrium.

**Proof.** See Appendix A.7.

Below is a figure illustrating the heuristics behind Theorem 1.2.
Figure 1-4 reproduces Figure 1-3, and it demonstrates the trajectories of outside options (and consequently, the PBE) in steady-state as the search frictions decrease from $K = \infty$ to $K = 1$. When $K = \infty$, the outside options are reduced to zero because returning to search market takes forever. Therefore, the PBE is a pooling equilibrium without layoffs. As $K$ decreases, the outside options for both parties increase, but the relative speed at which they increase depend on $L$. By Lemma 1.2, the flow rate for a firm is independent of $L$ and the flow rate for a worker is decreasing in $L$. As a result, a higher value of $L$ leads to a flatter trajectory because the outside option of the worker increases slower. When $L$ is high enough, the trajectory is sufficiently flat so that the region where there is a separating equilibrium is never visited. When $L$ is high enough, the locus is sufficiently steep so that the region where there is a pooling equilibrium with layoffs is never visited. For intermediate values of $L$, all three regions are visited in the order described in Theorem 1.2.

The economics behind Theorem 1.2 is as follows. When $L$ is high ($L \geq \bar{L}$), competition among unemployed workers is strong and it takes a long amount for an unemployed worker to find a match. Consequently, market opportunities for the workers remain unattractive, even when search frictions disappear. Therefore, a worker who finds a match always joins the firm. As a result, there is never a separating equilibrium, regardless of the level of search frictions. The level of search frictions, however, determine whether the firm quits after learning about the match quality. When search frictions are high ($K > K_1$), the firm’s market

---

\(^{21}\)The discontinuities at the boundaries are due to the capital choice and market tightness changing discontinuously as the PBE changes.
opportunities are also unattractive, so the firm does not break up a relationship with low match quality. When search frictions are low \((K \leq K_1)\), the firm’s market opportunities improve. This leads to the firm breaking up relationships with low match quality, and hence to a pooling equilibrium with layoffs.

In contrast, when \(L\) is sufficiently low \((L \leq L)\),\(^{22}\) competition among unemployed workers is weak. When search frictions decrease, market opportunities for the worker rapidly improve to the point where a worker in a relationship with low match quality does not join the firm \((K \leq K_3)\). In this case, separation is achieved for the values of search frictions where the firm continues after learning that the match quality is low. As a result, a pooling equilibrium with layoffs never appears: as search frictions decrease, the economy directly moves from a pooling equilibrium with layoffs to a separating equilibrium.

When \(L\) has an intermediate value \((L \in (L, \bar{L}))\), competition among unemployed workers is still not too strong. Separation is ultimately possible: when market is frictionless, a worker in a relationship with low match quality does not join the firm. Separation, however, requires the market to be near frictionless, and for high values of search frictions there is still pooling. In this case, a decrease in search frictions first increases the firm’s market opportunities to the point where there is a pooling equilibrium with layoffs \((K \leq K_1)\). When search frictions are reduced to sufficiently low levels \((K \leq K_3)\), a separating equilibrium is attained. As a result, the economy ends up in each of the three pure strategy PBE for various levels of search frictions.

1.5.1 Implications on Productivity

Given Theorem 1.2, one can now derive comparative statics on the total production in a partnership as a function of the search frictions. The following result is a continuation of Proposition 1.3.

**Proposition 1.4.** Let the production in a match be given by the lifetime discounted output in an employment relationship, as defined in Proposition 1.3.

- If the market is **thick** \((L \geq \bar{L})\), the total production in matches formed in a period is **increasing** in search frictions.
- If the market has **intermediate** thickness \((L \in (L, \bar{L}))\), the total production in matches formed in a period is **first decreasing and then increasing** in search frictions.
- If the market is **thin** \((L \leq L)\), the total production in matches formed in a period is **decreasing** in search frictions.

**Proof.** Proposition 1.4 directly follows from Proposition 1.3 and Theorem 1.2. \(\Box\)

\(^{22}\)Theorem 1.2 is stated for \(L \geq 1\), but the characterization of the \(L \leq L\) case extend to \(L < 1\). In that case, Lemma 1.2 would not hold, and instead one would have \(q(r) = \frac{L_L}{(1-L)L_k} \) and \(Tq(r) = 1\). The rest of the analysis follows the same steps.
Proposition 1.4 shows that the effect of a reduction in search frictions depends on the aggregate market conditions. In thick markets, the layoffs effect dominates and productivity decreases as search frictions decrease. In thin markets, the signaling effect dominates and productivity increases. In more balanced markets, the effect is nonmonotonic. When search frictions decrease from high to intermediate levels, layoffs effect dominates and productivity decreases. When they decrease to sufficiently low levels, signaling effect dominates and productivity increases. Therefore, productivity is u-shaped in search frictions.

Proposition 1.4 is a descriptive result, and similar descriptive results apply for other potential outcomes of interest such as the turnover rate and the average tenure length in a labor market.

1.5.2 Implications on Worker Welfare

In addition to the descriptive characterization results, Theorem 1.2 has normative implications as well. This subsection considers the welfare of a newborn worker in the steady-state and analyzes how it is affected by search frictions.\footnote{One could also consider the welfare of a newborn firm, but I am focusing on the worker welfare for two reasons. First, the welfare of a firm with respect to the market conditions has less interesting features than that of the worker. This is essentially because firm is the second mover in the employment relationship: because it responds optimally to the selection and its own outside option, a simple revealed preference argument shows that the firm cannot be worse off as a result of an increase in its own outside option. The second reason behind focusing on worker welfare is more policy-relevant: increasing the welfare of individual workers is arguably more important from a normative perspective.}

Since a newborn worker has the status of an unemployed worker, her welfare is equal to $J^U$. Rearranging Equation (1.10):

$$J^U = J^U \cdot \frac{\frac{\tau q(\tau)}{1 - \delta(1 - s) + \delta(1 - s)\tau q(\tau)}}{\frac{1}{1 - \delta(1 - s) + \delta(1 - s)\tau q(\tau)}} \cdot \frac{\Sigma_w(v, \pi)}{\text{match payoff component}}$$

This has two components:

1. The flow rate component depends on the market tightness in the steady-state, which is a function of $L$ and the turnover rate in the PBE. By Lemma 1.2, this is decreasing in $L$ and $\rho^*$.  

2. The match payoff component depends on the PBE of the employment relationship. It is increasing in the worker's outside option and decreasing in the firm's outside option.

How do these components change as the search frictions change?

1. By Theorem 1.2, as search frictions decrease the flow rate increases. The only PBE without any turnover (i.e. with $\rho^* = 1$) is the pooling equilibrium without layoffs, and this equilibrium ceases to exist for sufficiently low values of $K$. This is the market tightness effect: decreasing search frictions always help newborn workers to find their
first matches easier by increasing turnover.

Below is an illustration of the market tightness effect for the case of \( L \in (L, \bar{L}) \) (the alternative cases are similar, and only the cutoffs change):

![Graph showing market tightness effect](image)

Figure 1-5: Market tightness effect.

2. The change in match payoff is more delicate and possibly nonmonotonic. This is the **match payoff effect**, and it depends on the market thickness.

- If the market is thick \( (L \geq \bar{L}) \), for \( K > K_1 \), the payoff from match is constant in \( K \). For \( K \leq K_1 \), it is decreasing in \( K \) and is below the payoff when \( K > K_1 \).

- If the market is thin \( (L \leq L_0) \), for \( K > K_2 \), the payoff from match is constant in \( K \). For \( K \leq K_2 \), it is decreasing in \( K \) and is above the payoff when \( K > K_2 \).

- If the market has intermediate thickness \( (L \in (L, \bar{L})) \), for \( K > K_1 \), the payoff from match is constant in \( K \). For \( K_1 \geq K > K_2 \), it is decreasing in \( K \) and is below the payoff when \( K > K_1 \). For \( K \leq K_2 \), it is decreasing in \( K \) and is above the payoff when \( K > K_1 \).

Below is an illustration of the match payoff effect for the case of \( L \in (L, \bar{L}) \):

![Graph showing match payoff effect](image)

Figure 1-6: Match payoff effect.
For the case of \( L \in (\underline{L}, \overline{L}) \), the net welfare effect of decreasing the search friction can be observed by juxtaposing Figures 1-5 and 1-6. Consider a market that initially has very high search frictions (\( K > K_1 \)), but the search frictions decrease due to a technological shock or institutional reasons. Reduction of search frictions to \( K \geq K > K_2 \) has opposing effects: matches arrive faster to workers, but each individual match yields lower payoff. In other words, market tightness effect and match payoff effect operate in opposite directions. The net welfare effect depends on which effect dominates. When search frictions decrease to sufficiently low levels (\( K \leq K_3 \)), the trade-off disappears: market tightness effect and match payoff effect operate in the same direction. Therefore, the worker welfare unambiguously increases.

The welfare effect of reducing search frictions to intermediate levels depends on the strength of match payoff effect. The crucial parameter here is the responsiveness of capital choice to labor input. Consider the case where the capital choice is sufficiently responsive to labor:

\[
\frac{k(\phi)}{\bar{k}(\phi)} > \frac{(L-1)(1-\delta(1-s)) + \phi + (1-\phi)\eta_0}{(L-1)(1-\delta(1-s))\phi + \phi + (1-\phi)\eta_0} \quad (1.13)
\]

Under the cost function considered in Equation (1.1), Equation (1.13) holds when \( \Gamma \) is sufficiently small.

**Proposition 1.5.** Assume Equation (1.13) is satisfied.

**Thick markets.** When \( L \geq \overline{L} \),

- For \( K > K_1 \), the welfare of a newborn worker is constant in \( K \).
- For \( K \leq K_1 \), the welfare of a newborn worker is decreasing in \( K \), and below the value of welfare when \( K > K_1 \).

**Intermediate markets.** When \( L \in (\underline{L}, \overline{L}) \):

- For \( K > K_1 \), the welfare of a newborn worker is constant in \( K \).
- For \( K_1 \geq K > K_2 \), the welfare of a newborn worker is decreasing in \( K \), and below the value of welfare when \( K > K_1 \).
- For \( K \leq K_3 \), the welfare of a newborn worker is decreasing in \( K \), and above the value of welfare when \( K > K_1 \).

**Thin markets.** When \( L \leq \underline{L} \),

- For \( K > K_2 \), the welfare of a newborn worker is constant in \( K \).
- For \( K \leq K_3 \), the welfare of a newborn worker is decreasing in \( K \), and above the value of welfare when \( K > K_2 \).

**Proof.** See Appendix A.7. \( \square \)
Proposition 1.5 demonstrates that the effect of search frictions on worker welfare depends on aggregate market conditions as well. When the market is thick, the competition in the labor market is so strong that signaling opportunities never occur. In this case, high search frictions yield higher worker welfare by ensuring that the relationships are long-lasting. When the market is thin, low search frictions yield higher worker welfare by generating signaling opportunities.

When the market has intermediate thickness, worker welfare is u-shaped with respect to the search frictions. A frictionless market yields high welfare by providing signaling opportunities and inducing high investments. High search frictions yield a lower welfare because they lock workers in with their first matches, but they induce moderate capital levels. Intermediate search frictions are the worst for the workers: they do not generate any signaling opportunities and they induce low capital choice.

This discussion offers a framework for considering policy implications as well. A prevalent feature of U.S. labor markets is the high mobility of young workers: a typical male worker holds seven jobs in his first ten years in the labor market (Topel and Ward, 1992). This “job shopping” constitutes a stark contrast with the apprenticeship system in German labor market, which restricts a young worker’s mobility but ensures training by firms (Harhoff and Kane, 1997; Acemoglu and Pischke, 1998). Arguments for and against the U.S. system vis-à-vis the German system reduce to a discussion on the relative value of having higher match quality versus firm-specific training (Heckman, 1993; Felli and Harris, 2018). This paper argues that when information about match quality is asymmetric, ensuring mobility of workers can induce firm-specific training by generating signaling opportunities for workers. In loose labor markets, the German system (where the workers are locked in to their first matches) is more preferable for the workers. In tight labor markets, the U.S. system (where the workers can involve in excessive job shopping) is more preferable. If the market is balanced, each system has its merits but a mixed system (which neither ensures locking in, nor has frictionless labor markets) is a dismal combination: it is the one which yields lower worker welfare than either of these extremes.

1.6 Conclusion

This paper shows that a frictionless search market may promote productive partnerships by generating signaling opportunities about the match quality. It builds on the idea that information about the match quality is asymmetric: one party of the relationship (in this case, workers) have more information about the match quality than the other party (firms). My model focuses on employment relationships, but the insights extend to a wider range of settings where there is asymmetric information about match quality and investment by the (initially) uninformed party: business-to-business relationships, export development and pairwise partnerships are some examples.

The main takeaway is that reducing search frictions in a market helps relationships by im-
proving the outside option of the informed party, and simultaneously hurts relationships by improving the outside option of the uninformed party. When the market has intermediate thickness, these effects dominate for different values of search frictions. An optimal market has an “all-or-nothing” structure: markets help only when they are frictionless, and markets with moderate frictions are worse than no markets at all.

Another important insight is that the welfare effects of a reduction in search frictions depend on market thickness. For reduced search frictions to improve worker welfare, a market must have sufficiently many vacancies. Therefore, from a normative perspective, ensuring that firms can enter a market is as important as reducing search frictions. The policies that increase firms’ participation in the search market (such as decreasing vacancy costs) or decrease competition among workers (such as increasing unemployment benefits) are likely to help among these aspects. An investigation of the effects of such policies in a general labor search framework with free entry on the firm side (as in Diamond (1982), Mortensen (1982), Pissarides (1990) and Mortensen and Pissarides (1994)) warrants further analysis.
Chapter 2

Busing with Contracts:
Message Spaces as a Design Tool in School Choice

2.1 Introduction

School choice reforms are motivated in part by the desire to break the link between where families live and the schools their children can attend. Under residence-based school assignment, families purchase access to schools by buying into a neighborhood. Credit-constrained families may not be able to afford housing in places where there are desirable schools. Under school choice, families can send their children to schools outside their neighborhood without needing to purchase housing there. As a result, proponents believe choice generates a more equitable distribution of schooling. The expansion of choice, however, entails costs associated with busing students across neighborhoods. For example, Boston Public Schools regularly spends about $110 million a year (10 percent of the total school budget) on transportation, placing it among the highest in the nation (Russell and Ebbert, 2011). School districts struggle to find the right balance between expanding access through choice and managing the associated transportation costs.

Dur: Department of Economics, NC State University, e-mail: udur@ncsu.edu. Gitmez: Department of Economics, Massachusetts Institute of Technology, e-mail: agitmez@mit.edu. Pathak: Department of Economics, Massachusetts Institute of Technology, and NBER, e-mail: ppathak@mit.edu. Sönmez: Department of Economics, Boston College, e-mail: ssonmez@bc.edu. We are grateful to Glenn Ellison, Drew Fudenberg, and participants at MIT Theory Lunch for helpful comments and suggestions. Pathak and Sönmez acknowledge support of National Science Foundation grant SES-1426566. Sönmez also acknowledges support from the Goldman Sachs Gives via Dalinc Ariburnu - Goldman Sachs Faculty Research Fund. Pathak is on the scientific advisory board of the Institute for Innovation in Public School Choice.

1In 2017, Boston Public Schools announced a transportation challenge to optimize the city's bus pick up times and school start and end times (BPS, 2017; Bertsimas et al., 2018).

2Pathak and Sönmez (2013) describe how Seattle Public Schools scaled back choice because of high busing costs. Landsmark (2009) argued that the Boston Public Schools should eliminate busing stating: "We can no longer afford to spend millions a year to bus children across Boston to schools that are not demonstrably better than schools near their homes."
In this paper, we offer a new lever in the assignment mechanism to help districts to expand access, while keeping transportation costs low. Existing school assignment mechanisms elicit preferences over schools, but not about transportation. In most systems, if a student is assigned to a non-neighborhood school, the district provides transportation either through school buses or by paying the costs associated with taking public transportation. We propose a mechanism with a larger message space: a student can apply to a school either with or without transportation. These two options correspond to different types of “contracts” with a given school. In our model, a non-neighborhood applicant who is willing to be assigned a school without transportation is given a greater chance of being assigned to the school. This reduces transportation costs without reducing access for non-neighborhood applicants. We illustrate the logic of the proposal both in a decentralized and centralized setting, and formally demonstrate it expands access while not increasing transportation costs.

While our analysis applies to any choice plan, our specific motivation comes from a recent policy discussion in Boston Public Schools. In 2012, the Mayor’s State of the City Address resuscitated the debate between neighborhood and choice proponents.³ In the speech, Mayor Menino questioned the merits of assigning children outside their neighborhood:

“Something stands in the way of taking our [public school] system to the next level: a student assignment process that ships our kids to schools across our city. Pick any street. A dozen children probably attend a dozen different schools. Parents might not know each other; children might not play together. They can’t carpool, or study for the same tests. […] Boston will have a radically different school assignment process – one that puts priority on children attending schools closer to their homes.”

As part of a community engagement process, an (anonymous) commentator stated:

“I suggest we add an incentive to encourage parents and students to return to neighborhood schools and improve the schools themselves. I call this a “market based” approach, and it would not impact the substance of any particular zone design […] We can encourage all of these goals – more students closer to home, more money to improve those schools and more money overall to improve all schools - and reduce the skepticism […] Quite simply, the idea is to create a credit for the transportation costs of each student who gives up a voluntarily busing slot and selects a walk zone school.”

Our paper is inspired by this suggestion. We develop a model of school choice with or without transportation which marries the usual school choice model (Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003) with the matching with contracts framework (Kelso and Crawford, 1982; Hatfield and Milgrom, 2005).⁴ Transportation in a district can involve busing or subsidized public transportation. Since busing is common, we refer to our model a school choice problem with busing contracts. We also use the idea of neighborhood and walk-zone, a catchment area around the school, interchangeably. In the model, a slot in a school may be a walk-zone slot, an open slot, or a non-busing slot. A walk-zone slot favors

³For more on this debate, see the materials available at http://bostonschoolchoice.org and press accounts by Goldstein (2012) and Handy (2012).
the students in the walk-zone of a school, while an open slot does not discriminate between applicants. A non-busing slot favors the students who forgo transportation services to attend the school. Walk-zone and open slots are familiar from choice mechanisms used in practice. The non-busing slot is a new, and is only relevant because of the richer message space.

After defining the model in Section 2.2, we begin our analysis with the case of only one school with a given set of contracts, which can be seen as the admissions policy at a given school in a decentralized system. Our first result (Proposition 2.3) shows that when we replace an open slot with a non-busing slot, it reduces the number of students using transportation services, without reducing access for non-neighborhood applicants. That is, it is possible to keep the school accessible to non-neighborhood applicants, without increasing transportation costs using this new lever. Turning to a centralized system, it is well-known that distributional comparative statics in matching models with an arbitrary number of schools cannot be guaranteed without strong assumptions on student preferences (see, e.g., Dur et al. (2018)). However, in a model with two schools, which allows for interaction across schools, our main result is that converting an open slot to a non-busing slot does not harm access, but does reduce transportation costs (Theorem 2.1). Finally, in Section 2.3.3, we examine the alternative of replacing a walk-zone slot with a non-busing slot, and show that this need not increase access and reduce transportation costs.

2.2 Model

2.2.1 Primitives

A school choice problem with busing contracts consists of:

- a finite set of students \( I \);
- a finite set of schools \( A \);
- a finite set of slots \( S_a \) for each school \( a \in A \);\(^5\)
- a set of terms \( T = \{B, NB\} \) where \( B \) stands for busing and \( NB \) stands for non-busing,
- a correspondence \( w : A \to I \), where \( w(a) \) denotes the set of students who are in the walk-zone of school \( a \in A \);
- a preference profile \( P = (P_i)_{i \in I} \) where \( P_i \) is the strict preference relation of student \( i \) over \( (A \times T) \cup \{\emptyset\} \), such that \( \emptyset \), the "null contract", represents remaining unassigned for student \( i \);
- a precedence profile \( \succ = (\succ^a)_{a \in A} \) where \( \succ^a \) is the order of precedence for school \( a \in A \) such that \( s \succ^a s' \) implies that slot \( s \in S_a \) will be filled before slot \( s' \in S_a \) whenever possible.

\(^5\)Throughout the study, we will assume that \( \sum_{a \in A} |S_a| \geq |I| \), i.e. there are enough slots to accommodate all students. Here, \( |S_a| \) is the capacity of school \( a \).
Schools use a uniform tie-breaker denoted by $\pi$ over the students. The tie-breaker is represented by a function $\pi : I \rightarrow \mathbb{R}$. Student $i$ is favored over student $j$ under tie-breaker $\pi$ if $\pi(i) < \pi(j)$.\footnote{We assume that $\pi(i) = \pi(j)$ implies that $i = j$.}

A contract $x = (i, a, t)$ is a 3-tuple consisting of a student $i \in I$, a school $a \in A$, and a term $t \in T$. Let $X$ denote the set of all contracts. For a contract $x \in X$, $i(x)$ refers to the student associated with $x$. That is, if $x = (i^*, a, t)$, then $i(x) = i^*$. Similarly, $a(x)$ and $t(x)$ denote the school and the term associated with contract $x$, respectively. Let $i(X') = \bigcup_{x \in X'} i(x)$.

For any set of contracts $X' \subseteq X$ and for any $t \in T$, let $X'(t) = \{x \in X' : t(x) = t\}$, i.e. the set of contracts in $X'$ associated with the term $t$. Similarly, for any $I' \subseteq I$, let $X'(I') = \{x \in X' : i(x) \in I'\}$, and for any $A' \subseteq A$, let $X'(A') = \{x \in X' : a(x) \in A'\}$. Let $X'(I', t) = X'(I') \cap X'(t)$.

Suppose $i(x) = i(x') = i$. Then, $i$ prefers contract $x$ to $x'$ if $(a(x), t(x)) P_i (a(x'), t(x'))$. With slight abuse of notation, we use $P_i$ to denote student $i$'s preferences over contracts. Student $i$ considers any contract $x$ with $i(x) \neq i$ as unacceptable, i.e., $\emptyset, P_i x$ for any $x \in X(I \setminus \{i\})$. Given a strict preference relation $P_i$, let $R_i$ denote the associated weak preference relation.

An allocation $\mu \subset X$ is a set of contracts such that each student is assigned to at most one contract and no school is assigned to more contracts than its capacity. Given an allocation $\mu$, if $x = (i, a, t) \in \mu$, we refer to contract $x$ as the assignment of student $i$ under allocation $\mu$ and we denote $\mu(i) = x$.\footnote{If there does not exist $x \in \mu$ such that $i(x) = i$, then we say student $i$ is assigned to $\emptyset$, and denote it by $\mu(i) = \emptyset$.} Similarly, given an allocation $\mu$, for any $a \in A$, we define the set $\mu(a) = \{x \in \mu : a(x) = a\}$ as the set of contracts assigned to school $a$.

A mechanism is an ordered pair $(Z, \phi)$ of a strategy space $Z = Z_1 \times Z_2 \times \ldots \times Z_{|I|}$ and an outcome function $\phi : Z \rightarrow X$. A mechanism $\phi$ selects an allocation $\phi(z) = (\phi_1(z_1), \ldots, \phi_{|I|}(z))$ for each strategy vector $z = (z_1, \ldots, z_{|I|}) \in Z$. Let $\phi_i(z)$ and $\phi_a(z)$ denote the assignment of student $i$ and school $a$ under $\phi(z)$, respectively.

Given a student $i \in I$ and strategy profile $z \in Z$, let $z_{-i}$ denote the strategy of all students except $i$. Let $P_i$ denote the set of preferences over $A \times T \cup \{\emptyset\}$, and let $\mathcal{P} \equiv P_1 \times \ldots \times P_{|I|}$. A direct mechanism is a mechanism where $Z = \mathcal{P}$. In this paper, we focus on the direct mechanisms. A direct mechanism $(\mathcal{P}, \phi)$ is strategy-proof if $\forall i \in I, P_{-i}, P_i, P'_i$ such that $\phi_i(P'_i, P_{-i}) \neq P_i \phi_i(P)$.

Compared to the model in Abdulkadiroğlu and Sönmez (2003), a student’s preference is over $(A \times T) \cup \{\emptyset\}$ rather than $A \cup \{\emptyset\}$. Since we allow students to express preferences over a richer set, our setup allows the mechanism to exploit information about applicants who have different preferences over transportation at a given school. More specifically, in current practice, a student in the walk-zone of school $a$ cannot use busing services if she is assigned
to school \( a \) and a student not in the walk-zone of school \( a \) is assumed to use busing at school \( a \). Therefore, the mechanism does not allow a non walk-zone student with low priority for school \( a \) who does not need busing to give up her right to busing in exchange for a higher chance of being assigned to \( a \).

We do not take a position on the optimal mix of students from the walk-zone and outside the walk zone. In practice, this fraction is determined by a compromise across different stakeholders. Our argument is that enlarging the message space allows for a better chance to find the balance between competing factions.

School \( a \) may wish to allow for a non walk-zone student to give up her right to busing, so as to reduce transportation costs. Indeed, one reason a school prioritizes a walk-zone student over one outside the walk zone is that the latter may use busing, while the former does not. In terms of transportation costs, school \( a \) is indifferent between a non-walk-zone student who is assigned under no busing and a walk-zone student who is assigned without busing since both do not use busing services. However, school \( a \) may prefer a non-walk-zone student over a walk-zone student due to concerns about diversity or equity.

In our model, by allowing non-walk-zone students to be assigned without busing, school \( a \) can admit more non-walk-zone students to address these diversity concerns without increasing its transportation costs. That is, by enlarging the message space, the mechanism relaxes the trade-off between diversity and transportation cost objectives in a way that current mechanisms do not allow.

### 2.2.2 Priorities and Choice Functions

There are three types of slots at a school: walk-zone, open, and non-busing slots. Walk-zone and open slots are defined such that they correspond to the type of slots used in practice. The third type of slot, the non-busing slot, is now possible thanks to the enriched message space of our setting.

For any school \( a \in A \) and for any slot \( s \in S_a \), the priority ordering for a slot depends on its type. As in the case with students, we define \( \emptyset_s \) as the "null contract" which represents remaining empty for any slot \( s \). Each slot has a strict priority order over contracts involving \( a \) as well as null contract \( \emptyset_s \). In other words, slot \( s \in S_a \) has strict priority order, denoted by \( s \), over \( (I \times \{a\} \times T) \cup \{\emptyset_s\} \). If contract \( x \) is ranked below \( \emptyset_s \) (i.e. \( \emptyset_s \succ_s x \)), then it is considered unacceptable for slot \( s \). We use the notion of unacceptable contracts to construct projections of the slot types in the current mechanism, which does not even allow for certain contracts, hence marking them as unacceptable.

The priority order of each slot \( s \in S_a \) of each school \( a \in A \) over the contracts including school \( a \) and \( \emptyset_s \) is determined as follows:

47
Walk-Zone Slots | Open Slots | Non-busing Slots
--- | --- | ---
walk, non-busing | walk, non-busing | walk, non-busing
$\sim_s$ | $\sim_s$ | $\sim_s$
non-walk, busing | non-walk, busing | non-walk, non-busing
$\succ_s$ | $\succ_s$ | $\succ_s$
$\emptyset_s$ | $\emptyset_s$ | $\emptyset_s$
$\sim_s$ | $\sim_s$ | $\sim_s$
walk, busing | walk, busing | walk, busing

Table 2.1: Slot Types

1. **s is a walk-zone slot**: Any non-busing contract offered by a walk-zone student of school $a$ has priority over any busing contract offered by a non-walk-zone student of school $a$. The order within each group is determined according to $\pi$. The busing contracts offered by walk-zone students and non-busing contracts offered by non-walk-zone students are considered unacceptable for this type of slot, hence they are ranked below $\emptyset_s$.

2. **s is an open slot**: Any non-busing contract offered by a walk-zone student of school $a$ is tied with any busing contract offered by a non-walk-zone student of school $a$. The order within this group is determined according to $\pi$. The busing contracts offered by walk-zone students and non-busing contracts offered by non-walk-zone students are ranked below $\emptyset_s$.

3. **s is a non-busing slot**: Any non-busing contract has priority over any busing contract offered by a non-walk-zone student. The priority order within these two groups is determined according to $\pi$. Only the busing contracts including walk-zone students are ranked below $\emptyset_s$.

The priority orders of these three types of slots of school $a$ can be represented in the following table:

It is now possible to define the choice function of a school using the priority order of each slot and the precedence order of slots. Consider a school $a \in A$, and let its slots $S_a = \{s_1, s_2, \ldots, s_{|S_a|}\}$ be ordered according to the precedence order $s_1 \succ^a s_2 \succ^a \ldots \succ^a s_{|S_a|}$. Given a set of contracts $X' \subseteq X$, the chosen set of school $a$ contracts from $X'$ is denoted by $C_a(X')$ (and, similarly, the rejected set of a from $X'$ is denoted by $R_a(X') = X' \setminus C_a(X')$). Given $X'$, $C_a(X')$ is calculated via following procedure:

**Step 0.** Remove all contracts that involve another school $a' \neq a$ and add them to $R_a(X')$. Set $X'_1 = X' \setminus R_a(X') = X'(a)$.

**Step 1.** Start with $s_1$. If there is no acceptable contract for $s_1$ in $X'_1$, then keep $s_1$ unassigned and set $X'_2 = X'_1$. Otherwise, assign the contract $x_1$ which is $\succ_{s_1}$-maximal among

---

*The priority ordering within unacceptable contracts can be arbitrary.

*Here, $\sim$, represents indifference and any indifference is broken according to $\pi$.

48
contracts in \( X'_1 \). Remove all other contracts associated with \( i(x_1) \), and add them to \( R_a(X') \) and set \( X'_2 = X'_1 \setminus (R_a(X') \cup \{x_1\}) \).\(^{10}\)

In general,

**Step \( k > 1 \).** Consider \( s_k \). If there is no acceptable contract for \( s_k \) in \( X' \), then keep \( s_k \) unassigned and set \( X'_{k+1} = X'_k \). Otherwise, assign the contract \( x_k \) which is \( \succ_{X'_k} \)-maximal among contracts in \( X'_k \). Remove all other contracts associated with \( i(x_k) \), and add them to \( R_a(X') \) and set \( X'_{k+1} = X'_k \setminus (R_a(X') \cup \{x_k\}) \).

Continue until processing all slots in \( S_a \). When the process ends, define \( C_a(X') = X' \setminus (X'_{|S_a|+1} \cup R_a(X')) \), and \( R_a(X') = X' \setminus C_a(X') \).

The choice functions of students are simpler to define, because students have unit demand in this model. In particular, given \( X' \subseteq X \), student \( i \in I \) picks the maximal contract in \( X' \), so that \( C_i(X') = \{ x \in X' \cup \{0\} : x \mathrel{P} x' \ \forall x' \in X' \setminus \{x\} \}. \)

### 2.2.3 Relationship with Matching with Contracts

The school choice functions satisfy properties that have been studied in the literature on matching with contracts. We first define substitutability and unilateral substitutability conditions, which are sufficient for the existence of a stable matching (Hatfield and Milgrom (2005) and Hatfield and Kojima (2010)).\(^{11}\)

**Definition 2.1.** A choice function \( C \) is **substitutable** if for all \( x, x' \in X \) and \( Y \subseteq X \), \( x \notin C(Y \cup \{x\}) \) implies \( x \notin C(Y \cup \{x, x'\}) \).

**Definition 2.2.** A choice function \( C \) is **unilaterally substitutable** if \( x \notin C(Y \cup \{x\}) \) implies \( x \notin C(Y \cup \{x, x'\}) \) for all \( x, x' \in X \) and \( Y \subseteq X \) such that \( i(x) \notin i(Y) \).

Note that substitutability implies unilateral substitutability. In the following example, we show that unilateral substitutability condition (therefore, substitutability) fails under our setup.\(^{12}\)

**Example 2.1.** Take a school \( a \in A \), and students \( \{i_1, i_2, i_3\} \subseteq I \). Let \( \{i_1, i_2, i_3\} \subseteq (I \setminus w(a)) \) and \( \pi(i_1) < \pi(i_2) < \pi(i_3) \). Suppose that \( a \) has two slots, \( S_a = \{s_1, s_2\} \), where \( s_1 > a s_2 \), \( s_1 \) is a non-busing slot and \( s_2 \) is an open slot. Now, let \( X' \subset X'' \subset X \) be:

\[
X' = \{(i_1, a, B), (i_2, a, B), (i_3, a, NB)\}
\]
\[
X'' = \{(i_1, a, NB), (i_2, a, B), (i_2, a, B), (i_3, a, NB)\}.
\]

\(^{10}\)With slight abuse of notation, we set \( i(\emptyset_{an}) = 0 \).

\(^{11}\)See Section 2.2.4 for the formal definition of stability.

\(^{12}\)Under the current practice, for a given student only the contracts with one type of contract terms is acceptable for schools. Therefore, the school choice function induced by the current priority orders is always substitutable.
In this case, \( C_a(X') = \{(i_1, a, B), (i_3, a, NB)\} \) and \( C_a(X'') = \{(i_1, a, NB), (i_2, a, B)\} \). Hence, the contract \((i_2, a, B)\), previously rejected even if it is the only contract in \( X' \) that includes student \( i_2 \), is accepted when a new contract is added. This is a violation of unilateral substitutability.

Next, we define bilateral substitutability (Hatfield and Kojima (2010)), a weaker condition than unilateral substitutability.

**Definition 2.3.** A choice function \( C_a \) is **bilaterally substitutable** if \( x \not\in C_a(Y \cup \{x\}) \) implies \( x \not\in C_a(Y \cup \{x, x'\}) \) for all \( x, x' \in X \) and \( Y \subseteq X \) with \( i(x), i(x') \not\in i(Y) \).

The following proposition demonstrates that in our model, a school's choice function satisfies bilateral substitutability and the irrelevance of rejected contracts condition of Aygün and Sönmez (2013).

**Proposition 2.1.** Every choice function \( C_a \) satisfies the bilateral substitutes and irrelevance of rejected contracts conditions.

**Proof.** The model is a special version of the matching problem with slot-specific priorities. Lemma 1 and Lemma A.1 of Kominers and Sönmez (2016) immediately give the desired result.\( \square \)

### 2.2.4 Stability and Strategy-Proofness

With the bilateral substitutes and irrelevance of rejected contracts conditions established in the previous subsection, we now restrict attention to the stable allocations. An allocation \( \mu \) is **stable** if:

1. \( \bigcup_{a \in A} C_a(\mu) = \mu \) and \( \bigcup_{i \in I} C_i(\mu) = \mu \); and
2. there exists no student \( i \), school \( a \) and contract \( x = (i, a, t) \in X \setminus \mu \) such that \( x = C_i(\mu \cup \{x\}) \) and \( x \in C_a(\mu \cup \{x\}) \).

A stable mechanism is one that always chooses a stable allocation. Proposition 2.1 and Theorem 1 of Hatfield and Kojima (2010) imply the existence of stable outcomes in our setting.

To compute a stable allocation, we next consider the generalization of the Gale and Shapley (1962) deferred acceptance algorithm, known as the **cumulative offer algorithm:**

**Step 1** Start the process with an arbitrary student. Call this student \( i_1 \). Let \( x_1 \) be \( i_1 \)'s most-preferred contract. If \( x_1 = \emptyset_{i_1} \), then we assign \( i_1 \) to \( \emptyset_{i_1} \) and set \( A_a(1) = \emptyset \) for all \( a \in A \). Otherwise, student \( i_1 \) offers her most-preferred contract \( x_1 = (i_1, a_1, t) \) to

\(^{13}\)Recall that a choice function \( C_a \) satisfies irrelevance of rejected contracts if for any \( X' \subset X \), \( x \in X \setminus X' \), \( x \not\in C_a(X' \cup \{x\}) \), then \( C_a(X') = C_a(X' \cup \{x\}) \).
school $a_1$. School $a_1$ holds the contract if $x_1 \in C_{a_1}(\{x_1\})$ and rejects it otherwise. Let $A_{a_1}(1) = \{x_1\}$ and $A_a(1) = \emptyset$ for all $a \in A \setminus \{a_1\}$.

In general,

**Step k** Take an arbitrary student for whom no contract is currently held by any school or being unassigned option, call this student $i_k$. Let $x_k$ be $i_k$'s most-preferred contract that has not been rejected in previous steps. If $x_k = \emptyset_{i_k}$, then we assign $i_k$ to $\emptyset_{i_k}$ and set $A_a(k) = A_a(k-1)$ for all $a \in A$. Otherwise, student $i_k$ offers contract $x_k = (i_k, a_k, t)$ to school $a_k$. School $a_k$ holds the contract if $x_k \in C_{a_k}(A_{a_k}(k-1) \cup \{x_k\})$ and rejects it otherwise.\(^{14}\) Let $A_{a_k}(k) = A_{a_k}(k-1) \cup \{x_k\}$ and $A_a(k) = A_a(k-1)$ for all $a \in A \setminus \{a_k\}$.

The algorithm terminates when every student has an offer that is on hold by a school or the option of remaining unassigned. As there are a finite number of contracts, the algorithm terminates in some finite number $K$ of steps. The final allocation is $\mu = \bigcup_{a \in A} C_a(A_a(K))$. If there does not exist a contract $x \in \mu$ such that $i(x) = i$ for some $i \in I$, then we set $\mu(i) = \emptyset_i$.

**Proposition 2.2.** Under school choice problem with busing contracts, the cumulative offer mechanism is stable and strategy-proof.

**Proof.** Kominers and Sönmez (2016) show that under matching problem with slot-specific priorities, the cumulative offer algorithm is strategy-proof and stable. Since our model is a special case of theirs, Theorem 3 of Kominers and Sönmez (2016) holds under our setting too.

It is worth noting that the cumulative offer algorithm allows for any order of proposals. This description is without loss of generality since (Kominers and Sönmez, 2016) show that the outcome of the cumulative offer algorithm is independent of the order of proposals.

### 2.3 Non-Busing Slots

The open and walk-zone slots described in Section 2.2.2 represent the slots used in practice in many cities. In Boston, there was a mix of walk-zone and open slots for each school, and this represented a long-standing compromise between different interest groups (see, e.g., (Dur et al., 2018) for more details). One particular objective is minimizing transportation costs, and we now show how non-busing slots can reduce transportation, while not harming access.

Suppose the transportation cost incurred by a school is zero for any non-busing contract, and some fixed amount $\kappa > 0$ for any busing contract. This formulation does not account for the variable cost arising from the differences in the distance among several neighborhoods, nor does it internalize the cost incurred by a student non-busing contract who needs to use public transportation (or private means of transportation) nevertheless. With this formulation, decreasing transportation costs is equivalent to decreasing the number of students.

\(^{14}\)Since the choice function satisfies irrelevance of rejected contracts, it suffices to consider only the unrejected contracts in each step.
with busing contracts assigned to the school. If all schools reach capacity, this objective is the same as increasing the number of students with non-busing contracts assigned by the assignment mechanism.

Our ultimate aim is demonstrating that non-busing slot serves the purpose of increasing non-busing contracts. Specifically, we show that a non-busing slot decreases transportation costs more than an open slot.

2.3.1 Decentralized Admissions: The One School Case

We first examine the one school case, which is akin to a partial equilibrium analysis. Analysis of this case will be a building block for the centralized case with more than one school, where there are additional effects due to interactions across schools. The one-school environment is also relevant for decentralized admissions scenarios, where a school is balancing walk, non-walk and transportation objectives.

For a school \( a^* \in A \) and a slot \( s^* \in S_{a^*} \), suppose that \( s^* \) is an open slot which is then converted into a non-busing slot. Suppose furthermore that the precedence of \( a^*, b^* \), the ordering induced by the tiebreaker \( \pi \), and types of other slots remain the same in both cases. Let \( C_{a^*} \) be the choice function induced when \( s^* \) is an open slot, and \( D_{a^*} \) be the choice function induced when \( s^* \) is a non-busing slot.

Our first result is that when an open slot is replaced with a non-busing slot, a school (weakly) reduces transportation costs, and has fewer busing contracts. It follows under a mild restriction on the set of available contracts to choose from: if a non-busing contract with a student is available, then a busing contract with the same student is also available. Such an restriction is consistent with the interpretation that busing service can be a freely disposed for students, so any student who is willing to be admitted without transportation should also be willing to be admitted with transportation.

**Proposition 2.3.** Take any set of contracts \( \bar{X} \subseteq X \) with the following property: if \((i^*, a^*, t) \in \bar{X} \) for some \( i^* \in I \), then \((i^*, a^*, B) \in \bar{X} \). If \( D_{a^*}(\bar{X}) \neq C_{a^*}(\bar{X}) \), then

i) the number of non-busing contracts in \( D_{a^*}(\bar{X}) \) is higher than that in \( C_{a^*}(\bar{X}) \),

ii) the number of busing contracts in \( C_{a^*}(\bar{X}) \) is higher than that in \( D_{a^*}(\bar{X}) \).

**Proof.** See Appendix B.1.

In the proof, we show that the increase in the number of non-busing contract does not necessarily come from replacing a non-walk-zone student with a walk-zone student. It may be the case that the number of chosen non-walk-zone students remains the same, whereas the number of non-busing contracts increases. It is also possible that the same set of students assigned under both cases, where a non-walk-zone student is admitted with a busing contract under \( C_{a^*} \) and with a non-busing contract under \( D_{a^*} \).
We next show that replacing an open slot with a non-busing slot does not reduce access for non-neighborhood applicants. This result holds under a regularity assumption on the precedence order of slots.

**Definition 2.4.** The precedence order $\triangleright_a$ of school $a \in A$ is **regular** if:

(i) all the walk-zone slots precede all the open slots, i.e. for any walk-zone slot $s \in S_a$ and any open slot $s' \in S_a$, $s \triangleright_a s'$, or,

(ii) all the open slots precede all the walk-zone slots, i.e. for any walk-zone slot $s \in S_a$ and any open slot $s' \in S_a$, $s \triangleright_a s'$.

A regular precedence order subsumes the Walk-Open or Walk-Open precedence orders, the latter being nearly identical to the actual precedence used by Boston Public Schools during 1999-2013 (Dur et al., 2018). It does not allow for a rotating precedence order.

**Proposition 2.4.** Assume that school $a^*$ has a regular precedence order $\triangleright_a^*$. Take any set of contracts $\bar{X} \subseteq X$ with the following property: if $(i^*, a^*, t) \in \bar{X}$ for some $i^* \in I$, then $(i^*, a^*, B) \in \bar{X}$. If, for a student $i \in I \setminus w(a^*)$, $(i, a^*, NB) \in \bar{X}$, then the following holds:

$$i \in i(C_{a^*}(\bar{X})) \implies i \in i(D_{a^*}(\bar{X}))$$

**Proof.** See Appendix B.2.

Proposition 2.4 illustrates the sense in which our proposal does not restrict access to non-neighborhood applicants. Consider a non-neighborhood student is admitted to school $a^*$ when $s^*$ is an open slot. If such a student is willing to be admitted to $a^*$ under a non-busing contract, she will still be admitted when $s^*$ is converted to a non-busing slot. It is possible that, under our proposal, this student is admitted under a non-busing contract rather than a busing contract; but our proposal ensures that the school does not reject the student altogether if the student is willing to forgo transportation services. The regularity of precedence order is a necessary condition: in Appendix B.3, we provide an example where the precedence order of school $a^*$ is not regular and the conclusions of Proposition 2.4 does not hold.

### 2.3.2 Centralized Admissions: Multiple Schools

Proposition 2.3 implies that when choosing between open slots and non-busing slots, a designer with the objective of transportation-cost minimization should choose the latter in a single school environment. However, in an environment with more than two schools, replacing an open slot with a non-busing slot may result in a lower number of non-busing contracts and a higher number of busing contracts. The following example illustrates this.

**Example 2.2.** $A = \{a, b, c\}$ with $S_a = \{s_a\}$, $S_b = \{s_b\}$ and $S_c = \{s_c\}$. $I = \{i_1, i_2, i_3\}$ where $w(a) = \{i_1, i_3\}$, $w(b) = \emptyset$, and $w(c) = \{i_2\}$. The tie-breaker is such that $\pi(i_1) < \pi(i_2) < \pi(i_3)$. The preference profile is:
First, consider the case where all slots are open slots. The outcome of cumulative offer algorithm is:

$$\mu = \{(i_1, b, B), (i_2, c, NB), (i_3, a, NB)\}.$$ 

Now, consider the case where $s_b$ is replaced by a non-busing slot. The outcome of cumulative offer algorithm is:

$$\nu = \{(i_1, c, B), (i_2, a, B), (i_3, b, NB)\}.$$ 

In Example 2.2, when an open slot in school $b$ is replaced by a non-busing slot, the total number of busing contracts increases. Nevertheless, the number of busing contracts associated with school $b$ does not increase. In Appendix B.4, we demonstrate that this may not hold in general either: replacing an open slot with a non-busing slot in a school may result in a higher number of busing contracts assigned to that particular school.

The essential reason behind the phenomenon illustrated in Example 2.2 is the existence of long rejections chains. For our next result, we restrict ourselves to the case with two schools and assume $A = \{a, b\}$. The intuition behind why such a restriction restores the desired result is simple: the limited number of walk-zones puts an upper bound on the length of rejection chains. Consequently, the designer can now attain the improvements that one would expect as a result of such a policy change. The justification behind the case with two schools follows from the policy objective of providing students from poorer neighborhoods with the opportunity to be placed into schools at richer neighborhoods. In particular, one may consider the setup with two schools as a stylized representation of a city with substantial polarization of income.

Our main theoretical result follows under an assumption on students’ preferences, which is as follows:

**Assumption 2.1.** For each $a \in A$ and $i \in I$, $(i, a, B) P_i (i, a, NB)$.

In words, this assumption is “students prefer busing over non-busing.” It is plausible assumption, especially when one takes into account that even though a student is assigned to a busing contract by the central mechanism, she is merely eligible to use this right, and using this option is not mandatory. Assumption 2.1 is the counterpart of the restriction on the set of available contracts made in Proposition 2.3.

The following is our main theoretical result.
**Theorem 2.1.** Suppose there are two schools \((A = \{a, b\})\), each student belongs to at most one walk zone,\(^{15}\) and Assumption 2.1 holds. Keeping the precedence order of each school fixed, replacing an open slot of a school \(a \in A\) with a non-busing slot weakly increases the total number of non-busing contracts and weakly decreases the total number of busing contracts assigned by the cumulative offer mechanism.

*Proof.* See Appendix B.5.

Our next result shows that the counterpart of Proposition 2.4 holds in the centralized setup. Replacing an open slot with a non-busing slot would not reduce access for non-neighborhood applicants. That is, suppose a non-neighborhood applicant is admitted to school \(a\) before. If she is willing to attend school \(a\) without a busing contract, she will receive an outcome which is at least as preferable as the non-busing contract with school \(a\).

**Proposition 2.5.** Suppose there are two schools \((A = \{a, b\})\), each school has a regular precedence order, each student belongs to at most one walk zone, and Assumption 2.1 holds. Keeping the precedence order of each school fixed, consider replacing an open slot \(s^* \in S_a\) of a school \(a \in A\) with a non-busing slot. Let \(\mu\) be the outcome of cumulative offer algorithm when \(s^*\) is an open slot, and let \(\nu\) be the outcome when \(s^*\) is a non-busing slot.

*Proof.* See Appendix B.6.

Note that it is possible to have \(\mu(i) = (i, a, B)\) and \(\nu(i) = (i, a, NB)\) under our proposal. Nevertheless, student \(i\) is guaranteed to have an assignment that is at least as preferable to \((i, a, NB)\) under this proposal as long as she finds \((i, a, NB)\) acceptable. In other words, the only students who lose access to school \(a\) are those who find a non-busing contract at \(a\) unacceptable. By a revealed preference argument, one may argue that such students are less likely to gain from being admitted to school \(a\) anyway, compared to other students who find such contracts acceptable. Under our proposal, the students who find such contracts acceptable are more likely to be admitted at the expense of those who find such contracts unacceptable. Such an outcome generates opportunities for potential welfare gains in a centralized setting.

### 2.3.3 Extensions

A natural question to ask is whether Theorem 2.1 extends to the case of replacing a walk-zone slot with a non-busing slot in a school. The answer is negative, and the intuition is simple: the non-walk-zone student who now has higher priority for the new slot may replace a walk-zone student. The replaced student then proceeds to her next proposal, which may

\(^{15}\)It takes a simple modification of the proof to show that the result extends to the case where students may be in the walk-zone of both schools.
include a busing contract with another school. The following example illustrates this.

**Example 2.3.** \( A = \{a, b\} \) with \( S_a = \{s_a\} \) and \( S_b = \{s_b\} \). \( I = \{i_1, i_2\} \) where \( w(a) = \{i_1\} \) and \( w(b) = \{i_2\} \). The tie-breaker is such that \( \pi(i_1) < \pi(i_2) \). The preference profile is:

<table>
<thead>
<tr>
<th></th>
<th>( P_{i_1} )</th>
<th>( P_{i_2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b, B)</td>
<td>(b, B)</td>
<td></td>
</tr>
<tr>
<td>(b, NB)</td>
<td>(b, NB)</td>
<td></td>
</tr>
<tr>
<td>(a, B)</td>
<td>(a, B)</td>
<td></td>
</tr>
<tr>
<td>(a, NB)</td>
<td>(a, NB)</td>
<td></td>
</tr>
<tr>
<td>( \emptyset_{i_1} )</td>
<td>( \emptyset_{i_2} )</td>
<td></td>
</tr>
</tbody>
</table>

First, consider the case where \( s_a \) is an open slot and \( s_b \) is a walk-zone slot. The outcome of cumulative offer algorithm is:

\[
\mu = \{(i_1, a, NB), (i_2, b, NB)\}
\]

Now, consider the case where \( s_b \) is replaced by a non-busing slot. The outcome in this case is:

\[
\nu = \{(i_1, b, NB), (i_2, a, B)\}
\]

In the case of replacing a walk-zone slot with a non-busing slot, a result along the lines of Theorem 2.1 holds if one is willing to impose more restrictions on the priority order of non-busing slot. In particular, for the number of non-busing contracts to increase, the non-busing slot needs to discriminate among non-busing contracts by giving priority to the non-busing contracts offered by walk-zone students over non-busing contracts offered by non-walk-zone students. We define such a slot to be a **non-commuting slot**. Such a priority order can be defended on the grounds that the non-walk-zone students would need to use private means of transportation even though they do not use school buses. Therefore, a non-commuting slot aims to decrease the transportation cost in general, rather than considering the transportation cost incurred by the central authority only.

Formally, for a school \( a \in A \), if \( s \in S_a \) is a non-commuting slot: any non-busing contract offered by a walk-zone student has priority over any non-busing contract offered by a non-walk-zone student, which, in turn, has priority over a busing contract offered by a non-walk-zone student. The priority order within each group groups is determined with \( \pi \). The busing contracts offered by walk-zone students are ranked below \( \emptyset_s \).

**Proposition 2.6.** Suppose there are two schools, each student belongs to at most one walk zone, and Assumption 2.1 holds. Keeping the precedence order of each school fixed, replacing a walk-zone slot of a school \( a \in A \) with a non-commuting slot weakly increases the total number of non-busing contracts assigned by the cumulative offer mechanism.

**Proof.** The proof of this Proposition is almost identical to that of Theorem 2.1. It requires minor modifications in the proof of Proposition 2.3, which we explain in Footnotes 1 and 2. \( \square \)
2.4 Conclusion

One of the main purposes of this paper is to illustrate the potential gains from enlarging the message space in the school choice setting. The machinery of matching with contracts framework allows the market designer to utilize such gains. The matching with contracts framework, therefore, can be leveraged to complete the missing markets even in a centralized school choice setting without all of the price instruments.

Our main theoretical contribution is the introduction of a new type of slot, the non-busing slot, which is possible thanks to the richer message space. Theorem 2.1 demonstrates that introducing non-busing slots in a school choice problem reduces transportation costs without explicitly favoring the neighborhood students over non-neighborhood applicants.

Enlarging the message space in a school choice problem with busing contracts opens up potential avenues for further theoretical investigation. One possibility is considering non-busing slots that favor certain non-busing contracts over others (i.e. letting non-busing contracts offered by a non-walk-zone student have priority over non-busing contracts offered by a walk-zone student, or having priority orders within non-busing contracts be broken according to certain observable characteristics of students). Thanks to such slots, it is possible to ensure even greater access for non-neighborhood applicants while keeping busing costs down. An extensive theoretical investigation on such slots is left for future work.
Chapter 3

Media Capture: A Bayesian Persuasion Approach

3.1 Introduction

Modern democracies are increasingly distressed about media capture, a phenomenon where the politician has control over the editorial policies over a media source. However, economics literature has provided surprisingly few theoretical insights about media capture yet. The canonical model of the media capture, offered by Besley and Prat (2006), adopts a straightforward view about the way in which media operates. It assumes that the politician practices tyranny over the captured media sources and prevents them from sending any negative news about herself. According to tyranny view, if the media is captured, the citizens observe nothing but the positive news about the politician. Therefore, the argument goes, media capture prevents the citizens from accessing the information required to keep the politicians accountable, hindering the essence of democracy.

This paper proposes an alternative view about media capture. It is based on the idea that politician’s decisions of whether to capture media and, conditional on capturing, how to send news are strategic decisions. This is because the politician captures media in order to persuade the citizens to support her, which is a strategic process (Kamenica and Gentzkow, 2011; Rayo and Segal, 2010). According to this view, which we will denote the persuasion view, it is not always the positive news which will be produced when the media is captured. Instead, the politician would sometimes act strategically and send negative signals about herself in order to be more persuasive. Consequently, media can still be informative even when it is captured. The informativeness of captured media depends on how easy it is for the politician to persuade the citizens, which in turn depends on the perceptions of the citizens about the politician. This naturally bears the question of how a change in citizens’
perceptions (an increase in the popularity of politician or a polarization of opinions among the citizens) affects the occurrence of positive and negative news.

In order to model the persuasion view, we present a model of Bayesian persuasion as in Kamenica and Gentzkow (2011). The politician wants the citizens to take a certain action, which corresponds to supporting a policy proposed by the politician. The benefit of the policy is unknown to the citizens, but the politician can send messages about the benefit through the captured media. Unlike most models of Bayesian persuasion in the literature, we consider citizens who have heterogeneous priors about the benefit of the policy, because they have differing views about the politician. We characterize the equilibrium of this model and identify the conditions for media to always send truthful news (i.e. to be fully informative), or the conditions for the media to send the same type of news regardless of the truth (i.e. to be fully uninformative). We then investigate how the equilibrium depends on the distribution of prior beliefs. Interestingly, the politician is more likely to send negative news about herself when there are more citizens with extreme opinions. The occurrence of negative news is an outcome of strategic decisions: the politician realizes that the abundance of negative news makes the rare positive news more convincing for the extreme opposers.

At the heart of our model is a prominent trade-off faced by the politician. It is the result of a simple observation: even when the politician has full control over the media source, she cannot fully control the citizens, who are free in their decisions to pay attention to the media source. A citizen wants to pay attention to the media source, however, in order to be informed. If the media sends nothing but positive news, it will not be informative at all and the citizens would stop paying attention to the news. In order to convince citizens to listen to the news, then, the politician needs to ensure that the media is sufficiently informative: the media should commit to sending truthful news at least occasionally. On the other hand, the politician wants to send favorable news about herself whenever possible in order to persuade citizens. Our Bayesian persuasion model succinctly captures this trade-off. The politician ends up solving an optimization problem where she balances the objectives of “convincing more people to listen to the media” (i.e. gaining in the extensive margin) and “sending a favorable message to those who listen to the media” (i.e. gaining in the intensive margin).

After setting up the politician's problem, we begin the analysis by identifying the conditions for the captured media to commit to sending truthful news or sending the same type of news regardless of the state (Proposition 3.1). We show that if there are increasingly more supporters (represented by an increasing density of prior distributions), then the media always sends the same type of news, and no information is revealed. This is because the politician enjoys enough support and she wants to ride the priors of citizens, which can be achieved by providing them as little information as possible and having them not listen to the news at all. On the contrary, if there are increasingly more opposers (represented by a decreasing density), captured media will always send truthful news: there will be full information revelation. The intuition for this result is that, if there are more opposers in the society, the politician needs to persuade the citizens rather than confusing them. It turns out the most efficient way to persuade the citizens is to be truthful towards the citizens and having everyone listen to the media. Such a full revelation result typically does not appear in the
standard Bayesian persuasion results with one receiver: it occurs purely because there are multiple receivers with heterogeneous priors.

In our next set of results (Propositions 3.2 and 3.3), we provide characterizations of optimal media policies for single-peaked and single-dipped distributions of priors beliefs. A single-peaked distribution of prior beliefs corresponds to a society with many moderate citizens, whereas a single-dipped distribution suggests the existence of more extremists than moderates. An implication of the characterization results is that, in a society where citizens tend to hold extreme beliefs rather than moderate ones, one is more likely to encounter negative news rather than positive ones. The intuition for this result is that, when there are more people with extreme prior beliefs in the society, the politician faces a tough problem: she needs to convince the extreme opposers to listen to the news, which requires committing to a certain level of truthful revelation. But such a strategy has a risk of alienating the extreme supporters: they may listen to the media and stop supporting the politician if the news are negative. It turns out that the optimal strategy for the politician in this case is to allow a lot of negative news to be produced, so that the occurrence of positive news is an event rare enough to convince the extreme opposers to listen to the media. On the other hand, the negative news are abundant enough so that they do not sway the opinions of extreme supporters. This is the first analysis in the literature on Bayesian persuasion under a heterogeneous audience (Kolotilin et al., 2017) about the effect of changes in the distribution of receiver types on the optimal solution.1

Our results shed light on interesting and not entirely obvious relationships between media capture and the distribution of beliefs in society. A recent trend in the distribution of beliefs in modern democracies, which is a topic of increasing interest in the literature, is the increase in the number of people with extreme opinions (Abramowitz and Saunders, 2008; Bishop, 2009; Gentzkow, 2016). It is widely argued documented that this phenomenon and negative news are jointly observed (Bernhardt et al., 2008; Bail et al., 2018). Our finding suggests a surprising channel of causality in this relationship: it may be the politician herself, the target of negative news, who prefers the occurrence of negative news. Correspondingly, this finding suggests that the occurrence of media sources which send negative news about the politician is not necessarily indicative of media freedom.

Our next set of results are a group of comparative statics regarding the informative content of news. Proposition 3.4 demonstrates that it is relatively more popular politicians (indicated by the distribution of prior beliefs with more mass towards the right) who suppresses the informative content of the news. The intuition is related to the first set of results we discuss above: an unpopular politician wants to make sure that citizens listen to the media, so she would would allow for information revelation. On the other hand, a sufficiently popular politician suppresses the bad news, to the extent that it would make the media source uninformative, to ensure that citizens act based on their priors. This is a prediction that is not offered by the tyranny view of media capture. It is also suggestive of why some

1Kolotilin (2015) contains a comparative statics analysis on the distribution of types in a Bayesian persuasion model with heterogeneous audience, but its results are about the monotonicity of welfare, not the properties of solution.
leaders have been involved in capturing and suppressing the media only after they gained enough popularity. Proposition 3.5 identifies conditions on when polarization of opinions in a society (in the sense of replacing some moderate citizens with extreme opponents or supporters) leads to less informative content of media. It demonstrates that polarization leads to less informative news only if polarization generates more opposers than supporters. The intuition is as follows: when most of the moderate citizens turn into extreme opponents, the politician has fewer marginal citizens to convince to listen to the media. On the other hand, there is roughly the same number of supporters who already listen to the media. To avoid antagonizing the supporters, therefore, the politician chooses a less informative media source, and she is more lenient on losing the marginal citizens. If polarization generates more supporters than opposers, this result is reversed: there are already many extreme supporters, so the politician can afford to cater to the marginal citizens by increasing the informative content of the news.

Our final result (Proposition 3.6) demonstrates that if the media becomes less accessible to the citizens (due to cognitive or monetary costs associated with paying attention to the news), it becomes more informative. Intuitively, this is because a higher cost of listening to media implies that citizens with extreme opinions never follow the news. This leads to the politician being less concerned about losing the support of the extreme supporters, and she has incentives to make the media more informative to convince the moderate citizens to listen to the media. Interestingly, this finding suggests that accessibility of the media and its informativeness may be negatively associated. This finding can be coupled with the observation that the increased access to newsprint in 19th century led to a less informative press (Kaplan, 2002), or the social media as a news source suffers from the overwhelming existence of fake news (Allcott and Gentzkow, 2017). It is a prediction that models of media capture without any strategic considerations does not make, and, as it is the case with the other testable predictions of this model, opens up potential avenues for empirical research.

3.1.1 Relation to the Literature

Our model is a Bayesian persuasion model with heterogeneous priors. The two building blocks of the model (Bayesian persuasion, and heterogeneous priors) has been discussed in different literatures extensively, but ours is among the first ones which put these two pieces together. Prominent examples of the Bayesian persuasion literature rely on the common prior assumption, whereas the literature on heterogeneous priors has not been involved with the analysis of persuasion.

This paper is most closely connected to the Bayesian persuasion literature, where the sender is trying to persuade the receiver to take the action preferred by the sender (Rayo and Segal, 2010; Kamenica and Gentzkow, 2011). Such baseline models consider just one sender and one receiver, whereas we consider an audience with multiple heterogeneous receivers. In Kolotilin 2

---

2This result carries the same flavor as in Proposition 5 of Shadmehr and Bernhardt (2015), which is a model of information suppression rather than information design. In Shadmehr and Bernhardt (2015), the authors point out to the example of the Shah relaxing censorship during the unfolding of the 1979 Iranian Revolution when he lost popularity (Milani, 2011).

62
et al. (2017), the audience is heterogeneous because each receiver has private information about her (payoff-relevant) type. Whereas our model can be reformulated such that the prior beliefs can be interpreted as private types, our analysis is concerned with the properties of optimal solution and comparative statics with respect to the distribution of types. In contrast, Kolotilin et al. (2017)’s main contribution is demonstrating the equivalence of public and private persuasion mechanisms. The idea between such an equivalence also provides a foundation for our result on sufficiency of only two messages in our model (Lemma 3.1). Wang (2015), Alonso and Câmara (2016a), Bardhi and Guo (2018) and Chan et al. (2019) consider models of Bayesian persuasion with multiple receivers in a voting environment with common priors. Wang (2015) and Chan et al. (2019) compare the performance of public persuasion with various forms of private persuasion. Alonso and Câmara (2016a) focuses on public persuasion under different voting rules, and Bardhi and Guo (2018) considers private persuasion under unanimity rule. There is a growing literature on Bayesian persuasion with multiple receivers in non-voting environments with multiple audiences and various solution concepts (Mathevet et al., 2019; Taneva, 2019; Inostroza and Pavan, 2018).

The first model of Bayesian persuasion with heterogeneous priors is Alonso and Câmara (2016b), which is a model with one receiver. Laclau and Renou (2017) show that the techniques of Alonso and Câmara (2016b) extend to the case of multiple receivers with heterogeneous priors. The main focus of these papers is to provide conditions to guarantee that the sender benefits from persuasion, whereas we are concerned with the properties of optimal policy.

Our model is also reminiscent of the Gehlbach and Sonin (2014) model on government control of the media, where the authors assume that the politician has commitment power as well. The main difference between our setup and theirs is: we assume heterogeneous priors on the receiver side, so that a given signal distribution may have different effects in different people’s beliefs (and consequently, actions). This is not a very stringent approximation of reality, especially when one considers that typically people have diverse opinions about the quality of the politician or the potential outcomes of the actions taken by the politician. A sort of heterogeneity on the receiver side (which implies that only a certain subset of the population watches the news) is modeled through heterogeneous costs of listening to media in (Gehlbach and Sonin, 2014). Having heterogeneous priors instead of heterogeneous costs allows us to identify who watches the news and who does not, which is useful for our purposes. Differently from Gehlbach and Sonin (2014), we are interested in: how the distribution of priors (i.e. the initial popularity of politician) influences the informativeness of media, and how the informativeness of media influences who watches the media. None of these questions can be answered without having a model with heterogeneous priors. The canonical model of media capture, Besley and Prat (2006), also investigates a politician’s incentives to capture the

---

3Kolotilin (2015) is another model which allows private persuasion. Similar to our methodology, Kolotilin (2015) also investigates the potential effects of changes in the distribution of receiver types; but, it is exclusively focused on the effects on measures such as the ex ante welfare of the sender and receiver, rather than inspecting their effects on optimal disclosure strategies. Kolotilin (2018) also considers a model where the receiver has a private type. Differently from our model, the types and state of the world are jointly distributed. In other words, the type of receiver depends on the state of the world in a stochastic manner. Consequently, the sender would find it beneficial to condition the signal distribution on the state in order to elicit the private type.

3.2 The Model

In this section, we present the problem faced by a politician who has total control over a media source’s editorial policy and who wants to persuade the receivers to support her by sending positive news. We demonstrate that the persuasion setup can be reduced to a simple optimization problem where the politician trades off “convincing more people to listen to the media” and “sending a favorable message to those who listen to the media”.

3.2.1 Notation and Assumptions

We model the problem faced by a politician as a Bayesian persuasion problem towards an audience with heterogeneous priors. The politician wants the citizens to take an action which is favorable towards herself. This corresponds to supporting a policy proposed by the politician. Obtaining support requires persuading the citizens that the state is “good”, i.e. the policy is beneficial for the citizens. As standard in the Bayesian persuasion literature, we assume that the sender can commit to any distribution of messages conditional on the state. Unlike the canonical models of Bayesian persuasion, we assume that there are multiple receivers with heterogeneous priors. That is, the receivers disagree about the ex ante likelihood of the good state; hence, even though they observe the same message, their interpretations differ.

The model has two type of agents: the politician (denoted by \( p \)), and a continuum of citizens, denoted by \( I \). We will denote a generic citizen as \( i \in I \). The total measure of citizens is normalized to one.

There is an underlying state of the world, denoted by \( \theta \in \{0, 1\} \). Here, \( \theta = 1 \) corresponds to the “good” state where the policy is beneficial for the citizens. Conversely, \( \theta = 0 \) is the “bad” state where the policy is not beneficial. For the purposes of this model, what matters is that if the state is \( \theta = 0 \), the citizens will keep the status quo, and if the state is \( \theta = 1 \), they approve the policy. In other words, the politician and the citizens agree on state \( \theta = 1 \) and disagree on state \( \theta = 0 \).

To express this notion more formally, assume that the citizen \( i \in I \) takes an action \( a_i \in [0, 1] \). Here, \( a_i \) is the “support” that citizen \( i \) provides to the policy. Her utility is:
\[ u_i(a_i, \theta) = a_i(\theta - c) \]

where \( c \in (0, 1) \) is the cost of implementing the policy relative to the status quo, and is common knowledge. As a result, parameter \( c \) also corresponds to the posterior belief that a citizen needs to have to support the policy: it is a measure of how powerful the status quo is.

The politician cares about maximizing the support, so her utility is:

\[ u_p(\{a_i\}_{i \in I}) = \int_{i \in I} a_i di \]

This would be easy to analyze if the state is observed by all agents, but the state is not observed by the citizens. Only the politician observes the state, so the citizens need to act based on their beliefs about \( \theta \). The politician is able to send informative messages through the media. Assume that the politician can send a message \( s \in \{g, b\} \) in each state, where \( s = g \) corresponds to a "good" message which suggests that \( \theta = 1 \), and \( s = b \) is the "bad" message which suggests that \( \theta = 0 \).\(^4\) We assume that the politician can commit to generating any distribution of signals conditional on each state, before the state is realized. This implies that the politician’s strategy can be expressed as a pair:

\[ \sigma = (\sigma_0, \sigma_1) \in [0, 1]^2 \]

where \( \sigma_0 =: Pr(s = g|\theta = 0) \) is the probability of sending good message in the bad state, and \( \sigma_1 =: Pr(s = g|\theta = 1) \) is the probability of sending a good message in the good state.

The politician’s strategy is perfectly observable by each citizen.

Observing the politician’s strategy/editorial policy, each citizen \( i \in I \) decides whether to watch the news or not. Watching the news costs \( \kappa > 0 \) to each citizen. For the baseline model, we will assume that \( \kappa \) is an infinitesimally positive amount, which is the cognitive cost of watching the news. The implicit assumption here is that there is no monetary cost of following the news (consider the media source as public television), and processing the news does not impose a significant cost on the citizens. In Section 3.6, we provide a generalization to the case where listening to media costs any positive amount \( \kappa > 0 \).

Watching the news is helpful for the citizens to the extent that it potentially changes the action the citizen will take. This, clearly, depends on the distribution of posteriors of agent \( i \), which in turn depends on her prior. The citizens have heterogenous priors about the state of the world. That is, the ex ante probability of good state, \( Pr\{\theta = 1\} \), differs by citizen and is indexed by \( i \). For each \( i \in I \), let \( p_i := Pr\{\theta = 1\} \in [0, 1] \) denote the prior (subjective) probability that citizen \( i \) assigns to the good state. Assume that the priors are

\(^4\)As in every communication model, for each equilibrium, there is a corresponding “mirror-image” one which we don’t consider.
drawn independently from a distribution $F(.)$ with support $[0, 1]$, that is:

$$p_i \sim F(.) \ \forall i \in I$$

Along with the assumption that the total measure of citizens is normalized to one, this means that the measure of citizens with prior less than or equal to $p$ is $F(p)$. In order to avoid measurability issues, we assume that the pdf of $F(.)$, $f(.)$, is continuously differentiable. We will focus on single-peaked or single-dipped distributions throughout the analysis. Such a focus is appropriate for the types of questions we are asking (such as the effects of polarization), and is relevant for the analysis of political economy in many cases.

The politician's prior is:

$$p^* = 1 - c$$

which is inversely related to the cost of policy: if a policy is more costly (or if the status quo is stronger), the politician is less optimistic about its benefit.\(^5\)

**Timing** The timing of the game is as follows.

1. The prior of each citizen is drawn, and each citizen $i \in I$ observes her prior. The distribution $F(.)$ is common knowledge.
2. The politician commits to a strategy $(a_0, a_1)$, which is observed by each citizen.
3. Each citizen $i \in I$ decides whether to watch news or not.
4. The state is realized, and the media sends the message drawn according to the politician's strategy.
5. Each citizen $i \in I$ updates her prior, and chooses the action $a_i$ based on the posterior.
6. Payoffs are realized.

The solution concept we will adopt is Perfect Bayesian Equilibrium.

### 3.2.2 Discussion of The Model

The baseline model contains four crucial features of the model that are worthy of some discussion on their own. The first one is the ability of the politician to commit to any conditional distribution of messages. The second one is the assumption that the politician can generate only two types of messages (good or bad). Finally, the question of whether more general forms of payoff functions for the citizens and for the politician can be adopted remains. We discuss the justifications behind our modeling choices and the implications of relaxing them.

\(^5\)One may microfound this formulation by assuming that the politician considers the benefit from the policy to be uniformly distributed between 0 and 1.
Commitment The justification for the commitment assumption comes from how media capture works in practice. Even when the politician captures the media, she cannot check every single news produced by the state television. Nevertheless, she can have a general control over the media source’s editorial policy and the intensity of control is observable by the citizens. Of course, the intensity of control depends on the state of the world (i.e. the information that the state television has access to), hence the signal distribution depends on $\theta$. The commitment to a strategy $(\sigma_0, \sigma_1)$, then, should be interpreted as commitment to an editorial policy which is observable to the citizens.

The technical advantage of this assumption is that it allows us to reduce the decision of politician to a standard constrained optimization problem (Section 3.2.3). In this sense, it allows us to use the machinery used in the Bayesian persuasion literature (Kamenica and Gentzkow, 2011) which makes the analogous assumption.

Two Messages The assumption of having as many messages as the number of possible recommended actions is adopted by Kamenica and Gentzkow (2011). The authors justify their assumption by deriving a result which closely resembles revelation principle (Proposition 1). Their result depends on having only one receiver, and it breaks down in the case of multiple receivers with heterogeneous priors. If more messages are available for the politician, she would be able to induce some action profiles she is not able to induce with two messages only. The following result argues that, in the set of distributions we consider, having two messages does not impose a consequential restriction.

Lemma 3.1. Let $S$ be a set of signals with $|S| \geq 2$, and assume the politician commit to any signal distribution $(\sigma_0, \sigma_1) \in \Delta(S) \times \Delta(S)$ If $f(.)$ is single-peaked or single-dipped, the optimal strategy can be implemented using only two signals.

Proof. See Appendix C.1. The proof is an adaptation of Theorem 1 in Kolotilin et al. (2017), which proves an equivalence result in a Bayesian persuasion model with heterogeneous audience but common priors.

The assumption of having only two messages can also be justified on the grounds that the citizens are subject to some cognitive constraints. One may imagine that it is impossible for a citizen to distill every detail in the news and form a nuanced view of the state of the world. Instead, each citizen reads watches the news and leaves with some takeaway in the form of “the state is good” or “the state is bad”.

Payoff Functions of Citizens The current form of payoff functions of citizens, which adopts linear costs, tends to lead to corner solutions. That is, we will generically observe $a_i \in \{0, 1\}$.

---

6One alternative modeling strategy is to model this as a cheap talk game, which does not allow for commitment on the politician’s side. This requires that the signal distributions in each state must be the same ($\sigma_1 = \sigma_0$), thus we would be forgoing at least some of the richness of the model.
in the equilibrium. This is a deliberate modeling choice. Here is what goes wrong if an alternative payoff function is adopted. Suppose that we change the utility function of citizen $i$ to:

$$u_i(a_i, \theta) = -(a_i - \theta)^2$$

The quadratic loss function implies that the best action of the agent is no longer the “corner” action; rather, it equals to the posterior belief given the observed message. This has two implications.

1. Because the agents will necessarily act based on the posterior, everyone will watch the news even if they are a little bit informative. Consequently, the main trade-off of politician will be lost.

2. Because each citizen is Bayesian, her beliefs are a martingale with respect to her own prior. Therefore, when her action equals her posterior belief, the expected action for each citizen equals to her prior. But then, a politician with linear utility will be indifferent among any posterior distribution, so the politician will not care about the informativeness of media at all. We conclude that some level of “coarseness” of the action is necessary for the persuasion setup to be nontrivial. As long as such coarseness is maintained, the insights generated in specific case with only two actions apply.

Payoff Function of Politician The baseline model assumes linear utility on the politician’s side. That is, the politician cares about the sum of all $a_i$’s, rather than a more sophisticated functional form. This is assumption is crucial for the model to be operational. Suppose that we change the utility function of the politician to:

$$u_p(\{a_i\}_{i \in I}) = -\int_{i \in I} (1 - a_i)^2 di$$

In this case, the quadratic loss function would make the politician “risk-averse” with respect to each citizen’s action. Consequently, the politician would enjoy as little variation as possible in the action of each citizen from an ex ante perspective. She would therefore prefer to make the media as uninformative as possible, in order to keep the posteriors close to prior. In equilibrium, one would have an uninformative media which nobody follows.

3.2.3 The Optimization Problem

We now describe how the politician’s choice of editorial policy can be expressed as a constrained optimization problem, which simplifies the analysis considerably. The simplification works in two steps: we first identify which citizens would listen to the media given an editorial policy, and then present the problem of finding the optimal policy of the politician who takes the citizens’ decisions into account. The optimization problem we obtain succinctly

\[\text{This would change with } \kappa > > 0.\]

\[\text{This would change if the politician cares about the revenue obtained by media, as in Gehlbach and Sonin (2014).}\]

\[\text{This would probably change with coarser actions, but adopting a convex loss function instead of a linear one would not provide any insights not given by our baseline model.}\]
captures the main trade-off of media capture: the politician needs to make the media informative enough so that it convinces the citizens to listen to the news, but still distort the information such that good messages are sent frequently.

The Decision of Listening Given \((\sigma_0, \sigma_1)\): We begin by pinning down a strategy of the politician \((\sigma_0, \sigma_1)\) first; we will characterize program which picks the optimal pair \((\sigma_0, \sigma_1)\) later.

Consider a citizen \(i \in I\) and her posterior belief about the probability of good state following a message \(s\), \(\Pr_i(\theta = 1|s)\). The utility function of the citizen implies that her action is:\(^{10}\)

\[
\begin{align*}
a_i &= \begin{cases} 
1 & \text{if } \Pr_i(\theta = 1|s) \geq c \\
0 & \text{otherwise.}
\end{cases}
\end{align*}
\]

We now consider how the posteriors are calculated. If a citizen \(i\) does not watch the news, she will have her posterior equal to \(p_i\) following any message. If citizen \(i\) watches the news, her posterior will depend on the message \(s\) she observes. In this case:

\[
\begin{align*}
\Pr_i\{\theta = 1|s = g\} &= \frac{p_i\sigma_1}{p_i\sigma_1 + (1-p_i)\sigma_0} \\
\Pr_i\{\theta = 1|s = b\} &= \frac{p_i(1-\sigma_1)}{p_i(1-\sigma_1) + (1-p_i)(1-\sigma_0)}
\end{align*}
\]

Note that, given \((\sigma_0, \sigma_1)\), both values are increasing in \(p_i\). Moreover, assuming \(\sigma_1 \geq \sigma_0\),\(^{11}\) we have:

\[
\Pr_i\{\theta = 1|s = g\} \geq p_i \geq \Pr_i\{\theta = 1|s = b\}
\]

That is, good news update the prior upwards, and bad news update it downwards, as expected.

Now we characterize when citizen \(i\) decides to watch the news. Given the infinitesimal cost \(\kappa > 0\) of watching the news, a citizen \(i\) watches the news if and only if it would change the citizen's action with some probability. This is possible if and only if \(i\) takes different actions following each message. That is, \(i\) watches news if and only if:

\[
\Pr_i\{\theta = 1|s = g\} \geq c \geq \Pr_i\{\theta = 1|s = b\}
\]

\(^{10}\)The agent is indifferent between any actions when her posterior equals \(c\). Consequently, there is an indeterminacy in this case. We stay away from this indeterminacy by simply that the citizen takes politician's favorite action when indifferent, as in Kamenica and Gentzkow (2011). This assumption can be justified on the grounds that the politician can always invoke a posterior \(c + \epsilon\) for any \(\epsilon > 0\), thus having the agent strictly prefer \(a_i = 1\).

\(^{11}\)Without loss of generality, since in the opposite case, the politician can generate the same distribution of posteriors by adopting an alternative strategy where \(\sigma_1 \geq \sigma_0\).
Substituting the above expressions, this condition becomes:

\[
\frac{p_i \sigma_1}{p_i \sigma_1 + (1 - p_i) \sigma_0} \geq c \geq \frac{p_i (1 - \sigma_1)}{p_i (1 - \sigma_1) + (1 - p_i) (1 - \sigma_0)}
\]

Rearranging, we end up with the following Lemma.

**Lemma 3.2.** Given \((\sigma_0, \sigma_1)\) where \(\sigma_1 \geq \sigma_0\), a citizen \(i \in I\) with prior \(p_i\) watches the news if and only if \(p_i \in [\underline{p}, \bar{p}]\), where:

\[
\underline{p} = \frac{c \sigma_0}{c \sigma_0 + (1 - c) \sigma_1} \quad \bar{p} = \frac{c(1 - \sigma_0)}{c(1 - \sigma_0) + (1 - c)(1 - \sigma_1)}
\]

Those with \(p_i < \underline{p}\) do not watch the news because they will set \(a_i = 0\) even if they would observe the good news: they simply don’t believe in the politician because their initial beliefs are too pessimistic. Similarly, those with \(p_i > \bar{p}\) do not watch the news because they will set \(a_i = 1\) even if they would receive bad news: they support the politician regardless of the information obtained from the news.\(^2\)

Given this discussion, the measure of citizens who watch news is:

\[
F(\bar{p}) - F(\underline{p}) = F\left(\frac{c(1 - \sigma_0)}{c(1 - \sigma_0) + (1 - c)(1 - \sigma_1)}\right) - F\left(\frac{c \sigma_0}{c \sigma_0 + (1 - c) \sigma_1}\right)
\]

These citizens act based on the information obtained from the news: they set \(a_i = 1\) if and only if the message is good. In addition, a measure \(1 - F(\bar{p}) = 1 - F\left(\frac{c(1 - \sigma_0)}{c(1 - \sigma_0) + (1 - c)(1 - \sigma_1)}\right)\) of citizens always set \(a_i = 1\) without watching the news.

**Setting** \((\sigma_0, \sigma_1)\): Now we consider the politician who picks a strategy to commit to. If the politician’s prior about the state of the world is \(p^* = 1 - c \in [0, 1]\), by committing to \((\sigma_0, \sigma_1)\), the politician sends a good message with probability \(p^* \sigma_1 + (1 - p^*) \sigma_0\). All in all, the expected support for policy is:

\[
(F(\bar{p}) - F(\underline{p})) \left(p^* \sigma_1 + (1 - p^*) \sigma_0\right) + (1 - F(\bar{p}))
\]

\(^2\)As a sanity check, note that when \(\sigma_1 = \sigma_0\) (i.e. when media is totally uninformative), we have: \(p = \bar{p} = c\) – no one (except possibly for a measure zero of agents) watches the news. Conversely, when \(\sigma_1 = 1\) and \(\sigma_0 = 0\) (i.e. when media is fully informative), we have: \(p = 0\) and \(\bar{p} = 1\) – everyone watches the news.
A politician with prior $p^*$ therefore solves the following optimization problem when choosing the editorial policy $(\sigma_0, \sigma_1)$:

$$\max_{(\sigma_0, \sigma_1) \in [0,1]^2, \sigma_1 \geq \sigma_0} \left( F \left( \frac{c(1 - \sigma_0)}{c(1 - \sigma_0) + (1 - c)(1 - \sigma_1)} \right) - F \left( \frac{c\sigma_0}{c\sigma_0 + (1 - c)(1 - \sigma_1)} \right) \right) (p^* \sigma_1 + (1 - p^*)\sigma_0)$$

$$+ \left( 1 - F \left( \frac{c(1 - \sigma_0)}{c(1 - \sigma_0) + (1 - c)(1 - \sigma_1)} \right) \right)$$

The solution clearly depends on the value of $p^* = 1 - c$ and $F(.)$. Our approach would be fixing $c$ and observing how the solution changes with $F(.)$. On the practical level, this corresponds to investigating how the informativeness of media changes when people's opinions about a politician change.

The broad insight about the politician's problem is that: in the optimal solution, she finds the balance in an intensive versus extensive margin trade-off. In the absence of any strategic considerations, the politician would send only positive news, i.e. she would set $\sigma_0 = \sigma_1 = 1$. When citizens are strategic in their decisions, however, such a strategy would result in no one listening to the media, i.e. it would lead to $p = \bar{p} = c$. In this case, the politician would like the media to be more informative, which would be reflected in a decrease in $\sigma_0$. By decreasing $\sigma_0$, the politician would ensure more people listening to the media (in the form of a decrease in $p$), which correspond to gaining people in the extensive margin. On the other hand, the occurrence of good news would decrease, and each citizen who listens would be supporting the policy with a lower likelihood, i.e. the politician would lose people in the intensive margin. The optimal solution balances these two effects. The figure below illustrates this idea. For illustration purposes, it fixes $\sigma_1 = 1$ (which leads to $\bar{p} = 1$) and considers the optimal choice of $\sigma_0 \in [0,1]$, which maps into the optimal choice of $p$.

![Figure 3-1: The intensive versus extensive margin trade-off.](image)
3.3 The Solution

In this section, we sketch the universal features of the solution to the politician's problem. In particular, we demonstrate that for the class of prior distributions we consider, the solution lies in the extreme: it either reveals one of the states perfectly (i.e. one of the messages is fully informative) or it contains no information at all. To set up this result, we first provide a geometric interpretation of the politicians constrained optimization problem and offer an interpretation for the boundary conditions. We then demonstrate that if the prior distribution is single-peaked or single-dipped, the solution must lie in one of these boundaries: we cannot have an interior solution.

3.3.1 Simplifying the Problem

Let us begin by providing a geometric interpretation of the politicians problem. The constrained optimization problem can be represented as a problem on a two-dimensional plane whose constraints form a triangle. Each edge of the triangle represents an extreme information structure, and these extremes will correspond to potential solutions of the problem.

Substituting in \( p^* = 1 - c \) into the objective function of politician and rearranging, one can write the politician's problem as:

\[
\min_{(\sigma_0, \sigma_1) \in \Delta} \Pi(F, (\sigma_0, \sigma_1)) \tag{3.1}
\]

where

\[
\Pi(F, (\sigma_0, \sigma_1)) := F \left( \frac{c(1 - \sigma_0)}{c(1 - \sigma_0) + (1 - c)(1 - \sigma_1)} \right) (c(1 - \sigma_0) + (1 - c)(1 - \sigma_1)) + F \left( \frac{c\sigma_0}{c\sigma_0 + (1 - c)\sigma_1} \right) (c\sigma_0 + (1 - c)\sigma_1)
\]

and

\[\Delta := \{(x, y) : x \geq 0, 1 \geq y \geq 0, y \geq x\}\]

Note that \( \Delta \), the set of feasible \((\sigma_0, \sigma_1)\) pairs, is geometrically simply a triangle. The interior of this set corresponds to \((\sigma_0, \sigma_1)\) pairs such that \(0 < \sigma_0 < \sigma_1 < 1\). Any such pair maps to a range of priors where people listen to media \([\theta, \tilde{\theta}]\) with \(0 < \theta < c < \tilde{\theta} < 1\). In addition to the interior region, there are three boundaries of this set:

- The first boundary is the line where \(\sigma_0 = \sigma_1\). In this boundary, any message sent by the media is **fully uninformative**, i.e. posterior following any message equals the
prior for each citizen who listens to the media. In this boundary, we have $\bar{p} = \bar{p} = c$; consequently, the measure of citizens who listen to the media equals zero.

- The second boundary is the line where $\sigma_0 = 0$. In this boundary, the "good" message is fully informative, in the sense that $\Pr_i(\theta = 1|s = g) = 1$ for each $i \in I$ who listens to the media. In this boundary, we have $\bar{p} = 0$, i.e. even the most pessimistic citizen listens to the media.

- The third boundary is the line where $\sigma_1 = 1$. In this boundary, the "bad" message is fully informative, in the sense that $\Pr_i(\theta = 1|s = b) = 0$ for each $i \in I$ who listens to the media. In this boundary, we have $\bar{p} = 1$, i.e. even the most optimistic citizen listens to the media.

The figure provided below illustrates $\Delta$ and summarizes this discussion.

![Figure 3-2: The set of feasible ($\sigma_0, \sigma_1$) pairs.](image)

3.3.2 Discarding Interior Solutions

Having Equation 3.1 in hand, we now proceed to analyze the properties of the solution. The following result demonstrates that if the distribution of priors is single-peaked or single-dipped, the solution cannot be interior and we need to look for the boundaries of $\Delta$ to find the solution.

Lemma 3.3. Suppose $f(.)$ is single-peaked or single-dipped. Then, the politician's problem cannot have an interior solution. That is, the optimal solution has one of the following three forms.

- $\sigma_0 = \sigma_1$, where the messages are fully uninformative,
- $\sigma_0 = 0$, where the "good" message is fully informative,
- $\sigma_1 = 1$, where the "bad" message is fully informative.
Lemma 3.3 is extremely useful because it reduces the optimization problem into a one-dimensional problem: all one needs to do is to check the boundaries of \( \Delta \), discussed above. In Appendix C.4, we discuss how this problem can be reduced to a tractable one-dimensional optimization problem through appropriate changes of variables.

### 3.4 Characterization Results

With the simplified optimization problem in hand, it is now the time to characterize the solutions. In this section, we identify the general features of the optimal policy given the prior distribution. Proposition 3.1 characterizes the solution for the cases of monotonically increasing or monotonically decreasing densities: in the former case the optimal policy is to reveal nothing, and in the latter case the optimal policy is to reveal the state perfectly. Propositions 3.2 and 3.3 are weaker characterizations for the single-peaked and single-dipped distributions, respectively. In the case of single-peaked distributions, the “bad” message is fully informative; in the single-dipped case, the “good” message is.

**Proposition 3.1.** If \( f(.) \) is monotonically increasing in \([0, 1]\), then the optimal strategy has \( \sigma_0^* = \sigma_1^* \), i.e. the media is fully uninformative. If \( f(.) \) is monotonically decreasing in \([0, 1]\), then the optimal strategy has \( \sigma_0^* = 0 \) and \( \sigma_1^* = 1 \), i.e. the media is fully informative.

**Proof.** See Appendix C.5. \( \square \)

Proposition 3.1 offers sharp sufficient conditions for the shape of the solution. It suggests that if beliefs are skewed in favor of the politician, the politician makes the media fully uninformative, so that nobody follows the media. If, on the other hand, beliefs are skewed against the politician, the politician makes the media fully informative, so that everybody follows the media and they always learn the truth.

In order to see the intuition behind this results, one can revisit to the Kamenica and Gentzkow (2011)’s Bayesian persuasion setup. The following is a very simple but substantial insight offered by Kamenica and Gentzkow (2011): the extent of persuasion depends on the ex ante disagreement between the sender and the receiver. If the receiver agrees with sender ex ante (i.e. if she takes the sender’s preferred action in the absence of any information), no persuasion will be observed in equilibrium: indeed, the sender would like to reveal as little information as possible and “ride the prior” of the receiver. On the other hand, if the receiver and sender disagree ex ante, there will be some persuasion, and hence some information revelation, in equilibrium. As the receiver becomes more pessimistic about the state ex ante, the sender will choose a more informative revelation strategy. Looking at Proposition 3.1 through this lens makes the intuition clear. When the citizens are ex ante more in favor of the politician, there is no persuasion and information revelation – instead, what we will observe is people acting based on their priors. When the citizens are ex ante less in favor of the politician, the politician makes her best to convince people and reveal as
much information as possible.

The following is another result which has a similar flavor as Proposition 3.1. The conditions it employs are much weaker than the monotonicity conditions required in Proposition 3.1, hence the predictions are also weaker. Nevertheless, a tight characterization is still possible, and the general predictions are in line with those made in Proposition 3.1.

**Proposition 3.2.** Suppose \( f(.) \) is single-peaked.

- If \( F(c) > c \), then there will be some information revelation. That is, the messages cannot be fully uninformative. The optimal solution has \( \sigma_1^* = 1 \) (i.e. the “bad” message is fully informative) and \( \sigma_0^* \in [0, 1) \).

- If \( F(c) < c \), then there will never be full information revelation. The optimal solution has \( \sigma_1^* = 1 \) (i.e. the “bad” message is fully informative) and \( \sigma_0^* \in \left[ \frac{c-F(c)}{c}, 1 \right] \).

**Proof.** See Appendix C.6. \( \square \)

Note that the condition \( F(c) > c \) can equivalently be expressed as “less than \( c \) fraction of citizens support the policy ex ante”. Therefore, this should be read as a condition on the level of ex ante support the policy enjoys.

Note that any equilibrium of a single-peaked distribution is that has \( \sigma_1^* = 1 \). That is, as long as \( \sigma_0^* \in [0, 1) \), one has \( \bar{p} = 1 \) in equilibrium: even the most optimistic citizens will follow the news. The intuition behind this result is that: with a single-peaked distribution, there are more moderate citizens with priors close to \( c \). These citizens are the ones who are more inclined to listen to the news, and they are more sensitive to the news. As a result, the intensive margin becomes the more important one, and the setup becomes a generalization of Kamenica and Gentzkow (2011). The politician can set \( \sigma_1 = 1 \) and choose a sufficiently high \( \sigma_0 \in [0, 1] \) to ensure that positive news are generated sufficiently frequently (i.e. the gain the intensive margin is enough) and the losses in the extensive margin equals the gains in the intensive margin (i.e \( p \) is sufficiently high). The illustration in Figure 3-1, it turns out, represents the optimal choice of the politician. In the optimal policy, the gains from having more people listen to the media (the mass of citizens in the red slice) equals the expected loss of people who are informed about the media (the likelihood of citizens in the blue region supporting the policy).

Even though \( \sigma_1^* = 1 \) is pinned down in equilibrium, pinning down the exact value of \( \sigma_0^* \) is more difficult. Nevertheless, it is possible to identify some sufficient conditions. For instance, in the first case where the policy does not enjoy ex ante support, the following condition is sufficient to ensure that \( \sigma_0^* = 0 \):

\[
F(x) \geq x \quad \forall x \in [0, c]
\]
This is equivalent to $F(.)$ being first-order stochastically dominated by the uniform distribution in $[0, c]$. Intuitively, this means that $F(.)$ has “sufficient mass on the left”, i.e. there are enough people with unfavorable ex ante beliefs. In general, though, there are examples where $a_0 > 0$. An obvious one is the toy model discussed in the introduction of Kamenica and Gentzkow (2011), which is indeed a special case of the model considered here.

In the second case where the policy enjoys some ex ante support, it is possible to set a lower bound for $a_1^*$. As one can see in the statement of Proposition 3.2, $c - F(c) / e$ is such a lower bound. (This lower bound is not tight and in many cases, one may have $a_0^* = a_1^* = 1$.) Realize that as $F(c) \to 0$, this lower bound converges to one. That is, as the mass of citizens who ex ante oppose the policy shrinks, the media is likely to get less informative – in the limit, when $F(c) = 0$, no information is revealed and everybody supports the politician.

Below is a “mirror image” of Proposition 3.2, which investigates the case of single-dipped distributions.

**Proposition 3.3.** Suppose $f(.)$ is single-dipped.

- If $F(c) > c$ (i.e. if the policy does not enjoy ex ante support), then there will be **some** information revelation. That is, the messages cannot be fully uninformative. The optimal solution has $a_0^* = 0$ (i.e. the “good” message is fully informative) and $a_1^* \in \left[\frac{F(c) - c}{F(c) - F(c)c}, 1\right]$.

- If $F(c) < c$ (i.e. if the policy enjoys ex ante support), then there will never be full information revelation. The optimal solution has $a_0^* = 0$ (i.e. the “good” message is fully informative) and $a_1^* \in [0, 1)$.

**Proof.** See Appendix C.7.

Most of this result has an obvious parallel to Proposition 3.2, so all the comments we made earlier on Proposition 3.2 can be adapted to here. For instance, the first part of Proposition 3.3 provides a lower bound for $a_1^*$, and this lower bound converges to 1 as $F(c) \to 1$. That is, as the mass of citizens who ex ante oppose the policy grows, the media is more likely to get more informative.

Any equilibrium of a single-dipped distribution has $a_0^* = 0$ and, as long as $a_1^* \in (0, 1]$, one has $p = 0$ in equilibrium. Therefore, even the most pessimistic citizens will follow the news. In contrast to the case with a single-peaked distribution, there are not many moderate citizens; instead, now there is an overwhelming measure of extremist citizens with priors close to 0 or 1. These citizens are, in general, not inclined to listen to the news unless they are extremely informative in the opposite direction as their priors, e.g. an extremist opposer (with prior $p_i \approx 0$) would listen to the news only if the good news are informative. To convince these citizens to listen to the news, the politician offers informative good news with $a_0^* = 0$. She chooses a sufficiently high $a_1^* \in [0, 1]$ to ensure that positive news are generated sufficiently frequently, yet $\tilde{p}$ is sufficiently low so that extremist supporters still act based on their priors.
The figure below illustrates the optimal policy in this case. The citizens whose priors are larger than $\bar{p}$ act based on their priors (i.e. they set $a_i = 1$) and the citizens whose priors are smaller than $\bar{p}$ follow the news (i.e. they set $a_i = 1$ only if $s = g$).

![Figure 3-3: Optimal policy in a single-dipped distribution.](image)

3.5 Comparative Statics

Having characterized the optimal solution for different distributions of priors, we now conduct comparative statics exercises with respect to the distribution of the priors of the citizens.

3.5.1 Effect of Popularity

The first comparative statics exercise we will conduct is regarding the popularity of the politician. We model an increase in a politician's popularity as a shift in the citizen's priors over the right in the MLRP sense. The following result characterizes the changes in the informativeness of the media.

**Proposition 3.4.** Consider two distributions $f(.)$ and $\tilde{f}(.)$ satisfying the strict monotone likelihood ratio property, i.e., such that the ratio $f(x)/\tilde{f}(x)$ is strictly increasing in $x$.

(a) If $f(.)$ and $\tilde{f}(.)$ are both single-peaked, then $\sigma_0^* > \tilde{\sigma}_0^*$.

(b) If $f(.)$ and $\tilde{f}(.)$ are both single-dipped, then $\sigma_1^* < \tilde{\sigma}_1^*$.

**Proof.** See Appendix C.8. $\square$

In both cases covered in Proposition 3.4, the media becomes less informative as the citizens become more optimistic about the politician. The intuition follows from the one we provided after Proposition 3.1: a less popular politician needs to convince the citizens to listen to the news, which requires providing an informative media. On the other hand, a more popular politician can afford to provide a less informative media and ride the priors of citizens.
Our interpretation for Proposition 3.4 is that only sufficiently popular politicians are bothered by an independent media which provides informative news. Indeed, if there is a positive cost of capturing the independent media, only sufficiently popular citizens would be willing to pay this cost. This reasoning suggests that media capture is a concern only when the politician is popular. This is a prediction that the canonical view of media capture, the tyranny view, does not provide. In our model, media capture and popularity of politicians appear as complements rather than substitutes. This dynamic sheds light to an interesting dynamic observed in modern democracies, where the concerns about media capture became more prevalent as the politicians became more popular over time in their respective countries and enjoyed more vocal support. The empirical support behind this prediction warrants further analysis.

3.5.2 Effect of Polarization

One qualitative difference between the optimal policies in the single-peaked and single-dipped distributions is the abundance of negative news. A group of citizens which primarily involve extremists (as represented by a single-dipped prior distribution) encounters negative news more frequently than a group of citizens which primarily involve moderates (as represented by a single-peaked prior distribution). Interestingly, when the group of citizens involve extremists, it is the politician herself who sends the negative news. This stands in contrast with the common narrative about the effects of polarization.

Inspired by this observation, we study the effect of continuously deforming a single-peaked distribution of the priors to more closely resemble a uniform distribution. Practically, this corresponds to removing some moderates from the group of citizens and adding some extremists. We do this by considering a flexible family of distributions parameterized by two parameters:

\[ F_{\alpha, \rho}(x) = (\alpha S(x)^\rho + (1 - \alpha)x^\rho)^\frac{1}{\rho} \quad \alpha \in [0, 1], \rho \in \mathbb{R}. \]

Note that \( F_{\alpha, \rho}(0) = 0, F_{\alpha, \rho}(1) = 1, \) and \( F'_{\alpha, \rho}(x) > 0 \) for all \( \alpha \in [0, 1], \) so \( F_{\alpha, \rho}(0) = 0 \) is a proper cdf.

Given \((\alpha, \rho), F_{\alpha, \rho}(x)\) is a mixture of \(S(x)\) and (some transformation of) the uniform distribution. Because \(S(x)\) is single-peaked, this corresponds to putting more weight on the tails of \(S(x)\). Indeed, given \(\rho\), decreasing \(\alpha\) corresponds to taking some mass from the middle of \(S(x)\) and putting it on the extremes. Below is a figure illustrating the case where \(\rho = 1\) and \(\alpha \in (0, 1)\).
\[ p \in \mathbb{R} \] is a tractable way of parametrizing the weight one puts on either extreme. The easiest way of seeing this is to consider the limits of \( p \).

- When \( p = \infty \), \( F_{\alpha, \rho}(x) = \max\{S(x), x\} \). Therefore, \( F_{\alpha, \rho}(x) \) corresponds to taking some mass from the middle of \( S(x) \) and adding it to the left tail. Intuitively, then, \( \rho > 1 \) is the case where one is taking some of the moderate citizens in the population and turning them into extreme agents (and turning most of them into extreme opponents).

- In contrast, when \( p = -\infty \), \( F_{\alpha, \rho}(x) = \min\{S(x), x\} \). In this case, \( F_{\alpha, \rho}(x) \) is taking some mass from the middle of \( S(x) \) and adding it to the right tail. Intuitively, then, \( \rho < 1 \) is the case where one is taking some of the moderate citizens in the population and turning them into extreme agents (and turning most of them into extreme supporters).

Therefore, \( F_{\alpha, \rho}(x) \) is a more polarized version of \( S(x) \), where \( \rho > 1 \) implies more opponents, and \( \rho < 1 \) implies more supporters.

The following is the main result of this subsection.

**Proposition 3.5.** Assume \( S(x) \) the cdf of a single-peaked distribution. Consider the family of distributions defined as \( F_{\alpha, \rho}(x) = (\alpha S(x)^\rho + (1 - \alpha)x^\rho)^{\frac{1}{\rho}} \). For any \( \alpha \in [0, 1] \) and \( \rho \in \mathbb{R} \), \( f_{\alpha, \rho}(\cdot) \) is single-peaked, and thus \( \sigma_1^* = 1 \). Furthermore,

(a) If \( \rho = 1 \), then \( \sigma_0^* \) is independent of \( \alpha \).

(b) If \( \rho > 1 \), then \( \sigma_0^* \) is decreasing in \( \alpha \).

(c) If \( \rho < 1 \), then \( \sigma_0^* \) is increasing in \( \alpha \).
In light of the discussion above, the statement "If \( p > 1 \), then \( \sigma_1^* \) is decreasing in \( \alpha \)." should be read as follows: *When moderate citizens turn into extremist opponents, media becomes less informative.* The intuition behind this is closely related to the intensive versus extensive margin trade-off we presented earlier. The transformation turns some moderate citizens into extremist opponents who are *lost causes*: before they were keen to listen to the media, but now it is very costly for the politician to convince them to pay attention to the news. Consequently, there are now less people in the extensive margin, but still the same number of supporters who listen to the media already. The trade-off therefore leans toward the intensive margin: the politician would rather have a slightly less informative media and have a higher support from people who listen. For the \( p < 1 \) case, the opposite reasoning applies: some moderate people become fierce supporters with prior 1, so the politician can afford to cater a little bit towards opponents by offering a slightly more informative media.

The effect of polarization on the informativeness of news occurs through a channel that has not been analyzed before in the literature. There is an extensive discussion on polarization leading to less informative news (Sunstein, 2001). Proposition 3.5 characterizes under what conditions one would expect to see an association between polarization of opinions and less informative news. Moreover, it argues that if such an association is observed, it may be in the politician’s best interest to allow for less information revelation. The empirical assessment of this association and the channel through it occurs warrants further analysis.

### 3.6 Cost of Listening to Media

In this section, we provide a generalization of the baseline model where the cost of listening to the media can take a large positive value. The general insights from the baseline model go through in this case, but this extension allows us to conduct a comparative statics exercise on the cost of listening to the media.

Suppose that there is a cost \( \kappa > 0 \) of listening to the media. That is, a citizen has to pay \( \kappa > 0 \) if she chooses to observe the signal. This may correspond to the cognitive cost of listening to the media or the monetary cost, and \( \kappa \) in general parametrizes the accessibility of media. Following the same steps as in the derivation of Lemma 3.2, one can easily prove the following statement.

**Lemma 3.4.** Given \((\sigma_0, \sigma_1)\) where \( \sigma_1 \geq \sigma_0 \), a citizen \( i \in I \) watches the news if and only if \( p_i \in [\underline{p}, \bar{p}] \), where:

\[
\underline{p} = \min \left\{ \frac{c\sigma_0 + \kappa}{c\sigma_0 + (1 - c)\sigma_1}, c \right\}
\]

\[
\bar{p} = \max \left\{ \frac{c(1 - \sigma_0) - \kappa}{c(1 - \sigma_0) + (1 - c)(1 - \sigma_1)}, c \right\}
\]

Clearly, if \( \kappa \geq c(1 - c) \), one would have zero measure of citizens following the news: \( \kappa \) is simply too high in this case. Otherwise, an increase in \( \kappa \) leads to a *shrink* in the \([\underline{p}, \bar{p}]\) range.
As a matter of fact, one can simply repeat the analysis with a transformation of \( f(.) \), leaving the extremes values of \( p_i \) out. The transformation of a single-peaked \( f(.) \) is still single-peaked, so the analysis remains the same. The intensive versus extensive trade-off remains the same, and the solution has the same properties. In particular, the general insights from Propositions 3.1-3.3 continue to hold.

Introducing \( \kappa \) as an additional parameter of the model allows us to run comparative statics with respect to accessibility of the media. The following is the main result of this section.

**Proposition 3.6.** Suppose \( f(.) \) is single-peaked. There exists some \( \bar{\kappa} > 0 \) such that for all \( \kappa \in [0, \bar{\kappa}] \), \( \sigma_0^* \) is decreasing in \( \kappa \).

**Proof.** See Appendix C.10.

Proposition 3.6 should be read as: If the captured media is less accessible to citizens, it is more informative in equilibrium. The reason behind this is as follows: as \( \kappa \) increases, the range of individuals who listen to the media shrinks. That is, given a media policy, \( \bar{p} \) decreases and \( p \) increases. When \( \kappa \) is small, however, \( \bar{p} \approx 1 \) and the extensive margin does not change significantly (the blue region in the Figure 3-1 remains almost the same). Since \( p \gg 0 \), the effect on the intensive margin is larger: the politician convinces much more people in the margin when she chooses a slightly more informative media (the red slice in the Figure 3-1 has higher density, because \( p \) increases). As a result, the trade-off leans towards the extensive margin, and the politician sets a more informative media.

Intuitively, with a cost of listening: only more moderate citizens (whose priors are around the critical decision cutoff) listen to the media, because they are the ones who benefit from the news most. But when the distribution is single-peaked, there are many of them, and they are very responsive to the media policy (in the sense that many people start listening even with slight improvements in the informativeness). For the politician, therefore, there are more people to gain in the extensive margin by setting a slightly more informative media, and the optimal decision represents this change in the trade-off.

With a single-dipped density, the symmetric result holds: an increase in \( \kappa \) leads to an increase in \( \sigma_1^* \), once again leading to an increase in informativeness of the captured media. The intuition is similar: a higher \( \kappa \) implies extremist supporters listening to the media for a given policy. But then, the politician can more easily cater towards the other citizens by increasing the informativeness of media.

Proposition 3.6 suggests that a decrease in the accessibility of media may indeed be favorable for the citizens, because the moderate citizens will have access to a more informative media source. Most importantly, media capture may be more of a concern in environments where media is more easily accessible for citizens. The net welfare effects of an increase in the cost of media remains an open question.
3.7 Conclusion

This paper considers media capture as a strategic problem for the politician. On the technical front, it presents the first analysis of Bayesian persuasion towards an audience with heterogeneous priors. On the conceptual front, it extends out understanding of media capture by arguing when, and how, the informativeness of media is suppressed by a politician. In particular, media capture is a serious concern in environments where politician is more popular (Proposition 3.4), and in environments where media is more accessible for citizens (Proposition 3.6). Media capture can take the form of suppressing negative news (Proposition 3.2), positive news (Proposition 3.3), or suppressing the whole informative content of the news (Proposition 3.1) depending on the distribution of opinions in the society. Our model provides a framework for identifying these different cases and quantifying their implications.

On the practical front, our findings generate a group of interesting comparative statics. In addition to the comparative statics about the informativeness of the media with respect to different “types” of polarization of opinions (Proposition 3.5), a qualitative comparison of Proposition 3.2 and 3.3 suggests that, in societies with more extremists abundance of negative news may be preferable for the politician. These comparative statics yield testable implications, which can also be used to distinguish strategic and non-strategic models of media capture.

Our model is a static model with only one politician and one media source. Potential extensions with competition among politicians, with multiple media sources, and dynamic models with evolution of opinions will likely generate additional insights beyond the analysis considered here. Such extensions are left for future work.
Appendix A

Appendix to Chapter 1

A.1 Formal Definition of Equilibrium Concept

Given a history $h^t$ for $t \geq 1$, beliefs $\mu(h^t)$, and a strategy $\sigma_f$ for the firm, one can write the payoffs for the firm and each type of worker by the Law of Iterated Expectations:

$$u_f(\sigma_f| h^t, \mu(h^t)) := \mathbb{E} \left[ \sum_{s=0}^{\infty} \delta^s (\sigma_f(c|h^{t+s})(1-\beta)y^{t+s} + \sigma_f(q|h^{t+s})\theta) \right]$$
with $h^t, \mu(h^t)$

(A.1)

$$u_w(\sigma_f| h^t, \theta) := \mathbb{E} \left[ \sum_{s=0}^{\infty} \delta^s (\sigma_f(c|h^{t+s})\beta y^{t+s} + \sigma_f(q|h^{t+s})\nu) \right]$$
with $h^t, \theta$

(A.2)

Given history $h^0$, beliefs $\mu(h^0)$, and a strategy $\sigma_f$ for the firm, the payoff for the firm is:

$$u_f(\sigma_f| h^0, \mu(h^0), k) := \delta \mathbb{E} [u_f(\sigma_f| h^1 = (h^0, a_f^0 = k), \mu(h^0, a_f^0 = k)|h^0, \mu(h^0)] - c(k)$$

(A.3)

And finally, given a worker with type $\theta \in \Theta$, a strategy $\sigma_f$ for the firm and a strategy $\sigma_w$ for the worker, the payoffs for the worker of type $\theta$ is:

$$u_w(\sigma_w|\theta) := \sigma_w(j|\theta)\delta \mathbb{E}[u_w(\sigma_f| h^0, k = \sigma_f(h^0), \theta)] + \sigma_w(n|\theta)\nu$$

(A.4)

Definition A.1. A Perfect Bayesian Equilibrium (PBE) of this game is a strategy profile $(\sigma^*_w, \sigma^*_f)$ and beliefs $\{\mu(h^t)\}_{h^t \in H}$ such that:

- Given $(\sigma^*_f)$, $\sigma^*_w$ maximizes (A.4) for all $\theta \in \Theta$.
- Given $(\sigma^*_w)$ and $\{\mu(h^t)\}_{h^t \in H}$, $\sigma^*_f$ maximizes (A.3) for $h^0$ and (A.1) for all $h^t$ with $t \geq 1$.
- $\{\mu(h^t)\}_{h^t \in H}$ satisfies (1.3) and (1.4).
A.2 Proofs in Section 1.3.1

The analysis of the firm’s problem advances in backward induction. I begin by solving for the firm’s continuation decision following a capital choice (Section A.2.1). This yields a closed form expression for the firm’s continuation payoffs (Section A.2.2). I then solve for the firm’s optimal capital choice (Section A.2.3), taking the continuation payoffs into account. Section A.2.4 discusses the role played by Assumption 1.2. Proof of Proposition 1.1 follows from these arguments (Section A.2.5).

A.2.1 Continuation Decision

The game that begins in \( t = 1 \), following history \( h^1 = (a_0^0 = j, a_f^0 = k) \) with beliefs \( \mu(h^1) \), is a dynamic decision problem for the firm because the worker never makes a choice. Each period, the firm observes an output that is informative about the worker’s type and decides whether to quit the relationship or continue.

Let \( \mathcal{H}_k \) be the set of all histories under consideration in this subsection, following \( h^1 \):

\[
\mathcal{H}_k := \{ h^t \in \mathcal{H} : t \geq 1, h^t = (a_0^0 = j, a_f^0 = k; \ldots ; t^{t-1}) \}
\]

In any PBE, for any \( h^t \in \mathcal{H}_k \) and \( \mu(h^t) \), \( \sigma^*_t(c|h^t) \) must maximize firm’s payoff given in Equation (A.1). The maximization problem is:

\[
\sigma^*_t(c|h^t) \in \arg \max_{\sigma \in [0,1]} \sigma (1 - \beta) E[y^t|h^t, \mu(h^t)] \\
+ \sigma \mathbb{E} \left[ \sum_{s=1}^{\infty} \delta^s (\sigma^*_s(c|h^{t+s})(1 - \beta)y^{t+s} + \sigma^*_s(q|h^{t+s})\pi) \bigg| h^t, \mu(h^t) \right] \\
+ (1 - \sigma)\pi
\]

I will begin by losing the history, \( h^t \), as an argument in the firm’s problem.

Observation A.1. Due to the Markovian nature of the problem (imposed by the production function being i.i.d. conditional on type and capital choice), for any \( h^t \in \mathcal{H}_k \),

\[
E[y^t|h^t, \mu(h^t)] = (\eta_0 + (1 - \eta_0)\mu(h^t)) \cdot k
\]

Therefore, \( (\mu(h^t), k) \) is a sufficient statistic for \( h^t \in \mathcal{H}_k \).

Thanks to Observation A.1, the firm’s decision problem is reduced to a dynamic programming problem with state variables \( (\mu(h^t), k) \).

Consider \( \mu(h^{t+1}) \) conditional on \( h^t \). From the perspective of \( h^t \) (and consequently, conditional
on $\mu(h^t)), \mu(h^{t+1})$ is a random variable, whose distribution is given through Equation (1.4):

$$
\mu(h^{t+1}) = \begin{cases} 
\mu(h^t) \text{ w.p. } \eta_0, \\
0 \text{ w.p. } (1 - \mu(h^t)) \cdot (1 - \eta_0)
\end{cases}
$$

The next step is stating this optimization problem through its recursive formulation, which exists through standard arguments (Davis, 1993). It is:

$$
V(\mu(h^t), k) = \max_{\sigma \in \{0,1\}} \left( \left( (\mu(h^t) + (1 - \mu(h^t)) \eta_0) \cdot (1 - \beta) k + \delta \mathbb{E} [V(\mu(h^{t+1}), k)|\mu(h^t)] \right) + (1 - \sigma) \pi \right)
$$

(A.5)

where

$$
\mathbb{E} [V(\mu(h^{t+1}), k)|\mu(h^t)] = (\mu(h^t) + (1 - \mu(h^t)) \eta_0) \cdot V\left( \frac{\mu(h^t)}{\mu(h^t) + (1 - \mu(h^t)) \eta_0}, k \right) + (1 - \mu(h^t))(1 - \eta_0)V(0, k)
$$

(A.6)

The optimal policy of the dynamic programming problem is:

$$
\sigma(\mu(h^t), k) := \sigma_f(c|h^t, \mu(h^t))
$$

The following Lemma characterizes the value function and the optimal policy.

**Lemma A.1.** Given a value of $k$, there is a cutoff belief $\mu^* \in [0,1]$ such that

$$
\sigma(\mu, k) = \begin{cases} 
1, & \text{if } \mu \geq \mu^* \\
0, & \text{otherwise}
\end{cases}
$$

$V(\mu, k)$ is continuous and nondecreasing in $\mu$, and strictly increasing in $\mu$ for $\mu \geq \mu^*$. 

**Proof.** Continuity of $V(\mu, k)$ follows from the flow payoff being continuous and bounded. $V(\mu, k)$ being increasing in $\mu$ follows from the flow payoff, $(\mu + (1 - \mu) \eta_0)(1 - \beta)k$, being increasing in $\mu$. These two features, and linearity of the objective function in $\sigma$, implies that the policy must be of the form characterized in the statement of the Lemma. $V(\mu, k)$ being strictly increasing in $\mu$ for $\mu \geq \mu^*$ follows from the form of the optimal policy, and the flow payoff being strictly increasing in $\mu$. 

Below is a picture of $V(\mu, k)$ for a given value of $k$, to make the exposition easier.
Figure A-1: The value function in Lemma A.1.

This figure is the mirror image of Figure 1 in Keller and Rady (2015), and serves the same purpose. The arrows to the right of $\mu^*$ are deliberate: by Lemma A.1, this is the region where the firm experiments. As long as high output is observed, firm’s beliefs improve, so $\mu$ increases and the firm keeps experimenting. As soon as low output is observed, firm’s beliefs drop to $\mu = 0$.

Whereas Lemma A.1 proposes a general shape for the value function, it does not rule out the possibility of corner cases: $\mu^* \in \{0,1\}$ is possible. That is, the firm may always continue regardless of the output ($\mu^* = 0$) or immediately quit without observing any output ($\mu^* = 1$). To characterize the ranges where these cases occur, I will partition $\mathcal{H}_k$ into two classes. The first class involves the set of histories where the output has always been high (including the initial history where no output is observed yet). The second class involves the set of histories that contain at least one low output, i.e. the firm has conclusively learned that the worker is low-type (and thus $\mu = 0$).

Finally, I will define a quantity that is critical for the next result. Given $\mu^1 \in [0, 1]$, define the experimentation threshold as:

$$
\gamma(\mu^1) := \frac{\mu^1(1 - \eta_0) + \eta_0(1 - \delta)}{\mu^1(1 - \eta_0)\delta + (1 - \delta)} \in [0, 1]
$$

(A.7)

The following is the main result of this subsection.
Proposition A.1. In the Perfect Bayesian Equilibrium of the continuation game, following history \( h^1 = (a_0^0 = j, a_f^0 = k) \) with beliefs \( \mu^1 := \mu(h^1) \),

- For any \( h^t, h'^t \in H^1_k \) with \( t' \geq t \), \( \sigma_j^*(c|h'^t) \geq \sigma_j^*(c|h^t) \).

  - Moreover, for \( h^1 \in H^1_k \),
    \[
    \sigma_j^*(c|h^1) = \begin{cases} 
    1, & \text{if } \pi \leq \frac{1}{1-\delta}(1-\beta)k \cdot \gamma(\mu^1) \\
    0, & \text{otherwise}.
    \end{cases}
    \]

- For any \( h^t, h'^t \in H^2_k \), \( \sigma_j^*(c|h^t) = \sigma_j^*(c|h'^t) \).

  - Moreover, for all \( h^t \in H^2_k \),
    \[
    \sigma_j^*(c|h^t) = \begin{cases} 
    1, & \text{if } \pi \leq \frac{1}{1-\delta}(1-\beta)k \cdot \eta_0 \\
    0, & \text{otherwise}.
    \end{cases}
    \]

Proposition A.1 says that the firm's decision problem following a capital decision is characterized by two continuation decisions.

1. One of these is a continuation decision following a history of only high outputs. If the outside option is sufficiently attractive given \( k \), the firm quits immediately, without observing any output. If the outside option is not attractive enough, the firm continues as long as the high output is observed. In this case, the firm effectively adopts an experimentation strategy: the firm learns about the worker's type over time by continuing with the relationship. The cutoff value of \( \pi \) for experimentation depends on \( \mu^1 \) through \( \gamma(\mu^1) \).

2. The other continuation decision is the one following a history with at least one low output, i.e. after the firm has conclusively learned that the worker’s type is low. Similar to the case above, if the outside option is sufficiently attractive, the firm quits after such a history. Otherwise, the firm continues even after the low output is observed. In this case, the relationship is never terminated, regardless of the output.

Below is a graphical illustration of Proposition A.1 for different values of \( \pi \) and \( \mu^1 \) for a given \( k \).
Proof. (of Proposition A.1.)

Lemma A.1 yields the form of the optimal policy for the firm. To characterize the cutoff $\mu^* \in [0, 1]$, first consider the optimal policy in the two extremes: $\mu \in \{0, 1\}$.

- For $\mu = 0$, since there is no further Bayesian updating, Equation (A.5) simplifies to:
  \[ V(0, k) = \max_{\sigma \in [0, 1]} \sigma \left( (1 - \beta)\eta_0 k + \delta V(0, k) \right) + (1 - \sigma)\pi \]
  Clearly, $\sigma(0, k) = 1$ if and only if $\frac{1}{1 - \delta} (1 - \beta)\eta_0 k \geq \pi$. Consequently, $\mu^* = 0$ when $\frac{1}{1 - \delta} (1 - \beta)\eta_0 k \geq \pi$. In this case,
  \[ V(\mu, k) = \frac{1}{1 - \delta} (1 - \beta)(\mu + (1 - \mu)\eta_0)k \]
  for all $\mu \in [0, 1]$.

- For $\mu = 1$, Equation (A.5) simplifies to:
  \[ V(1, k) = \max_{\sigma \in [0, 1]} \sigma \left( (1 - \beta)k + \delta V(1, k) \right) + (1 - \sigma)\pi \]
  Clearly, $\sigma(1, k) = 0$ if and only if $\frac{1}{1 - \delta} (1 - \beta)k < \pi$. Consequently, $\mu^* = 1$ when $\frac{1}{1 - \delta} (1 - \beta)k < \pi$. In this case, \[ V(\mu, k) = \pi \]
  for all $\mu \in [0, 1]$. 

Figure A-2: Employment dynamics.
It is immediate from the discussion above that,

- If \( \frac{1}{1-\delta}(1-\beta)\eta_0 k \geq \pi, \mu^* = 0 \) and \( \sigma_f(c|h^t) = 1 \) for all \( h^t \in \mathcal{H}_k = \mathcal{H}_k^1 \cup \mathcal{H}_k^2 \).
- If \( \frac{1}{1-\delta}(1-\beta)k < \pi, \mu^* = 1 \) and \( \sigma_f(c|h^t) = 0 \) for all \( h^t \in \mathcal{H}_k = \mathcal{H}_k^1 \cup \mathcal{H}_k^2 \).
- Otherwise, \( \mu^* \in (0,1) \).

For the rest of the proof, consider the case \( \frac{1}{1-\delta}(1-\beta)k \in [\pi, \frac{\pi}{\eta_0}) \).

Recall that, for any \( h^t \in \mathcal{H}_k^2, \mu(\theta = 1|h^t) = 0 \). Since \( \frac{1}{1-\delta}(1-\beta) < \frac{\pi}{\eta_0} \), \( \sigma_f(c|h^t) = 0 \) and \( V(0,k) = \pi \). This completes the proof of the second part of Proposition A.1.

Take two histories \( h^t, h'^t \in \mathcal{H}_k^1 \) with \( t' \geq t \). In particular, since \( h'^t \in \mathcal{H}_k^1 \), high output must be observed in periods \( \tau \in \{1, \ldots, t, \ldots, t'-1\} \). Through repeated application of Bayesian updating in Equation (1.4), one has:

\[
\mu(h'^t) = \frac{\mu(h^t)}{\mu(h^t) + (1-\mu(h^t))\eta_0^{\mu^*}} \geq \mu(h^t)
\]

The cutoff structure of optimal policy in Lemma A.1 immediately implies one must have:

\[
\sigma_f^*(c|h'^t) \geq \sigma_f^*(c|h^t)
\]

What remains is to prove that \( \sigma_f^*(c|h^t) = 1 \) if and only if \( \pi \leq \frac{1}{1-\delta}k \cdot \gamma(\mu^*). \) Note that the firm has an option to quit following history \( h^t \), which yields a payoff of \( \pi \). Alternatively, the firm may set \( \sigma_f^*(c|h^t) = 1 \). If this is the case, it must be that:

- \( \sigma_f^*(c|h^t) = 1 \) for all \( h^t \in \mathcal{H}_k^1 \), because \( t \geq 1 \) and \( \sigma_f^*(c|h^t) \geq \sigma_f^*(c|h^1) \).
- \( \sigma_f^*(c|h^t) = 0 \) for all \( h^t \in \mathcal{H}_k^2 \), because \( \frac{1}{1-\delta}(1-\beta)k < \frac{\pi}{\eta_0} \).

If the worker has type \( \theta = 1 \) (with probability \( \mu^0 \)), this yields a payoff of:

\[
V_f(k,1,\pi) := \frac{1}{1-\delta}(1-\beta)k
\]

and, if the worker has type \( \theta = 1 \) (with probability \( 1-\mu^0 \)), this yields a payoff of:

\[
V_f(k,0,\pi) := \sum_{s=1}^{\infty} \frac{(\eta_0)^s(1-\eta_0)}{Pr\{\text{worker produces } s \text{ high outputs}\}} + \sum_{s=0}^{\infty} \frac{\delta^{s+1}(1-\beta)k}{Pr\{\text{worker produces a high outputs}\}} \quad \text{payoff after quitting}
\]

89
Therefore,

\[ V_f(k, 0, \pi) = (1 - \eta_0)(1 - \beta)k \sum_{s=1}^{\infty} \left( \eta_0 \right)^s \sum_{\tau=0}^{s-1} \delta^\tau + (1 - \eta_0)\pi \delta \sum_{s=0}^{\infty} \left( \eta_0 \right)^s \delta^s \]

\[ = \frac{1 - \delta}{1 - \eta_0 \delta} \frac{1}{1 - \delta} \eta_0 (1 - \beta)k + \frac{(1 - \eta_0)\delta}{1 - \eta_0 \delta} \pi \]

Therefore, the expected payoff of the firm from \( \sigma_f(c|h^1) = 1 \) is:

\[
E[V_f(k, \theta, \pi)|\mu^0] := \mu^1 V(k, 1, \pi) + (1 - \mu^1) V(k, 0, \pi)
\]

\[ = \mu^1 \frac{1}{1 - \delta} (1 - \beta)k + (1 - \mu^1) \left( \frac{1 - \delta}{1 - \eta_0 \delta} \frac{1}{1 - \delta} (1 - \beta)\eta_0 k + \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi \right) \]

For \( \sigma^*_f(c|h^1) = 1 \) to be optimal, this payoff must prevail the one from quitting, i.e. \( \pi \). Therefore, \( \sigma^*_f(c|h^1) = 1 \) if and only if:

\[
\mu^1 \frac{1}{1 - \delta} (1 - \beta)k + (1 - \mu^1) \left( \frac{1 - \delta}{1 - \eta_0 \delta} \frac{1}{1 - \delta} (1 - \beta)\eta_0 k + \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi \right) \geq \pi
\]

\[
\iff \frac{1}{1 - \delta} (1 - \beta)k \mu^1 (1 - \eta_0) + \eta_0 (1 - \delta) \geq \pi
\]

\[
\iff \frac{1}{1 - \delta} (1 - \beta)k \cdot \gamma(\mu^1) \geq \pi
\]

where \( \gamma(\mu^1) \) is as defined in Equation (A.7). The result follows. \Box

### A.2.2 The Firm’s Continuation Payoffs

With Proposition A.1 in hand, one can now derive the expected payoffs for the firm following \( h^1 = (a^0_w = j, a^0_f = k) \) with beliefs \( \mu^1 \). By the discussion in proof of Proposition A.1, its continuation payoff is given by:

\[
V_f(k, \mu^1, \pi) = \begin{cases} 
\frac{1 - \delta}{1 - \beta} (1 - \beta)(\mu^1 + (1 - \mu^1)\eta_0)k, & \text{if } k \geq \frac{1 - \delta}{1 - \beta} \frac{1}{\eta_0} \pi \\
\frac{1 - \delta}{1 - \beta} (1 - \beta) \left( \mu^1 + (1 - \mu^1)\eta_0 \frac{1 - \delta}{1 - \eta_0 \delta} \right) k + (1 - \mu^1) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi, & \text{if } k \in \left[ \frac{1 - \delta}{1 - \beta} \frac{1}{\gamma(\mu^1) \eta_0}, \frac{1 - \delta}{1 - \beta} \frac{1}{\eta_0} \pi \right] \\
\pi, & \text{if } k \leq \frac{1 - \delta}{1 - \beta} \gamma(\mu^1) \pi 
\end{cases}
\]
Note that $V_f(k, \mu^1, \pi)$ is a piecewise linear and continuous function of $k$. Below is an illustration of Equation (A.8), demonstrating firm’s continuation payoffs as a function of $k$.

![Figure A-3: Continuation payoffs for the firm.](image)

### A.2.3 Capital Choice

This section characterizes the firm’s capital choice following a history $h^0 = (a^0_w = j)$ with beliefs $\mu(h^0)$. Since the worker does not take any action after $h^0$ and since no output is observed until after $h^1$, Bayesian updating does not occur between $h^0$ and $h^1 = (a^0_w = j, a^0_f = k)$. Consequently:

$$\mu(h^0) = \mu(h^1) = \mu^1$$

Let:

$$\mu^0 := \mu(h^0)$$

Since $\mu^1 = \mu^0$, the firm knows about its continuation payoffs following a capital choice $k \in \mathbb{R}^+$: it is given by Equation (A.8). The investment is made only once, so the capital choice is a static optimization problem. Given $\mu^0$ and $\pi$, the problem is:

$$k^\ast(\mu^0, \pi) = \arg \max_{k \in \mathbb{R}^+} \delta V_f(k, \mu^0, \pi) - c(k) \quad (A.9)$$

A detailed exploration of the solution requires defining some objects.

---

1Piecewise linearity is immediate, and continuity can be observed by checking the left and right limits at the kinks whereas using Equation (A.7).
• Given a belief $\mu^0 \in [0, 1]$, define the unconstrained investment problem as:

$$\overline{S}(\mu^0) = \max_{k \in \mathbb{R}^+} \frac{1}{1 - \delta} \frac{1}{1 - \delta} (1 - \beta) (\mu^0 + (1 - \mu^0) \eta_0) k - \frac{c(k)}{\delta}$$  \hspace{1cm} (A.10)

This is the investment problem faced by a firm whose belief is $\mu^0$, and who knows that the relationship will continue forever, i.e. the quitting will never occur in the future. Under Assumption 1.1, the objective function is strictly concave, the optimizer is unique and strictly positive. Its optimizer is:

$$\overline{k}(\mu^0) := \arg \max_{k \in \mathbb{R}^+} \frac{\delta}{1 - \delta} (1 - \beta) \cdot (\mu^0 + (1 - \mu^0) \eta_0) \cdot k - c(k)$$  \hspace{1cm} (A.11)

• Given a belief $\mu^0 \in [0, 1]$, define the constrained investment problem as:

$$\underline{S}(\mu^0) = \max_{k \in \mathbb{R}^+} \frac{1}{1 - \delta} \frac{1}{1 - \delta} (1 - \beta) (\mu^0 + (1 - \mu^0) \eta_0) \frac{1 - \delta}{1 - \eta_0 \delta} k - \frac{c(k)}{\delta}$$  \hspace{1cm} (A.12)

This is the investment problem faced by a firm whose belief is $\mu^0$, and who knows that the relationship will be terminated following a low output. Under Assumption 1.1, the objective function is also strictly concave, the optimizer is unique and strictly positive. Its optimizer is:

$$\underline{k}(\mu^0) := \arg \max_{k \in \mathbb{R}^+} \frac{\delta}{1 - \delta} (1 - \beta) \cdot (\mu^0 + (1 - \mu^0) \eta_0) \frac{1 - \delta}{1 - \eta_0 \delta} \cdot k - c(k)$$  \hspace{1cm} (A.13)

The following observation easily follows from the fact that the return to $k$ in (A.11) is larger than the return to $k$ in (A.13) for any $k$.

**Observation A.2.** For any $\mu^0 \in [0, 1)$, $\overline{S}(\mu^0) > \underline{S}(\mu^0)$ and $\overline{k}(\mu^0) > \underline{k}(\mu^0)$.

The following observation follows from the complementarity of capital and labor, and usual supermodularity arguments (Topkis, 2011).

**Observation A.3.** $\overline{S}(\mu^0)$, $\underline{k}(\mu^0)$, $\underline{S}(\mu^0)$ and $\underline{k}(\mu^0)$ are strictly increasing in $\mu^0$.

Given Proposition A.1, the optimal capital choice essentially involves identifying the maximizer of the optimization problem in Equation (A.9), which has a continuous and piecewise concave objective function. Below is an illustration of the maximization problem the firm faces.

---

\footnote{The normalization of $c(k)$ by $\frac{1}{\delta}$ appears due to the timing assumption: the firm chooses the capital level at $t = 0$ and the returns do not accrue until $t = 1$.}
The objective function is a combination of three different concave functions:

1. The leftmost piece (the one that applies when $k \leq \frac{1-\delta}{1-\beta \eta_0} \pi$) is the curve
   \[
   \delta \pi - c(k)
   \]
   whose maximizer is $k = 0$ and whose maximum value is $\delta \pi$.

2. The piece in the middle (the one that applies when $k \in [\frac{1-\delta}{1-\beta \eta_0} \pi, \frac{1-\delta}{1-\beta \eta_0} \pi]$) is the curve
   \[
   \frac{\delta}{1-\delta} (1-\beta)(\mu^0 + (1-\mu^0) \frac{1-\delta}{1-\eta_0 \delta} \eta_0) k + \delta (1-\mu^0) \frac{\delta - \eta_0 \delta}{1-\eta_0 \delta} \pi - c(k)
   \]
   whose maximizer is $k(\mu^0)$ and whose maximum value is $\delta \overline{S}(\mu^0) + \delta (1-\mu^0) \frac{\delta - \eta_0 \delta}{1-\eta_0 \delta} \pi$, where $\overline{S}(\mu^0)$ is as defined in Equation (A.12) and $k(\mu^0)$ is as defined in Equation (A.13).

3. The rightmost piece (the one that applies when $k \geq \frac{1-\delta}{1-\beta \mu_0} \pi$) is the curve
   \[
   \frac{\delta}{1-\delta} (1-\beta)(\mu^0 + (1-\mu^0) \eta_0) k - c(k)
   \]
   whose maximizer is $\overline{k}(\mu^0)$ and whose maximum value is $\delta \overline{S}(\mu^0)$, where $\overline{S}(\mu^0)$ is as defined in Equation (A.10) and $\overline{k}(\mu^0)$ is as defined in Equation (A.11).

The derivation of optimal capital choice proceeds in two steps. I will begin by solving an
auxiliary maximization problem, given by:

\[ V_f(\mu^0, \pi) := \max \{ \delta \pi, \delta \bar{S}(\mu^0) + \delta(1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi, \delta \bar{S}(\mu^0) \} \]  \hspace{1cm} (A.14)

\[ V_f(\mu^0, \pi) \text{ corresponds to comparing the "unconstrained" maxima of three pieces. I will then demonstrate that } V_f(\mu^0, \pi) \text{ coincides with the maximum value of the objective function in Equation (A.9) (Lemma A.3).} \]

Given \( \mu^0 \), let:

\[ \pi_1(\mu^0) := \frac{\bar{S}(\mu^0) - \bar{S}(\mu^0)}{(1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta}} \]  \hspace{1cm} (A.15)

\[ \pi_2(\mu^0) := \frac{\bar{S}(\mu^0)}{\mu^0 + (1 - \mu^0) \frac{1 - \delta}{1 - \eta_0 \delta}} \]  \hspace{1cm} (A.16)

**Lemma A.2.** If

\[ \pi_1(\mu^0) \geq \pi_2(\mu^0) \]

then

\[ V_f(\mu^0, \pi) = \max \{ \delta \pi, \bar{S}(\mu^0) \} \]

Otherwise,

\[ V_f(\mu^0, \pi) = \begin{cases} 
\delta \bar{S}(\mu^0), & \text{if } \pi \leq \pi_1(\mu^0) \\
\delta \bar{S}(\mu^0) + \delta(1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi, & \text{if } \pi \in [\pi_1(\mu^0), \pi_2(\mu^0)] \\
\delta \pi, & \text{if } \pi \geq \pi_2(\mu^0) 
\end{cases} \]

The proof consists of straightforwardly checking the optimality conditions, and has an intuitive pictorial illustration, provided below.
Proof. (of Lemma A.2.) The proof considers the two cases separately.

1. Assume

\[
\frac{S(\mu^0) - S(\mu^0)}{(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta}} \geq \frac{S(\mu^0)}{\mu^0 + (1 - \mu^0)\frac{1 - \delta}{1 - \eta_0\delta}}
\] (A.17)

Two preliminary observations in this case are as follows.

**Claim A.1.** \(\delta S(\mu^0) + \delta(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta} \pi > \delta\pi \implies \delta\bar{S}(\mu^0) > \delta S(\mu^0) + \delta(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta} \pi.\)

Proof. Assume \(\delta S(\mu^0) + \delta(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta} \pi > \delta\pi.\) Rearranging gives: \(\pi < \frac{S(\mu^0)}{\mu^0 + (1 - \mu^0)\frac{1 - \delta}{1 - \eta_0\delta}}.\)

By Equation (A.17), this implies \(S(\mu^0) - S(\mu^0) > \bar{S}(\mu^0) > \pi.\) Rearranging gives: \(\delta\bar{S}(\mu^0) > \delta S(\mu^0) + \delta(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta} \pi.\)

**Claim A.2.** \(\delta S(\mu^0) + \delta(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta} \pi > \delta\bar{S}(\mu^0) \implies \delta\pi > \delta S(\mu^0) + \delta(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta} \pi.\)

Proof. Assume \(\delta S(\mu^0) + \delta(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta} \pi > \bar{S}(\mu^0).\) Rearranging gives: \(\pi > \frac{S(\mu^0) - S(\mu^0)}{(1 - \mu^0)\frac{1 - \delta}{1 - \eta_0\delta}}.\)

By Equation (A.17), this implies \(\pi > \frac{S(\mu^0)}{\mu^0 + (1 - \mu^0)\frac{1 - \delta}{1 - \eta_0\delta}}.\) Rearranging gives: \(\delta\pi > \delta S(\mu^0) + \delta(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta} \pi.\)

Claim A.1 and A.2 combined implies \(\delta S(\mu^0) + \delta(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta} \pi \leq \max\{\delta\pi, \delta\bar{S}(\mu^0)\}.\) It follows that in this case, \(\bar{V}(\mu^0, \pi) = \max\{\delta\pi, \delta\bar{S}(\mu^0)\}.\)

2. Assume

\[
\frac{S(\mu^0) - S(\mu^0)}{(1 - \mu^0)\frac{\delta - \eta_0\delta}{1 - \eta_0\delta}} < \frac{S(\mu^0)}{\mu^0 + (1 - \mu^0)\frac{1 - \delta}{1 - \eta_0\delta}}
\] (A.18)
In this case,

(a) Suppose

\[
\pi \leq \frac{\bar{S}(\mu^0) - S(\mu^0)}{1 - \mu^0 \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta}} \tag{A.19}
\]

By Equation (A.18), this implies: \( \pi < \frac{S(\mu^0)}{\mu^0 + (1 - \mu^0) \frac{1 - \eta_0 \delta}{1 - \eta_0 \delta}} \). Rearranging gives:

\[
\delta S(\mu^0) + \delta (1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi > \delta \pi \tag{A.20}
\]

Also, rearranging Equation (A.19) gives:

\[
\delta S(\mu^0) + \delta (1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi \geq \delta \pi \tag{A.21}
\]

Combining Equation (A.20) and (A.21), one concludes: \( \nu_f(\mu^0, \pi) = \delta \bar{S}(\mu^0) \).

(b) Suppose

\[
\pi \in \left[ \frac{\bar{S}(\mu^0) - S(\mu^0)}{1 - \mu^0 \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta}}, \frac{\bar{S}(\mu^0)}{\mu^0 + (1 - \mu^0) \frac{1 - \delta}{1 - \eta_0 \delta}} \right] \tag{A.22}
\]

Rearranging \( \pi \geq \frac{\bar{S}(\mu^0) - S(\mu^0)}{1 - \mu^0 \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta}} \) yields: \( \delta S(\mu^0) + \delta (1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi \geq \delta \bar{S}(\mu^0) \). Rearranging \( \pi \leq \frac{S(\mu^0)}{\mu^0 + (1 - \mu^0) \frac{1 - \delta}{1 - \eta_0 \delta}} \) yields: \( \delta S(\mu^0) + \delta (1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi \geq \delta \pi \). Combining, one concludes: \( \nu_f(\mu^0, \pi) = \delta \bar{S}(\mu^0) + \delta (1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi \).

(c) Suppose

\[
\pi \geq \frac{\bar{S}(\mu^0)}{\mu^0 + (1 - \mu^0) \frac{1 - \delta}{1 - \eta_0 \delta}} \tag{A.23}
\]

By Equation (A.18), this implies: \( \pi > \frac{\bar{S}(\mu^0) - S(\mu^0)}{1 - \mu^0 \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta}} \). Rearranging gives:

\[
\delta S(\mu^0) + \delta (1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi > \delta \bar{S}(\mu^0) \tag{A.24}
\]

Also, rearranging Equation (A.23) gives:

\[
\delta \pi \geq \delta S(\mu^0) + \delta (1 - \mu^0) \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} \pi \tag{A.25}
\]

Combining Equation (A.24) and (A.25), one concludes: \( \nu_f(\mu^0, \pi) = \delta \pi \).

Since all the cases considered above are exhaustive, the proof follows. \( \square \)
For ease of exposition, the maximization problem in Equation (A.9) is reproduced below:

\[ k^*(\mu^0, \pi) := \arg \max_{k \in \mathbb{R}^+} \delta V_f(k, \mu^0, \pi) - c(k) \]

where \( V_f(k, \mu^0, \pi) \) is as given in Equation (A.8). Therefore:

\[ \delta V_f(k, \mu^0, \pi) - c(k) = \begin{cases} \frac{1}{1-\delta}(1-\beta)(\mu^0 + (1-\mu^0)p_0)k - c(k), & \text{if } k > \frac{1-\delta}{1-\beta} \frac{1}{\eta_0} \pi \\ \frac{1-\delta}{1-\beta}(1-\beta)(\mu^0 + (1-\mu^0)\frac{1-\eta_0}{1-\eta_0} \eta_0)k + \delta(1-\mu^0)\frac{1-\eta_0}{1-\eta_0} \pi - c(k), & \text{if } k \in \left[ \frac{1-\delta}{1-\beta} \frac{1}{\eta_0} \pi, \frac{1-\delta}{1-\beta} \frac{1}{\eta_0} \pi \right] \\ \delta \pi - c(k), & \text{if } k \leq \frac{1-\delta}{1-\beta} \frac{1}{\eta_0} \pi \end{cases} \]

Also, define the maximum of this problem as

\[ V_f^*(\mu^0, \pi) := \max_{k \in \mathbb{R}^+} \delta V_f(k, \mu^0, \pi) - c(k) \]

Lemma A.3. \( V_f(\mu^0, \pi) = V_f^*(\mu^0, \pi) \). Moreover,

- If \( \pi_1(\mu^0) \geq \pi_2(\mu^0) \)
  then
  \[ k^*(\mu^0, \pi) = \begin{cases} \kappa(\mu^0), & \text{if } \pi < \underline{\pi}(\mu^0) \\ \kappa(\mu^0), & \text{if } \pi = \underline{\pi}(\mu^0) \\ 0, & \text{if } \pi \geq \underline{\pi}(\mu^0) \end{cases} \]

- Otherwise,
  \[ k^*(\mu^0, \pi) = \begin{cases} \kappa(\mu^0), & \text{if } \pi < \pi_1(\mu^0) \\ \kappa(\mu^0), & \text{if } \pi = \pi_1(\mu^0) \\ \kappa(\mu^0), & \text{if } \pi \in (\pi_1(\mu^0), \pi_2(\mu^0)) \\ \kappa(\mu^0), & \text{if } \pi = \pi_2(\mu^0) \\ 0, & \text{if } \pi > \pi_2(\mu^0) \end{cases} \]

The proof consists of verifying that the unconstrained maxima of the auxiliary maximization problem is attained in the constrained problem as well.

**Proof. (of Lemma A.3.)** For expositional simplicity throughout this proof, let:

\[ \alpha := \frac{1-\delta}{1-\eta_0 \delta}, \quad 1-\alpha := \frac{\delta - \eta_0 \delta}{1-\eta_0 \delta} \]

The proof considers three cases.

1. Suppose \( V_f(\mu^0) = \delta \pi \), i.e. \( \pi \geq \max \{ \delta \underline{\pi}(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi, \delta \bar{\pi}(\mu^0) \} \).

I will demonstrate that \( k^*(\mu^0, \pi) = 0 \), i.e. \( \delta V_f(0, \mu^0, \pi) - c(0) \geq \delta V_f(k, \mu^0, \pi) - c(k) \) for
all $k \geq 0$.

Note that in this case, $\delta V_f(0, \mu^0, \pi) - c(0) = \delta \pi$, because:

$$0 \leq \frac{1 - \delta}{1 - \frac{1}{\gamma(\mu^0)}} \pi$$

and $c(0) = 0$ by Assumption 1.1.

- For $k \in (0, \frac{1 - \delta}{1 - \frac{1}{\gamma(\mu^0)}} \pi)$,

  $\delta V_f(0, \mu^0, \pi) - c(0) = \delta \pi$

  $> \delta \pi - c(k)$

  $= \delta V_f(k, \mu^0, \pi) - c(k)$

  where the inequality follows by Assumption 1.1.

- For $k \in \left[\frac{1 - \delta}{1 - \frac{1}{\gamma(\mu^0)}} \pi, \frac{1 - \delta}{1 - \frac{1}{\gamma(\mu^0)}} \pi\right]$,

  $\delta V_f(0, \mu^0, \pi) - c(0) = \delta \pi$

  $\geq \delta S(\mu^0) + \delta(1 - \mu^0)(1 - \alpha)\pi$

  $= \delta \left(\frac{1}{1 - \delta}(1 - \beta)(\mu^0 + (1 - \mu^0)\alpha \eta_0)k(\mu^0) - \frac{c(k(\mu^0))}{\delta}\right) + \delta(1 - \mu^0)(1 - \alpha)\pi$

  $\geq \delta \left(\frac{1}{1 - \delta}(1 - \beta)(\mu^0 + (1 - \mu^0)\alpha \eta_0)k - \frac{c(k)}{\delta}\right) + \delta(1 - \mu^0)(1 - \alpha)\pi$

  $= \frac{\delta}{1 - \delta}(1 - \beta)(\mu^0 + (1 - \mu^0)\alpha \eta_0)k + \delta(1 - \mu^0)(1 - \alpha)\pi - c(k)$

  $= \delta V_f(k, \mu^0, \pi) - c(k)$

  where the first inequality holds in the case considered here, and the second inequality follows by Equation (A.13).

- For $k \geq \frac{1 - \delta}{1 - \frac{1}{\gamma(\mu^0)}} \pi$,

  $\delta V_f(0, \mu^0, \pi) - c(0) = \delta \pi$

  $\geq \delta S(\mu^0)$

  $= \delta \left(\frac{1}{1 - \delta}(1 - \beta)(\mu^0 + (1 - \mu^0)\eta_0)k(\mu^0) - \frac{c(k(\mu^0))}{\delta}\right)$

  $\geq \delta \left(\frac{1}{1 - \delta}(1 - \beta)(\mu^0 + (1 - \mu^0)\eta_0)k - \frac{c(k)}{\delta}\right)$

  $= \frac{\delta}{1 - \delta}(1 - \beta)(\mu^0 + (1 - \mu^0)\eta_0)k - c(k)$

  $= \delta V_f(k, \mu^0, \pi) - c(k)$

  where the first inequality holds in the case considered here, and the second inequality follows by Equation (A.11).
2. Suppose $V_f(\mu^0) = \delta S(\mu^0)$, i.e. $\delta S(\mu^0) \geq \max\{\delta \pi, \delta S(\mu^0) + \delta (1 - \mu^0)(1 - \alpha)\pi\}$.

I will demonstrate that $k^*(\mu^0, \pi) = \bar{k}(\mu^0)$, i.e. $\delta V_f(\bar{k}(\mu^0), \mu^0, \pi) - c(\bar{k}(\mu^0)) \geq \delta V_f(k, \mu^0, \pi) - c(k)$ for all $k \geq 0$.

My first claim is that in this case, $\delta V_f(\bar{k}(\mu^0), \mu^0, \pi) - c(\bar{k}(\mu^0)) = \frac{\delta}{1 - \delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\eta_0)\bar{k}(\mu^0) - c(\bar{k}(\mu^0))$. This is equivalent to showing that $\bar{k}(\mu^0) \geq \frac{1 - \delta}{1 - \delta} \frac{1}{\eta_0}$. Suppose not, i.e. suppose $\pi > \frac{1 - \delta}{1 - \delta} \frac{1}{\eta_0} \bar{k}(\mu^0)$. But then,

$\delta S(\mu^0) + \delta (1 - \mu^0)(1 - \alpha)\pi = \frac{\delta}{1 - \delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\alpha \eta_0)\bar{k}(\mu^0) - c(\bar{k}(\mu^0)) + \delta (1 - \mu^0)(1 - \alpha)\pi$

$> \frac{\delta}{1 - \delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\alpha \eta_0)\bar{k}(\mu^0) - c(\bar{k}(\mu^0)) + \delta (1 - \mu^0)(1 - \alpha)\frac{1 - \beta}{1 - \delta} \eta_0 \bar{k}(\mu^0)$

$= \frac{\delta}{1 - \delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\eta_0)\bar{k}(\mu^0) - c(\bar{k}(\mu^0))$

$= \delta S(\mu^0)$

where the first inequality follows by Equation (A.13). This contradicts the case considered here. Therefore,

$\delta V_f(\bar{k}(\mu^0), \mu^0, \pi) - c(\bar{k}(\mu^0)) = \frac{\delta}{1 - \delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\eta_0)\bar{k}(\mu^0) - c(\bar{k}(\mu^0)) = \delta S(\mu^0)$

- For $k \in \left[0, \frac{1 - \delta}{1 - \beta} \frac{1}{\gamma(\mu^0)} \pi\right]$, we have:

$\delta V_f(\bar{k}(\mu^0), \mu^0, \pi) - c(\bar{k}(\mu^0)) = \delta S(\mu^0)$

$\geq \delta \pi$

$\geq \delta \pi - c(k)$

$= \delta V_f(k, \mu^0, \pi) - c(k)$

where the first inequality holds in the case considered here, and the second inequality holds by Assumption 1.1.

- For $k \in \left[\frac{1 - \delta}{1 - \beta} \frac{1}{\gamma(\mu^0)} \pi, \frac{1 - \delta}{1 - \beta} \frac{1}{\eta_0} \pi\right]$, we have:

$\delta V_f(\bar{k}(\mu^0), \mu^0, \pi) - c(\bar{k}(\mu^0)) = \delta S(\mu^0)$

$\geq \delta S(\mu^0) + \delta (1 - \mu^0)(1 - \alpha)\pi$

$= \delta \left(\frac{1}{1 - \delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\alpha \eta_0)\bar{k}(\mu^0) - c(\bar{k}(\mu^0))\right) + \delta (1 - \mu^0)(1 - \alpha)\pi$

$\geq \delta \left(\frac{1}{1 - \delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\alpha \eta_0)k - \frac{c(k)}{\delta}\right) + \delta (1 - \mu^0)(1 - \alpha)\pi$

$= \delta \left(\frac{1 - \delta}{1 - \beta} (1 - \beta)(\mu^0 + (1 - \mu^0)\alpha \eta_0)k + \delta (1 - \mu^0)(1 - \alpha)\pi - c(k)\right)$

$= \delta V_f(k, \mu^0, \pi) - c(k)$
where the first inequality holds in the case considered here, and the second inequality follows by Equation (A.13).

- For $k \geq \frac{1-\delta}{1-\beta} \eta_0 \pi$,

$$\delta V_f(\bar{k}(\mu^0), \mu^0, \pi) - c(\bar{k}(\mu^0)) = \delta \bar{S}(\mu^0)$$

$$= \delta \left( \frac{1}{1-\delta} (1-\beta)(\mu^0 + (1-\mu^0)\eta_0)\bar{k}(\mu^0) - \frac{c(\bar{k}(\mu^0))}{\delta} \right)$$

$$\geq \delta \left( \frac{1}{1-\delta} (1-\beta)(\mu^0 + (1-\mu^0)\eta_0)k - \frac{c(k)}{\delta} \right)$$

$$= \frac{\delta}{1-\delta} (1-\beta)(\mu^0 + (1-\mu^0)\eta_0)k - c(k)$$

$$= \delta V_f(k, \mu^0, \pi) - c(k)$$

where the inequality follows by Equation (A.11).

3. Suppose $V_f(\mu^0) = \delta \bar{S}(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi$, i.e. $\bar{S}(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi \geq \max\{\delta \pi, \delta \bar{S}(\mu^0)\}$.

I will demonstrate that $k^*(\mu^0, \pi) = \bar{k}(\mu^0)$, i.e. $\delta V_f(\bar{k}(\mu^0), \mu^0, \pi) - c(\bar{k}(\mu^0)) \geq \delta V_f(k, \mu^0, \pi) - c(k)$ for all $k \geq 0$.

My first claim is that in this case, $\delta V_f(\bar{k}(\mu^0), \mu^0, \pi) - c(\bar{k}(\mu^0)) = \frac{\delta}{1-\delta} (1-\beta)(\mu^0 + (1-\mu^0)(\alpha\eta_0)\bar{k}(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi - c(\bar{k}(\mu^0)))$. This is equivalent to showing that $\bar{k}(\mu^0) \in [\frac{1-\delta}{1-\beta} \gamma(\mu^0) \pi, \frac{1-\delta}{1-\beta} \pi]$.

To see $\bar{k}(\mu^0) \leq \frac{1-\delta}{1-\beta} \eta_0 \pi$, suppose not, i.e. suppose $\pi < \frac{1-\beta}{1-\delta} \eta_0 \bar{k}(\mu^0)$. Then,

$$\delta \bar{S}(\mu^0) = \frac{\delta}{1-\delta} (1-\beta)(\mu^0 + (1-\mu^0)\eta_0)\bar{k}(\mu^0) - c(\bar{k}(\mu^0))$$

$$> \frac{\delta}{1-\delta} (1-\beta)(\mu^0 + (1-\mu^0)\eta_0)\bar{k}(\mu^0) - c(\bar{k}(\mu^0))$$

$$> \frac{\delta}{1-\delta} (1-\beta)(\mu^0 + (1-\mu^0)\alpha\eta_0)\bar{k}(\mu^0) - c(\bar{k}(\mu^0)) + \delta(1-\mu^0)(1-\alpha)\pi$$

$$= \delta \bar{S}(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi$$

where the first inequality follows by Equation (A.11). This contradicts the case considered here.

To see $\bar{k}(\mu^0) \geq \frac{1-\delta}{1-\beta} \frac{1}{\gamma(\mu^0)} \pi$, suppose not, i.e. suppose $\pi > \frac{1-\beta}{1-\delta} \gamma(\mu^0) \bar{k}(\mu^0)$. Then,
\[
\delta \pi > \frac{\delta}{1-\delta} (1-\beta) k(\mu^0) \gamma(\mu^0) \\
= \frac{\delta}{1-\delta} (1-\beta) k(\mu^0) \frac{\mu^0(1-\eta_0) + \eta_0(1-\delta)}{\mu^0(1-\eta_0)\delta + (1-\delta)} \\
= \frac{\delta}{1-\delta} (1-\beta) k(\mu^0) \frac{\mu^0 + (1-\mu^0)\alpha \eta_0}{1-(1-\mu^0)(1-\alpha)} \\
= \frac{\delta}{1-\delta} (1-\beta) (\mu^0 + (1-\mu^0)\alpha \eta_0) k(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi \\
> \frac{\delta}{1-\delta} (1-\beta) (\mu^0 + (1-\mu^0)\alpha \eta_0) k(\mu^0) - c(k(\mu^0)) + \delta(1-\mu^0)(1-\alpha)\pi \\
= \delta \mathcal{S}(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi
\]

which contradicts the case considered here. Therefore,

\[
\delta V_f(k(\mu^0), \mu^0, \pi) - c(k(\mu^0)) = \frac{\delta}{1-\delta} (1-\beta) (\mu^0 + (1-\mu^0)\alpha \eta_0) k(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi - c(k(\mu^0)) \\
= \delta \mathcal{S}(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi
\]

- For \( k \in [0, \frac{1-\delta}{1-\beta} \eta_0, \frac{1-\delta}{1-\beta} \pi] \),

\[
\delta V_f(k(\mu^0), \mu^0, \pi) - c(k(\mu^0)) = \delta \mathcal{S}(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi \\
\geq \delta \pi \\
\geq \delta \pi - c(k) \\
= \delta V_f(k, \mu^0, \pi) - c(k)
\]

where the first inequality holds in the case considered here, and the second inequality holds by Assumption 1.1.

- For \( k \in \left[\frac{1-\delta}{1-\beta} \eta_0, \frac{1-\delta}{1-\beta} \pi, \frac{1-\delta}{1-\beta} \gamma(\mu^0) \pi\right] \),

\[
\delta V_f(k(\mu^0), \mu^0, \pi) - c(k(\mu^0)) = \delta \mathcal{S}(\mu^0) + \delta(1-\mu^0)(1-\alpha)\pi \\
= \frac{\delta}{1-\delta} (1-\beta) (\mu^0 + (1-\mu^0)\alpha \eta_0) k(\mu^0) - c(k(\mu^0)) + \delta(1-\mu^0)(1-\alpha)\pi \\
\geq \frac{\delta}{1-\delta} (1-\beta) (\mu^0 + (1-\mu^0)\alpha \eta_0) k - c(k) + \delta(1-\mu^0)(1-\alpha)\pi \\
= \delta V_f(k, \mu^0, \pi) - c(k)
\]

where the inequality follows by Equation (A.13).
• For \( k \geq \frac{1-\delta}{1-\beta \eta_0} \pi \),
\[
\delta V_f(\bar{k}(\mu^0), \mu^0, \pi) - c(\bar{k}(\mu^0)) = \delta S(\mu^0) + \delta(1 - \mu^0)(1 - \alpha)\pi \\
\geq \delta S(\mu^0) \\
= \delta \left( \frac{1}{1-\delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\eta_0)\bar{k}(\mu^0) - \frac{c(\bar{k}(\mu^0))}{\delta} \right) \\
\geq \delta \left( \frac{1}{1-\delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\eta_0)k - \frac{c(k)}{\delta} \right) \\
= \frac{\delta}{1-\delta} (1 - \beta)(\mu^0 + (1 - \mu^0)\eta_0)k - c(k) \\
= \delta V_f(k, \mu^0, \pi) - c(k)
\]
where the first inequality holds in the case considered here, and the second inequality follows by Equation (A.11).

\[\square\]

A.2.4 Role of Assumption 1.2

One takeaway from Lemma A.2 and A.3 is that: the range where the firm sets \( k(\mu^0) \) and quits following a low output appears only if \( \pi_1(\mu^0) < \pi_2(\mu^0) \). This section discusses the meaning of this condition and argues that it can be guaranteed by Assumption 1.2.

Lemma A.4. Given \( \mu^0 \in [0, 1] \), there exists an \( \eta_0(\mu^0) > 0 \) such that:
\[
\eta_0 \in (0, \eta_0(\mu^0)) \implies \pi_1(\mu^0) < \pi_2(\mu^0)
\]

Proof. Take some \( \mu^0 \in [0, 1] \). Investigation of (A.10) and (A.12) reveals that, when \( \eta_0 = 0 \), \( \bar{S}(\mu^0) = S(\mu^0) \). Therefore, \( \pi_1(\mu^0) = 0 \) for any \( \mu^0 > 0 \). Under Assumption 1.1, \( S(\mu^0) > 0 \) for any \( \mu^0 > 0 \) and thus \( \pi_2(\mu^0) > 0 \) for any \( \mu^0 \geq 0 \). It follows that \( \pi_1(\mu^0) < \pi_2(\mu^0) \) when \( \eta_0 = 0 \). By Berge’s maximum theorem, both \( \bar{S}(\mu^0) \) and \( S(\mu^0) \) are continuous in \( \eta_0 \). Consequently, both \( \pi_1(\mu^0) \) and \( \pi_2(\mu^0) \) are continuous in \( \eta_0 \), and one concludes that there exists some \( \eta_0^*(\mu^0) > 0 \) such that:
\[
\eta_0 \in (0, \eta_0^*(\mu^0)) \implies \pi_1(\mu^0) < \pi_2(\mu^0)
\]

\[\square\]

It is possible to obtain a tighter characterization by imposing more structure on the cost of capital. For instance, consider the cost function in Equation (1.1). In this case, it is trivial to verify that:
\[
\eta_0 < \frac{\mu^0}{1 - \mu^0} \left( \frac{1 - (\mu^0 + (1 - \mu^0)\frac{1-\delta}{1-\eta_0 \delta})^{\frac{1}{1-\delta}}}{\frac{1-\delta}{1-\eta_0 \delta}} \right) \iff \pi_1(\mu^0) < \pi_2(\mu^0)
\]

102
This expression carries the same flavor as in Lemma A.4, but is a tighter characterization because it is a necessary and sufficient condition.

A.2.5 Proof of Proposition 1.1

Proof. Take \( \mu^0 \in [0, 1] \), and let \( \eta_0^* < \eta^*(\mu^0) \). Under Assumption 1.2, Lemma A.4 applies and \( \pi_1(\mu^0) < \pi_2(\mu^0) \). In this case, Lemma A.2 and A.3 yield the firm’s optimal capital choice given in Proposition 1.1. Firm’s optimal continuation decision in Proposition 1.1 follows from Proposition A.1. \( \square \)

A.3 Proofs in Section 1.3.2

Given \( \mu^0 := \mu(h_0) \) with \( \pi_1(\mu^0) < \pi_2(\mu^0) \) and \( (\pi, v) \), payoffs from participation for both types can be derived in a manner analogous to the firm’s continuation payoffs. For \( \theta \in \Theta \), let \( V_w(\theta, \mu^0, \pi) \) denote the payoff from participation for a worker with type \( \theta \), which will be accrued to the worker in \( t = 1 \) if she participates. Proposition 1.1 yields:

\[
V_w(1, \mu^0, \pi) = \begin{cases} \frac{1}{1-\delta} \beta \tilde{k}(\mu^0), & \text{if } \pi < \pi_1(\mu^0) \\ \omega \frac{1}{1-\delta} \beta \tilde{k}(\mu^0) + (1-\omega) \frac{1}{1-\delta} \beta \tilde{k}(\mu^0), & \text{if } \pi = \pi_1(\mu^0), \omega \in [0, 1] \\ \frac{1}{1-\delta} \beta \tilde{k}(\mu^0), & \text{if } \pi \in (\pi_1(\mu^0), \pi_2(\mu^0)) \\ \omega \frac{1}{1-\delta} \beta \tilde{k}(\mu^0) + (1-\omega)v, & \text{if } \pi = \pi_2(\mu^0), \omega \in [0, 1] \\ v, & \text{if } \pi > \pi_2(\mu^0) \end{cases}
\]  

(A.26)

and

\[
V_w(0, \mu^0, \pi) = \begin{cases} \frac{1}{1-\delta} \beta \eta_0 \tilde{k}(\mu^0), & \text{if } \pi < \pi_1(\mu^0) \\ \omega \frac{1}{1-\delta} \beta \eta_0 \tilde{k}(\mu^0) + (1-\omega) \left( \frac{1}{1-\delta} \beta \eta_0 \frac{1-\delta}{1-\eta_0} k(\mu^0) + \frac{\delta-\eta_0 \delta}{1-\eta_0} v \right), & \text{if } \pi = \pi_1(\mu^0), \omega \in [0, 1] \\ \frac{1}{1-\delta} \beta \eta_0 \frac{1-\delta}{1-\eta_0} k(\mu^0) + \frac{\delta-\eta_0 \delta}{1-\eta_0} v, & \text{if } \pi \in (\pi_1(\mu^0), \pi_2(\mu^0)) \\ \omega \left( \frac{1}{1-\delta} \beta \eta_0 \frac{1-\delta}{1-\eta_0} k(\mu^0) + \frac{\delta-\eta_0 \delta}{1-\eta_0} v \right) + (1-\omega)v, & \text{if } \pi = \pi_2(\mu^0), \omega \in [0, 1] \\ v, & \text{if } \pi > \pi_2(\mu^0) \end{cases}
\]  

(A.27)

Optimality of worker’s strategy requires:

\[
\sigma_w^*(j|\theta) \in \arg \max_{\sigma \in [0,1]} \left( 1 - \sigma \right) \cdot v + \sigma \cdot \delta V_w(\theta, \mu^0, \pi) \quad \forall \theta \in \{0, 1\}
\]  

(A.28)

3The restriction that the firm continues when indifferent between \( c \) and \( q \) comes into play in this statement by simplifying this expression.

4Recall that, by Proposition 1.1, the firm is indifferent between at most two capital levels. These statements take into account that in the case of indifference, the firm may assign any probability to the two capital levels in the support. Consequently, the worker’s payoff from participation takes the form of an upper hemicontinuous correspondence.

103
Proof. (of of Lemma 1.1). Take a PBE (\(\{\sigma^*_w, \sigma^*_j, \{\mu(h^t)\}_{h^t} \}_{h^t} \)), and suppose \(\sigma^*_w(j|\theta = 0) > 0\). By Equation (A.28), for this to be an optimal strategy, one must have:

\[
\delta V_w(0, \mu^0, \pi) \geq v
\]  

(A.29)

I will first argue that \(\pi \leq \pi_2(\mu^0)\). Suppose, towards a contradiction, that \(\pi > \pi_2(\mu^0)\). By Equation (A.27), \(V_w(0, \mu^0, \pi) = v > 0\). But since \(\delta \in (0, 1)\), this contradicts Equation (A.29). I conclude that \(\pi \leq \pi_2(\mu^0)\).

Now, consider the remaining four cases.

1. If \(\pi < \pi_1(\mu^0)\), by Equation (A.26) and (A.27):

\[
V_w(\theta, \mu^0, \pi) = \frac{1}{1 - \delta} \beta \eta_0 \bar{k}(\mu^0) \quad \theta \in \Theta
\]

By Equation (A.29), this implies:

\[
\frac{\delta}{1 - \delta} \beta \eta_0 \bar{k}(\mu^0) \geq v
\]

Since \(v > 0\), one must have \(\bar{k}(\mu^0) > 0\). Moreover, because \(\eta_0 < \eta_1 = 1\),

\[
\frac{1}{1 - \delta} \beta \bar{k}(\mu^0) > \frac{1}{1 - \delta} \beta \eta_0 \bar{k}(\mu^0)
\]

And therefore:

\[
V_w(1, \mu^0, \pi) > V_w(0, \mu^0, \pi)
\]

Combining this with Equation (A.29) gives \(\delta V_w(1, \mu^0, \pi) > v\). Since \(\sigma^*_w(j|\theta = 1)\) solves the optimization problem in Equation (A.28), \(\sigma^*_w(j|\theta = 1) = 1\).

2. If \(\pi = \pi_1(\mu^0)\), by Equation (A.26):

\[
V_w(1, \mu^0, \pi) = \frac{1}{1 - \delta} \beta(\omega \bar{k}(\mu^0) + (1 - \omega)\bar{k}(\mu^0))
\]

and by Equation (A.27):

\[
V_w(0, \mu^0, \pi) = \omega \frac{1}{1 - \delta} \beta \eta_0 \bar{k}(\mu^0) + (1 - \omega) \left( \frac{1}{1 - \delta} \beta \eta_0 \frac{1 - \delta}{1 - \eta_0 \delta} \bar{k}(\mu^0) + \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} v \right)
\]

By Equation (A.29), this implies:

\[
\frac{\delta}{1 - \delta} \beta \eta_0 (\omega \bar{k}(\mu^0) + (1 - \omega)\bar{k}(\mu^0)) \geq v
\]
Since $v > 0$, and because $\eta_0 < 1$,
\[
\frac{1}{1 - \delta} \beta (\omega k(\mu^0) + (1 - \omega) k(\mu^0)) > v
\]
And therefore:
\[
\delta V_w(1, \mu^0, \pi) > v
\]
Since $\sigma^*_w(j|\theta = 1)$ solves the optimization problem in Equation (A.28), $\sigma^*_w(j|\theta = 1) = 1$.

3. If $\pi \in (\pi_1(\mu^0), \pi_2(\mu^0))$, by Equation (A.26):
\[
V_w(1, \mu^0, \pi) = \frac{\delta}{1 - \delta} \beta k(\mu^0)
\]
and by Equation (A.27):
\[
V_w(0, \mu^0, \pi) = \frac{1}{1 - \delta} \beta \eta_0 \frac{1 - \delta}{1 - \eta_0 \delta} k(\mu^0) + \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} v
\]
By Equation (A.29), this implies:
\[
\frac{\delta}{1 - \delta} \beta \eta_0 k(\mu^0) \geq v \left(1 - \delta \frac{1 - \delta}{1 - \eta_0 \delta}\right)
\]
Which is rearranged to:
\[
\frac{\delta}{1 - \delta} \beta \eta_0 k(\mu^0) \geq v(1 + \delta(1 - \eta_0))
\]
Therefore,
\[
\frac{\delta}{1 - \delta} \beta \eta_0 k(\mu^0) > v
\]
Since $v > 0$, one must have $k(\mu^0) > 0$. Moreover, because $\eta_0 < 1$,
\[
\frac{\delta}{1 - \delta} \beta k(\mu^0) > \frac{\delta}{1 - \delta} \beta \eta_0 k(\mu^0)
\]
And therefore:
\[
\delta V_w(1, \mu^0, \pi) > \frac{\delta}{1 - \delta} \beta \eta_0 k(\mu^0) > v
\]
Since $\sigma^*_w(j|\theta = 1)$ solves the optimization problem in Equation (A.28), $\sigma^*_w(j|\theta = 1) = 1$.

4. If $\pi = \pi_2(\mu^0)$, by Equation (A.26):
\[
V_w(1, \mu^0, \pi) = \omega \frac{\delta}{1 - \delta} \beta k(\mu^0) + (1 - \omega)v
\]
and by Equation (A.27):

\[ V_w(0, \mu^0, \pi) = \omega \left( \frac{1}{1 - \delta} \beta \eta_0 \frac{1 - \delta}{1 - \eta_0 \delta} k(\mu^0) + \frac{\delta - \eta_0 \delta}{1 - \eta_0 \delta} v \right) + (1 - \omega) v \]

By Equation (A.29), this implies:

\[ \frac{1}{1 - \delta} \beta \eta_0 k(\mu^0) > v \]

and because \( \eta_0 < 1 \),

\[ V_w(1, \mu^0, \pi) > V_w(0, \mu^0, \pi) \]

Combining this with Equation (A.29):

\[ \delta V_w(1, \mu^0, \pi) > v \]

Since \( \sigma^*_w(j|\theta = 1) \) solves the optimization problem in Equation (A.28), \( \sigma^*_w(j|\theta = 1) = 1 \). Since the four cases considered above are exhaustive, the proof follows.

Full characterization of the participation decision requires one to define two objects. These are the critical values of outside option for the worker which determine what type of a PBE exists. Let:

\[ v_2(\mu^0, \pi) := \begin{cases} \frac{\delta}{1 - \delta} \beta k(\mu^0), & \text{if } \pi \leq \pi_1(\mu^0) \\ \frac{\delta}{1 - \delta} \beta k(\mu^0), & \text{if } \pi > \pi_1(\mu^0) \end{cases} \]

and

\[ v_1(\mu^0, \pi) := \begin{cases} \frac{\delta}{1 - \delta} \beta \eta_0 k(\mu^0), & \text{if } \pi \leq \pi_1(\mu^0) \\ \frac{\delta}{1 - \delta} \beta \eta_0 \frac{1}{1 + \delta(1 - \eta_0)} k(\mu^0), & \text{if } \pi > \pi_1(\mu^0) \end{cases} \]

Proof. (of Proposition 1.2). Assume \( \pi_1(\mu^0) < \pi_2(\mu^0) \). Recall that payoffs upon participation is given by Equations (A.26) and (A.27).

The proof analyzes different cases separately.

1. When \( \pi \leq \pi_1(\mu^0) \), by Proposition 1.1, the optimal strategy of the firm involves choosing \( k(\mu^0) \) and never quitting regardless of the output. By the optimization problem in Equation (A.28):

\[ \sigma^*_f(j|\theta) \begin{cases} = 1, & \text{if } v \leq \frac{\delta}{1 - \delta} \beta \eta_0 k(\mu^0) \\ \in [0, 1], & \text{if } v = \frac{\delta}{1 - \delta} \beta \eta_0 k(\mu^0) \forall \theta \in \Theta \\ = 0, & \text{otherwise.} \end{cases} \]
2. When \( \pi \in [\pi_1(\mu^0), \pi_2(\mu^0)] \), by Proposition 1.1, the optimal strategy of the firm involves choosing \( k(\mu^0) \) and quitting following a low output. By the optimization problem in Equation (A.28),

\[
\sigma^*_j(\theta = 1) = \begin{cases} 
1, & \text{if } v \leq \frac{\delta}{1-\delta} \beta k(\mu^0) \\
1, & \text{if } v = \frac{\delta}{1-\delta} \beta k(\mu^0) \\
0, & \text{otherwise.}
\end{cases}
\]

and

\[
\sigma^*_j(\theta = 0) = \begin{cases} 
1, & \text{if } v \leq \frac{\delta}{1-\delta} \beta k(\mu^0) \eta_0 \frac{1}{1+\delta(1-\eta_0)} \\
1, & \text{if } v = \frac{\delta}{1-\delta} \beta k(\mu^0) \frac{1}{1+\delta(1-\eta_0)} \\
0, & \text{otherwise.}
\end{cases}
\]

3. When \( \pi > \pi_2(\mu^0) \), by Proposition 1.1, the firm chooses 0 as the capital level and immediately quits. By the optimization problem in Equation (A.28), \( \sigma^*_j(\theta) = 0 \) for \( \theta \in \Theta \).

Combining the three cases yields the desired result on the pooling and separating equilibria, as well as the semi-separating equilibria where the firm uses a pure strategy. The final possibility not considered yet is the possibility of firm mixing between different levels of capital choices. By Proposition 1.1, the firm has at most two capital choices in its support, and this indifference condition is satisfied only for particular values of its outside option. 

- When \( \pi = \pi_1(\mu^0) \), by Proposition 1.1, the firm is indifferent between \( \{k(\mu^0), 0\} \). If the firm chooses \( k(\mu^0) \), it never quits in the following periods, regardless of the output. If the firm chooses \( 0 \), it continues as long as high output is observed and immediately quits following a low output.

Suppose the firm chooses \( k(\mu^0) \) with probability \( \omega \in (0, 1) \), and chooses \( 0 \) with probability \( 1 - \omega \). By Proposition 1.1, this semi-separating equilibrium exists if and only if \( \pi = \pi_1(\mu^0) \). By Equation A.27, semi-separating equilibrium requires:

\[
\omega \frac{\delta}{1-\delta} \beta \eta_0 k(\mu^0) + (1 - \omega) \left( \frac{\delta}{1-\delta} \beta \eta_0 \frac{1}{1-\eta_0} k(\mu^0) + \delta \frac{\delta - \eta_0 \delta}{1-\eta_0} v \right) = v
\]

For this equality to be satisfied for some \( \omega \in (0, 1) \), one needs:

\[
v \in \left( \frac{\delta}{1-\delta} \beta \eta_0 \frac{1}{1+\delta(1-\eta_0)} k(\mu^0), \frac{\delta}{1-\delta} \beta \eta_0 k(\mu^0) \right)
\]

- When \( \pi = \pi_2(\mu^0) \), by Proposition 1.1, the firm is indifferent between \( \{k(\mu^0), 0\} \). If the firm chooses \( k(\mu^0) \), it continues as long as high output is observed and immediately quits following a low output. If the firm chooses 0, it immediately quits.

Suppose the firm chooses \( k(\mu^0) \) with probability \( \omega \in (0, 1) \), and chooses 0 with probability \( 1 - \omega \). By Proposition 1.1, this semi-separating equilibrium exists if and only if
\( \pi = \pi_2(\mu^0). \) By Equation A.27, semi-separating equilibrium requires:

\[
\omega \left( \frac{\delta}{1-\delta} \beta \eta_0 \frac{1-\delta}{1-\eta_0 \delta} k(\mu^0) + \delta \frac{\delta - \eta_0 \delta}{1-\eta_0 \delta} v \right) + (1-\omega)\delta v = v
\]

For this equality to be satisfied for some \( \omega \in (0, 1), \) one needs:

\[
v \in (0, \frac{\delta}{1-\delta} \beta \eta_0 k(\mu^0) \frac{1}{1+\delta(1-\eta_0)})
\]

The result follows. \( \square \)

### A.4 Proof of Theorem 1.1

Let

\[
\eta_0^* < \min_{\mu^0 \in [\phi, 1]} \eta^*_0(\mu^0)
\]

(A.30)

By Lemma 1.1, in any PBE with positive participation, \( \mu^0 \in [\phi, 1]. \) Therefore, under Assumption 1.2, Proposition 1.1 and 1.2 for \( \mu^0 \in [\phi, 1] \) apply.

**Proof.** In a separating equilibrium, the participation result follows from Proposition 1.2 for \( \mu^0 = 1, \) and characterization of the capital choice and quitting behavior follow from Proposition 1.1 for \( \mu^0 = 1. \)

In a pooling equilibrium, once again, the participation result follows from Proposition 1.2 for \( \mu^0 = \phi, \) and characterization of the capital choice and quitting behavior follow from Proposition 1.1 for \( \mu^0 = \phi. \)

For the rest of the proof, I construct the semi-separating equilibria in the remaining regions. To make the illustration easier, below is an adapted version of Figure 1-3.
This is the same figure as Figure 1-3, except that Region D is divided into four subregions.

- The boundary between regions E and F is the curve:
  \[
  \{(\pi, v) : \pi = \pi_1(\mu^0), v = \frac{\delta}{1 - \delta} \beta \eta_0 \bar{k}(\mu^0), \text{ for } \mu^0 \in (\phi, 1)\}
  \]

- The boundary between regions F and G is the curve:
  \[
  \{(\pi, v) : \pi = \pi_1(\mu^0), v = \frac{\delta}{1 - \delta} \beta \eta_0 \bar{k}(\mu^0) \frac{1}{1 + \delta(1 - \eta_0)}, \text{ for } \mu^0 \in (\phi, 1)\}
  \]

- The boundary between regions G and H is the curve:
  \[
  \{(\pi, v) : \pi = \pi_2(\mu^0), v = \frac{\delta}{1 - \delta} \beta \eta_0 \bar{k}(\mu^0) \frac{1}{1 + \delta(1 - \eta_0)}, \text{ for } \mu^0 \in (\phi, 1)\}
  \]

A different type of semi-separating equilibria exists in each of these subregions.

- Take \(\mu^0 \in (\phi, 1)\), and consider a semi-separating equilibrium without layoffs. In such an equilibrium, the firm chooses a capital level \(\bar{k}(\mu^0)\) and never quits, regardless of the output. By Proposition 1.1, this semi-separating equilibrium exists if and only if \(\pi \leq \pi_1(\mu^0)\). By Proposition 1.2, this semi-separating equilibrium exists if and only
if $v = \frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0)$. Combining, the set of $(\pi, v)$ pairs where such a semi-separating equilibria exists corresponds to:

$$\{(\pi, v) : v = \frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0), \pi \leq \pi_1(\mu^0) \text{ for } \mu^0 \in (\phi, 1)\}$$

This region exactly corresponds to the region marked E in Figure A-6 such a semi-separating equilibrium exists in, and only in, Region E.

- Take $\mu^0 \in (\phi, 1)$, and consider a semi-separating equilibrium with layoffs. In such an equilibrium, the firm chooses a capital level $k(\mu^0)$ and quits immediately following a low output. By Proposition 1.1, this semi-separating equilibrium exists if and only if $\pi \in [\pi_1(\mu^0), \pi_2(\mu^0)]$. By Proposition 1.2, this semi-separating equilibrium exists if and only if $v = \frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0) \frac{1}{1+\delta(1-\eta_0)}$. Combining, the set of $(\pi, v)$ pairs where such a semi-separating equilibria exists corresponds to:

$$\{(\pi, v) : v = \frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0) \frac{1}{1+\delta(1-\eta_0)}, \pi \in [\pi_1(\mu^0), \pi_2(\mu^0)] \text{ for } \mu^0 \in (\phi, 1)\}$$

This region exactly corresponds to the region marked G in Figure A-6 such a semi-separating equilibrium exists in, and only in, Region G.

- Take $\mu^0 \in (\phi, 1)$, and consider a semi-separating equilibrium where the firm mixes in $\{k(\mu^0), k(\mu^0)\}$. By Proposition 1.1, this semi-separating equilibrium exists if and only if $\pi = \pi_1(\mu^0)$. By Proposition 1.2, this semi-separating equilibrium exists if and only if

$$v \in \left[\frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0) \frac{1}{1+\delta(1-\eta_0)}, \frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0)\right]$$

Combining, the set of $(\pi, v)$ pairs where such a semi-separating equilibria exists corresponds to:

$$\{(\pi, v) : \pi = \pi_1(\mu^0), v \in \left[\frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0) \frac{1}{1+\delta(1-\eta_0)}, \frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0)\right], \text{ for } \mu^0 \in (\phi, 1)\}$$

This region exactly corresponds to the region marked F in Figure A-6 such a semi-separating equilibrium exists in, and only in, Region F.

- Take $\mu^0 \in (\phi, 1)$, and consider a semi-separating equilibrium where the firm mixes in $\{k(\mu^0), 0\}$. By Proposition 1.1, this semi-separating equilibrium exists if and only if $\pi = \pi_2(\mu^0)$. By Proposition 1.2, this semi-separating equilibrium exists if and only if

$$v \leq \frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0) \frac{1}{1+\delta(1-\eta_0)}$$

Combining, the set of $(\pi, v)$ pairs where such a semi-separating equilibria exists corresponds to:

$$\{(\pi, v) : \pi = \pi_2(\mu^0), v \in (0, \frac{\delta}{1-\delta} \beta \eta_0 \bar{k}(\mu^0) \frac{1}{1+\delta(1-\eta_0)}], \text{ for } \mu^0 \in (\phi, 1)\}$$
This region exactly corresponds to the region marked H in Figure A-6 such a semi-
separating equilibrium exists in, and only in, Region \( H \).

Combined, these regions comprise Region \( D \) in Figure 1-3, and the result follows. \( \square \)

### A.5 Proof of Proposition 1.3

Let

\[
\eta^*_0 < \min \{ \min_{\mu^0} \eta^*_0(\mu^0), \frac{\phi}{1 - \phi} \left( \frac{k(1)}{k(\phi)} - 1 \right) \} \tag{A.31}
\]

The \( \eta^*_0 < \min_{\mu^0} \eta^*_0(\mu^0) \) restriction is already imposed in the proof of Theorem 1.1. The \( \eta^*_0 < \frac{\phi}{1 - \phi} \left( \frac{k(1)}{k(\phi)} - 1 \right) \) restriction is a very mild one, and is not binding if the capital choice is sufficiently responsive. Indeed, if the cost function is as given in Equation (1.1), this restriction is:

\[
\eta^*_0 < \frac{\phi}{1 - \phi}
\]

which is implied by \( \eta^*_0 < \eta^*_0(\phi) \) (See Section A.2.4).

**Proof.** Consider the case when \( \pi \leq \pi_1(\phi) \). In this case,

- When \( v \leq v(0) = \frac{\delta}{1 - \delta} \beta \eta_0 k(\phi) \), the PBE is a pooling equilibrium without layoffs and

  \[
  Y(\pi, v) = \frac{\delta}{1 - \delta}(\phi + (1 - \phi)\eta_0)k(\phi)
  \]

- When \( \pi \leq v(v(0), \bar{v}(0)) = (\frac{\delta}{1 - \delta} \beta \eta_0 k(\phi), \frac{\delta}{1 - \delta} \beta \eta_0 k(1)) \), the PBE is a semi-separating equilibrium without layoffs. In this case, the low-type worker joins the firm with probability \( \sigma \in (0, 1) \) and following equalities apply:

  \[
  v = \frac{\delta}{1 - \delta} \beta \eta_0 k(\mu^0)
  \]

  \[
  \mu^0 = \frac{\phi}{\sigma(1 - \phi)k(1)}
  \]

  \[
  Y(\pi, v) = \frac{\delta}{1 - \delta}(\phi + (1 - \phi)\sigma\eta_0)k(\mu^0)
  \]

- When \( v \in [\bar{v}(0), v^*] = (\frac{\delta}{1 - \delta} \beta \eta_0 k(1), \frac{\delta}{1 - \delta} \beta k(1)) \), the PBE is a separating equilibrium and

  \[
  Y(\pi, v) = \frac{\delta}{1 - \delta} \phi k(1)
  \]

111
Equation (A.31) and Assumption 1.2 ensures that the output in a separating equilibrium is larger than the output in a pooling equilibrium with layoffs. By continuity and monotonicity, \( Y(\pi, v) \) is increasing in the region in between and the result follows.

Now, consider the case when \( v \leq \bar{v}(\pi_2) \). In this case,

- When \( \pi \leq \pi_1(\phi) \), there is a pooling equilibrium without layoffs and

  \[
  Y(\pi, v) = \frac{\delta}{1 - \delta}(\phi + (1 - \phi)\eta_0)k(\phi)
  \]

- When \( \pi \in [\pi_1(\phi), \pi_2(\phi)] \), there is a pooling equilibrium with layoffs and

  \[
  Y(\pi, v) = \frac{\delta}{1 - \delta}(\phi + (1 - \phi)\eta_0)k(\phi)
  \]

- When \( \pi \in (\pi_2(\phi), \pi_2(1)) \), there is a semi-separating equilibrium where the low-type worker joins the firm with probability \( \sigma \in (0, 1) \). The firm mixes in \( \{k(\mu^0), 0\} \), where the probability assigned to \( k(\mu^0) \) is \( \omega \in (0, 1) \). Following equalities apply:

  \[
  \pi = \pi_1(\mu^0)
  \]

  \[
  v = \omega \left( \frac{\delta}{1 - \delta}\eta_0 \frac{1 - \delta}{1 - \eta_0\delta}k(\mu^0) + \frac{\delta - \eta_0\delta}{1 - \eta_0\delta}\delta v \right) + (1 - \omega)\delta v
  \]

  \[
  \mu^0 = \frac{\phi}{\phi + (1 - \phi)\sigma}
  \]

  \[
  Y(\pi, v) = \frac{\delta}{1 - \delta}\omega(\phi + (1 - \phi)\sigma)k(\mu^0)
  \]

Investigation of equations reveals that in this case, \( \mu^0 \) is increasing in \( \pi \) whereas \( \omega \) and \( \sigma \) is decreasing in \( \pi \). By the second equation, \( \omega k(\mu^0) \) is decreasing in \( \pi \), and therefore \( Y(\pi, v) \) is decreasing in \( \pi \).

A.6 Proof of Lemma 1.2

Proof. The first claim is that it one must have \( u \geq \nu \) in the steady state. Suppose not, i.e. suppose \( u < \nu \). Then, \( \tau > 1 \) and one must have \( q(\tau) = \frac{u}{\nu} \) and \( \tau q(\tau) = 1 \). Equations (1.8) and (1.9) then become:

\[
L = u \cdot \rho^*
\]

\[
1 = u \cdot \rho^* + (\nu - u)s
\]

Since \( L \geq 1 \), these equations cannot be satisfied simultaneously for \( \nu - u > 0 \), a contradiction. Therefore, \( u \geq \nu \) in the steady state and \( \tau \leq 1 \). Consequently, \( q(\tau) = 1 \) and \( \tau q(\tau) = \frac{\nu}{u} \).
Equations (1.8) and (1.9) imply:

\[ L = \nu \cdot \rho^* + (u - \nu)s \]
\[ 1 = \nu \cdot \rho^* \]

Solving these equations yields the desired result and verifies that \( u \geq \nu \). \( \square \)

## A.7 Proofs in Section 1.5

By Equation (1.10):

\[ J^U = \tau q(\tau) \Sigma_w(\pi, v) + (1 - \tau q(\tau)) \delta(1 - s) J^U \]

where \( \tau = \frac{u}{v} \) is the market tightness in the steady-state. Denoting \( \tilde{\delta} = \tilde{\delta}(1 - s) \) for the effective discount factor, one has:

\[ J^U = \frac{\tau q(\tau)}{1 - \tilde{\delta} + \tilde{\delta} q(\tau)} \Sigma_w(\pi, v) \]  

(A.32)

An identical argument applies to the firm as well, and thus the net present value of being a vacancy is:

\[ J^V = \frac{q(\tau)}{1 - \tilde{\delta} + \tilde{\delta} q(\tau)} \Sigma_f(\pi, v) \]  

(A.33)

Equation (1.12) for the unemployed worker, therefore, is:

\[ \nu = \delta^K J^U = \delta^K \frac{\tau q(\tau)}{1 - \tilde{\delta} + \tilde{\delta} q(\tau)} \Sigma_w(\pi, v) \]

Similarly, for the firm:

\[ \pi = \delta^K J^U = \delta^K \frac{q(\tau)}{1 - \tilde{\delta} + \tilde{\delta} q(\tau)} \Sigma_f(\pi, v) \]

One can indeed \( \tau \) from these equations by appealing to Lemma 1.2. If the PBE under \((\pi, v)\) invokes \( \rho^* \), the following equations must hold:

\[ \nu = \delta^K \frac{s}{(L - 1)(1 - \delta) \rho^* + s} \Sigma_w(\pi, v) \]  

(A.34)

\[ \pi = \delta^K \Sigma_f(\pi, v) \]  

(A.35)

Before providing the formal proof of Theorem 1.2, I will begin by outlining the general idea. Below is a demonstration of the critical values of the search frictions for the markets with intermediate thickness.
This figure is a concise illustration of how different values of $K$ map into different market opportunities and equilibrium behavior for the firm and the worker. When $K \approx \infty$, both $v \approx 0$ and $\pi \approx 0$. As $K$ decreases, $\pi$ improves and eventually reaches a value that induces the firm to quit when the worker’s type is revealed to be low ($K_1$). For $K < K_1$, the search market is sufficiently attractive for the firm so that the firm chooses a capital level $k(\phi)$ and quits following a low output if the workers pool. Similarly, as $K$ decreases, $v$ improves as well. Eventually, it reaches a point where the workers no longer pool ($K_2(L)$). For sufficiently low values of $K$ ($K \leq K_3(L)$), the search market is sufficiently attractive for the workers such that they separate in joining the firm: a low-type worker strictly prefers to keep searching for a better match.

It should also be noted that $K_1$ does not depend on $L$, because by Lemma 1.2, the flow rate of matches for a vacancy is independent of $L$. Moreover, it can be shown that both $K_2(L)$ and $K_3(L)$ is decreasing in $L$: when the market is thicker, a lower level of search frictions is required to separate the workers. The $L < L$ condition ensures that $K_3(L) \geq 1$, so that for sufficiently low $K$ separation is possible. The $L > L$ condition ensures that $K_1 > K_2(L)$, so that a pooling equilibrium with layoff exists.

**Proof. (of Theorem 1.2.)** The proof follows the roadmap discussed at the end of Section 1.4.3. I begin by considering each of the three pure strategy equilibria separately and deriving the conditions for their existence.

**Pooling equilibrium without layoffs.** This is the PBE which exists when

$$v \leq \frac{\delta}{1 - \delta} \beta \eta_0 \bar{k}(\phi)$$

$$\pi \leq \pi_1(\phi)$$

The behavior in this equilibrium is such that:

$$\sigma_w^*(j|\theta = 0) = 1$$

$$\zeta^* = 1$$
In this equilibrium, the ex ante payoffs conditional on being matched are:

\[
\Sigma_w(v, \pi) = \frac{\delta}{1 - \delta} \beta \phi + (1 - \phi) \eta \bar{k}(\phi)
\]

\[
\Sigma_f(v, \pi) = \delta \bar{S}(\phi)
\]

**Pooling equilibrium with layoffs.** This is the PBE which exists when

\[
v \leq \frac{\delta}{1 - \delta} \beta \eta k(\phi) - \frac{1}{1 + \delta(1 - \eta)}
\]

\[
\pi \in [\pi_1(\phi), \pi_2(\phi)]
\]

In this equilibrium, the ex ante payoffs conditional on being matched are:

\[
\Sigma_w(v, \pi) = \frac{\delta}{1 - \delta} \beta \phi + (1 - \phi) \eta \frac{1 - \delta}{1 - \eta \delta} k(\phi) + \delta(1 - \phi) \frac{\delta - \eta \delta}{1 - \eta \delta} v
\]

\[
\Sigma_f(v, \pi) = \delta \bar{S}(\phi) + \delta(1 - \phi) \frac{\delta - \eta \delta}{1 - \eta \delta} \pi
\]

**Separating equilibrium.** This is the PBE which exists when

\[
v \in \left[ \frac{\delta}{1 - \delta} \beta \eta k(1), \frac{\delta}{1 - \delta} \beta k(1) \right]
\]

\[
\pi \leq \bar{S}(1)
\]

or

\[
v \in \left[ \frac{\delta}{1 - \delta} \beta \eta \frac{1}{1 + \delta(1 - \eta)} k(1), \frac{\delta}{1 - \delta} \beta \eta k(1) \right]
\]

\[
\pi \in [\pi_1(1), \bar{S}(1)]
\]

The behavior in this equilibrium is such that:

\[
\sigma^*_w(j|\theta = 0) = 0
\]

\[
\zeta^* \in [0, 1]
\]

In this equilibrium, the ex ante payoffs conditional on being matched are:

\[
\Sigma_w(v, \pi) = \phi \frac{\delta}{1 - \delta} \beta \bar{k}(1) + (1 - \phi)v
\]

\[
\Sigma_f(v, \pi) = \phi \delta \bar{S}(1) + (1 - \phi)\pi
\]

Using Equations (1.7), (A.34) and (A.35), the conditions for existence of each of these equilibria are as follows.
1. A pooling equilibrium without layoffs exists when:

\[ \delta^K \leq \frac{(L - 1)(1 - \delta) + s}{s} \frac{1}{\phi \eta_0 + (1 - \phi)} \]  
(A.36)

\[ \delta^K \leq \frac{\pi_1(\phi)}{\delta S(\phi)} \]  
(A.37)

2. A pooling equilibrium with layoffs exists when:

\[ \delta^K \leq \frac{(L - 1)(1 - \delta) + s}{s} \frac{1}{\phi \eta_0 (1 + \delta(1 - \eta_0)) + (1 - \phi)} \]  
(A.38)

\[ \delta^K \geq \frac{\pi_1(\phi)}{\delta S(\phi)} \]  
(A.39)

3. A separating equilibrium exists when

\[ \delta^K \geq \frac{(L - 1)(1 - \delta) + s}{s} \frac{1}{\phi \eta_0 (1 + \delta(1 - \eta_0)) + (1 - \phi)} \]  
(A.40)

or

\[ \delta^K \geq \frac{\pi_1(1)}{\phi \delta S(1) + (1 - \phi) \pi_1(1)} \]  
(A.41)

Let \( K_1 \) be the smallest value of \( K \) which satisfies Equation (A.37):

\[ \delta^{K_1} \leq \frac{\pi_1(\phi)}{\delta S(\phi)} \]

\[ \delta^{K_1-1} \geq \frac{\pi_1(\phi)}{\delta S(\phi)} \]

When \( \eta_0 < \eta_0^* \), \( \pi_1(\phi) < \delta S(\phi) \) so that \( K_1 \geq 1 \). Also, note that \( K_1 \) is independent of the value of \( L \).

Let \( K_2(L) \) be the smallest value of \( K \) which satisfies Equation (A.38):

\[ \delta^{K_2(L)} \leq \frac{(L - 1)(1 - \delta) + s}{s} \frac{1}{\phi \eta_0 (1 + \delta(1 - \eta_0)) + (1 - \phi)} \]

\[ \delta^{K_2(L)-1} \geq \frac{(L - 1)(1 - \delta) + s}{s} \frac{1}{\phi \eta_0 (1 + \delta(1 - \eta_0)) + (1 - \phi)} \]
Let $K^*_3(L)$ be the largest value of $K$ which satisfies Equation (A.40):

$$
\delta_{K^*_3(L)} \geq \left( L - 1 \right) \left( 1 - \delta \phi \right) + s \frac{1}{1 - \phi} \frac{1}{\eta_0 + (1 - \phi)}
$$

$$
\delta_{K^*_3(L)+1} \leq \left( L - 1 \right) \left( 1 - \delta \phi \right) + s \frac{1}{1 - \phi} \frac{1}{\eta_0 + (1 - \phi)}
$$

Let $K^*_3$ the largest value of $K$ which satisfies Equation (A.42):

$$
\delta_{K^*_3} \geq \frac{\pi_1(1)}{\phi \delta S(1) + (1 - \phi) \pi_1(1)}
$$

$$
\delta_{K^*_3+1} \leq \frac{\pi_1(1)}{\phi \delta S(1) + (1 - \phi) \pi_1(1)}
$$

Finally, let

$$
K_3(L) = \begin{cases} 
K^*_3(L), & \text{if } K^*_3 < K^*_3(L) \\
K_2(L), & \text{otherwise.}
\end{cases}
$$

One can immediately see that the right hand-side of Equation (A.38) and Equation (A.40) are increasing in $L$, so that both $K_2(L)$ and $K_3(L)$ are decreasing in $L$. Moreover, right hand-side of Equation (A.38) is smaller than right hand-side of Equation (A.40), so that $K_2(L) > K_3(L)$.

Set $\bar{L}$ such that:

$$
K_3(\bar{L}) = 1
$$

and set $L$ such that:

$$
K_2(L) = K_1
$$

By Equations (A.36)-(A.40),

- When $L \geq \bar{L}$,
  - A pooling equilibrium without layoffs exists when $K > K_1$.
  - A pooling equilibrium with layoffs exists when $K \leq K_1$.
  - A separating equilibrium does not exist because $K_3(L) < 1$.

- When $L \in (L, \bar{L})$,
  - A pooling equilibrium without layoffs exists when $K > K_1$.
  - A pooling equilibrium with layoffs exists when $K_1 \geq K > K_2(L)$.
  - A separating equilibrium exists when $K_3(L) \geq K$.

- When $L \leq L$,
  - A pooling equilibrium without layoffs exists when $K > K_2(L)$.
-- A pooling equilibrium with layoffs does not exist because when $K_2(L) > K_1$.
-- A separating equilibrium exists when $K_3(L) \geq K$.

Theorem 1.2 follows from these statements. □

**Proof.** (of Proposition 1.5.) When the PBE is a pooling equilibrium without layoffs, the welfare of a newborn worker is:

$$J_U = \frac{s}{(L-1)(1-\delta)} + s \frac{\delta}{1-\delta} \beta(\phi + (1-\phi)\eta(1-\delta)}k(\phi)$$
(A.43)

which does not depend on $K$.

When the PBE is a pooling equilibrium with layoffs, the welfare of a newborn worker is:

$$J_U = \frac{s}{(L-1)(1-\delta)} + s \frac{\delta}{1-\delta} \beta(\phi + (1-\phi)\eta(1-\delta)}k(\phi) + (1-\phi)\delta - \frac{\eta \delta}{1-\eta \delta} v$$
(A.44)

By Equation (A.34), $v$ is decreasing in $K$, and therefore $J_U$ is decreasing in $K$. Moreover, since this is a pooling equilibrium with layoffs, by Theorem 1.1:

$$v \leq \frac{\delta}{1-\delta} \beta \eta k(\phi) \frac{1}{1+\delta(1-\eta)}$$

Therefore, $J_U$ satisfies:

$$J_U \leq \frac{s}{(L-1)(1-\delta)} + s \frac{\delta}{1-\delta} \beta(\phi + (1-\phi)\eta(1-\delta)}k(\phi)$$

Then, Equation (1.13) guarantees that $J_U$ is lower than the value defined in Equation (A.43).

When the PBE is a separating equilibrium, the welfare of a newborn worker is:

$$J_U = \frac{s}{(L-1)(1-\delta)} + s \frac{\delta}{1-\delta} \beta(\phi + (1-\phi)\eta)k(1) + (1-\phi)v$$
(A.45)

By Equation (A.34), $v$ is decreasing in $K$, and therefore $J_U$ is decreasing in $K$. Moreover, since this is a pooling equilibrium with layoffs, by Theorem 1.1:

$$v \geq \frac{\delta}{1-\delta} \beta \eta k(1)$$

Therefore, $J_U$ satisfies:

$$J_U \geq \frac{s}{(L-1)(1-\delta)} + s \frac{\delta}{1-\delta} \beta(\phi + (1-\phi)\eta)k(1)$$
Since $k(1) > k(\phi)$, $J^U$ is higher than the value defined in Equation (A.43).
Appendix B

Appendix to Chapter 2

B.1 Proof of Proposition 2.3

Proposition 2.3 is a corollary of the following proposition.

**Proposition B.1.** Take any set of contracts $\tilde{X} \subseteq X$ with the following property: if $(i^*, a^*, l) \in \tilde{X}$ for some $i^* \in I$, then $(i^*, a^*, B) \in \tilde{X}$.

If $D_{a^*}(\tilde{X}) \neq C_{a^*}(\tilde{X})$, then $|D_{a^*}(\tilde{X}) \setminus C_{a^*}(\tilde{X})| = |C_{a^*}(\tilde{X}) \setminus D_{a^*}(\tilde{X})| = 1$. Let $x \in (D_{a^*}(\tilde{X}) \setminus C_{a^*}(\tilde{X}))$ and $x' \in (C_{a^*}(\tilde{X}) \setminus D_{a^*}(\tilde{X}))$. Then,

- $x \in \tilde{X}(NB)$, and,
- $x' \in \tilde{X}(I \setminus w(a^*), B)$.

**Proof.** We first note that any $x \in \tilde{X}(w(a^*), B)$ is unacceptable for any slot in $S_{a^*}$, hence these contracts will be rejected. Since both choice functions satisfy irrelevance of rejected contracts, we remove all such contracts from $\tilde{X}$. With a slight abuse of notation, we denote the remaining contracts with $\tilde{X}$. We proceed by induction on the number of slots at $a^*$, $|S_{a^*}|$.

The base case $|S_{a^*}| = 1$ is immediate, as then $S_{a^*} = \{a^*\}$ and $C_{a^*}(\tilde{X}) \neq D_{a^*}(\tilde{X})$ if and only if a busing contract with a non-walk-zone student is selected under $C_{a^*}$, but a non-busing contract is selected under $D_{a^*}$. It follows immediately from this observation that statement is true.

Given the result for the base case $|S_{a^*}| = 1$, we suppose that the result holds for all $|S_{a^*}| < l$ for some $l > 1$, and show that this implies the result for $|S_{a^*}| = l$. Let $|S_{a^*}| = l$, and let $\tilde{s} \in S_{a^*}$ be the slot which is minimal (i.e., processed last) under the precedence order $\triangleright_{a^*}$.

If $\tilde{s} = s^*$, then the result follows as in the base case: It is clear from the algorithms defining $C_{a^*}$ and $D_{a^*}$ that the same contracts are chosen for slots $s \triangleright_{a^*} s^* = \tilde{s}$ in the computations of $C_{a^*}(\tilde{X})$ and $D_{a^*}(\tilde{X})$, as those slots’ priorities and the precedence order are fixed. Hence, the same set of contracts will be considered for slot $s^*$ under both $C_{a^*}$ and $D_{a^*}$. Thus, as in the base case, $C_{a^*}(\tilde{X}) \neq D_{a^*}(\tilde{X})$ if and only if a busing contract with non-walk-zone student
is selected for $s^*$ under $C_{a^*}$ and a non-busing contract is selected for $s^*$ under $D_{a^*}$.

If $\bar{s} \neq s^*$, then $s^* \triangleright s^* \triangleright \bar{s}$. Let $J \subset \bar{X}$ and $\bar{J} \subset \bar{X}$ be the set of contracts selected for slots in $S_{a^*} \setminus \{\bar{s}\}$ in the computation of $C_{a^*}(\bar{X})$ and $D_{a^*}(\bar{X})$, respectively. If $J = \bar{J}$, then the last slot is filled by the same contract in the computation of both choice functions and $C_{a^*}(\bar{X}) = D_{a^*}(\bar{X})$. Otherwise, the inductive hypothesis, in the case $|S_{a^*}| = l - 1$, implies that $|\bar{J} \setminus J| = |J \setminus \bar{J}| = 1$, $x \in (\bar{J} \setminus J)$, and $x' \in (J \setminus \bar{J})$ and either:

1. $x \in \bar{X}(I \setminus w(a^*), NB)$ and $x' \in \bar{X}(I \setminus w(a^*), B)$, or
2. $x \in \bar{X}(w(a^*), NB)$ and $x' \in \bar{X}(I \setminus w(a^*), B)$.

This is because the first $l - 1$ slots of $a^*$ can be treated as a school with slots $S_{a^*} \setminus \{\bar{s}\}$ (under the precedence order induced by $\triangleright s^*$).

**Case 1:** $x \in \bar{X}(I \setminus w(a^*), NB)$ and $x' \in \bar{X}(I \setminus w(a^*), B)$.

In this case, $i(x) \in I \setminus w(a^*)$ has higher priority under $\pi$ than all students with non-busing contract in $i(\bar{X}) \setminus i(J)$. By the assumption on $\bar{X}$ stated in the proposition, $i(x)$ has a busing contract in $\bar{X} \setminus J$ as well. Similarly, $i(x') \in I \setminus w(a^*)$ has higher priority under $\pi$ than all students in $I \setminus w(a^*)$ with busing contract in $i(\bar{X}) \setminus i(J)$ and all students in $w(a^*)$ with non-busing contract in $i(\bar{X}) \setminus i(J)$. By the assumption on $\bar{X}$ stated in the proposition, and because we removed the busing contracts with walk-zone students, we conclude that $i(x')$ has higher priority under $\pi$ than all students with a contract in $i(\bar{X}) \setminus i(J)$.

- If $\bar{s}$ is an open slot, then $x'$ is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$ and a busing contract with a non-walk-zone student, possibly $(i(x), a^*, B)$, is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$.
- If $\bar{s}$ is a walk-zone slot, then
  - a non-busing contract with a walk-zone student is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$ and $D_{a^*}$, or,
  - $x'$ is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$ and another busing contract with non-walk-student, possibly $(i(x), a^*, B)$, is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$.
- If $\bar{s}$ is non-busing slot, then $x$ is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$ and
  - a non-busing contract is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$, or,
  - $x'$ is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$.

In any case, the statement is satisfied.

**Case 2:** $x \in \bar{X}(w(a^*), NB)$ and $x' \in \bar{X}(I \setminus w(a^*), B)$.

---

1 For the proof of Proposition 2.6, change this bullet point as follows: If $\bar{s}$ is a non-commuting slot, a non-busing contract with walk-zone student is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$ and $D_{a^*}$, or $x$ is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$ and either a non-busing contract or $x'$ is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$.
In this case, $i(x) \in w(a^*)$ has higher priority under $\pi$ than all students in $w(a^*)$ with a contract in $i(X) \setminus i(J)$. By the same reasoning as in Case 1, $i(x') \in I \setminus w(a^*)$ has higher priority under $\pi$ than all students with a contract in $i(X) \setminus i(J)$.

- If $\bar{s}$ is an open slot, then $x'$ is selected in the computation of $D_{a^*}$ and
  - a busing contract with non-walk-zone student is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$, or,
  - $x$ is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$.
- If $\bar{s}$ is a walk zone slot, then $x$ is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$ and
  - a non-busing contract with a walk-zone student is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$, or,
  - $x'$ is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$.
- If $\bar{s}$ is a non-busing slot, then
  - a non-busing contract with a non-walk-zone student is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$ and $D_{a^*}$, or,
  - $x$ is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$ and a non-busing contract is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$, or,
  - $x$ is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$ and $x'$ is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$.

In all possible cases the statement is satisfied, and the proof follows. \[\square\]

### B.2 Proof of Proposition 2.4

**Proof.** Suppose that $s^* \in S_{a^*}$ is an open slot which is then converted to a non-busing slot. Take a student $i \in I \setminus w(a^*)$ with $(i, a^*, NB) \in \hat{X}$. By the assumption on $\hat{X}$, $(i, a^*, B) \in \hat{X}$.

Assume, towards a contradiction, that $i \in i(C_{a^*}(x))$ but $i \notin i(D_{a^*}(x))$. Note that this immediately implies: $D_{a^*}(x) \neq C_{a^*}(x)$. By Proposition B.1:

- $C_{a^*}(\hat{X}) \setminus D_{a^*}(\hat{X}) = \{(i, a^*, B)\}$, and,
- $D_{a^*}(\hat{X}) \setminus C_{a^*}(\hat{X}) = \{(j, a^*, NB)\}$ for some $j \in i(\hat{X}) \setminus \{i\}$.

Let $s \in S_{a^*}$ be the slot $(j, a^*, NB)$ is chosen for under $D_{a^*}$, and let $s' \in S_{a^*}$ be the slot $(i, a^*, B)$ is chosen for under $C_{a^*}$. Consider two cases, mutually exclusive under regularity of the precedence order:

1. Suppose all the walk-zone slots precede all the open slots. By the same reasoning as in the proof of Proposition B.1, $i$ has higher priority than all students in $i(\hat{X}) \setminus i(C_{a^*}(\hat{X}))$.

\[\text{Footnote: For the proof of Proposition 2.6, change this bullet point as follows: If $\bar{s}$ is non-commuting slot, then $x$ is chosen for slot $\bar{s}$ in the computation of $C_{a^*}$ and a non-busing contract or $x'$ is chosen for slot $\bar{s}$ in the computation of $D_{a^*}$.}\]
In particular, \( \pi(i) < \pi(j) \).

By the algorithms defining \( C_{a^*} \) and \( D_{a^*} \), all the walk-zone slots are processed before \( s^* \), and the same set of contracts are chosen for these slots. Consequently, \( s \) must be an open slot or a non-busing slot.

- If \( s \) is an open slot, because \( \pi(i) < \pi(j) \), \((i, a^*, B)\) must be chosen for \( s \) under \( D_{a^*} \), a contradiction.
- If \( s \) is a non-busing slot, because \( \pi(i) < \pi(j) \), \((i, a^*, NB)\) must be chosen for \( s \) under \( D_{a^*} \), a contradiction.

2. Suppose all the open slots precede all the walk-zone slots. It must be the case that \( j \in w(a^*) \), because otherwise some contract with \( i \) would be chosen for \( s \) under \( D_{a^*} \). Then, \( s' \) cannot be a walk-zone slot, because otherwise \((j, a^*, NB)\) would be chosen for \( s' \) under \( C_{a^*} \). Moreover, \( s' \) cannot be a non-busing slot either, because otherwise \((i, a^*, NB)\) would be chosen for \( s' \) under \( C_{a^*} \). We conclude that \( s' \) must be an open slot.

Let \( k \leq |S_{a^*}| \) be the processing order of \( s' \) under the precedence order \( \triangleright a^* \). Let \( J \subset \bar{X} \) be the set of contracts selected for the first \( k \) slots in the calculation of \( C_{a^*} \). Similarly, let \( \bar{J} \subset \bar{X} \) be the set of contracts selected for the first \( k \) slots in the calculation of \( D_{a^*} \). Since \( J \neq \bar{J} \), by Proposition B.1:

- \( J \setminus \bar{J} = \{(i, a^*, B)\} \), and,
- \( \bar{J} \setminus J = \{(m, a^*, NB)\} \) for some \( m \in i(\bar{X}) \setminus \{i\} \).

Note that \( i \) has higher priority than all students in \( i(\bar{X}) \setminus i(J) \). In particular, \( \pi(i) < \pi(m) \).

Let \( s'' \in S_{a^*} \) be the slot \((m, a^*, NB)\) is chosen for under \( D_{a^*} \). Because all the open slots precede all the walk-zone slots, \( s'' \) must be an open or a non-busing slot.

- If \( s'' \) is an open slot, because \( \pi(i) < \pi(m) \), \((i, a^*, B)\) must be chosen for \( s'' \) under \( D_{a^*} \), a contradiction.
- If \( s'' \) is a non-busing slot, because \( \pi(i) < \pi(m) \), \((i, a^*, NB)\) must be chosen for \( s'' \) under \( D_{a^*} \), a contradiction.

In any case, we obtain a contradiction, and the result follows.

\[ \square \]

### B.3 Irregular Precedence Orders

**Example B.1.** Take a school \( a^* \in A \) and students \( \{i_1, i_2, i_3, i_4\} \subseteq I \). Let \( \{i_2, i_4\} \subseteq w(a^*) \), \( \{i_1, i_3\} \subseteq (I \setminus w(a^*)) \), and \( \pi(i_1) < \pi(i_2) < \pi(i_3) < \pi(i_4) \).

Suppose \( S_{a^*} = \{s_{a^*}^1, s_{a^*}^2, s_{a^*}^3\} \), and \( s_{a^*}^1 \triangleright a^* s_{a^*}^2 \triangleright a^* s_{a^*}^3 \). Let \( s_{a^*}^1 \) be an open slot (which is then converted to a non-busing slot), \( s_{a^*}^2 \) be a walk-zone slot, and \( s_{a^*}^3 \) be an open slot. Note that
the precedence order is not regular. Let

\[ \tilde{X} = \{(i_1, a^*, B), (i_2, a^*, B), (i_2, a^*, NB), (i_3, a^*, B), (i_3, a^*, NB), (i_4, a^*, B), (i_4, a^*, NB)\} \]

Note that \( \tilde{X} \) satisfies the conditions specified in Proposition 2.4. Let \( C_{a^*} \) be the choice function induced when \( s^1_{a^*} \) is an open slot. In this case:

\[ C_{a^*}(\tilde{X}) = \{(i_1, a^*, B), (i_2, a^*, NB), (i_3, a^*, B)\} \]

Let \( D_{a^*} \) be the choice function induced when \( s^1_{a^*} \) is a non-busing slot. In this case:

\[ D_{a^*}(\tilde{X}) = \{(i_1, a^*, B), (i_2, a^*, NB), (i_4, a^*, NB)\} \]

Note that \( i_3 \in I \setminus w(a^*) \) and \( (i_3, a^*, NB) \in \tilde{X}. \) However, \( i_3 \notin i(D_{a^*}(\tilde{X})) \) even though \( i_3 \in i(C_{a^*}(\tilde{X})) \). Thus, the conclusion of Proposition 2.4 does not hold.

### B.4 Omitted Example

**Example B.2.** \( A = \{a, b, c, d, e\} \) with \( (|S_a|, |S_b|, |S_c|, |S_d|, |S_e|) = (5, 1, 1, 1, 1) \).

\( I = \{i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8, i_9\} \) where \( w(a) = \{i_1, i_2, i_3, i_4\} \), \( w(b) = \{i_5\} \), \( w(d) = \{i_7\} \), and \( w(e) = \{i_9\} \). The tie-breaker is such that

\[ \pi(i_6) < \pi(i_8) < \pi(i_1) < \pi(i_2) < \pi(i_5) < \pi(i_9) < \pi(i_4) < \pi(i_3) < \pi(i_7) \]

The preference profile is:

<table>
<thead>
<tr>
<th></th>
<th>( P_{i_1} )</th>
<th>( P_{i_2} )</th>
<th>( P_{i_3} )</th>
<th>( P_{i_4} )</th>
<th>( P_{i_5} )</th>
<th>( P_{i_6} )</th>
<th>( P_{i_7} )</th>
<th>( P_{i_8} )</th>
<th>( P_{i_9} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a, B)</td>
<td>(a, B)</td>
<td>(a, B)</td>
<td>(a, B)</td>
<td>(a, B)</td>
<td>(b, B)</td>
<td>(c, B)</td>
<td>(d, B)</td>
<td>(a, B)</td>
<td></td>
</tr>
<tr>
<td>(a, NB)</td>
<td>(a, NB)</td>
<td>(a, NB)</td>
<td>(a, NB)</td>
<td>(b, B)</td>
<td>(b, NB)</td>
<td>(c, NB)</td>
<td>(d, NB)</td>
<td>(a, NB)</td>
<td></td>
</tr>
<tr>
<td>( \emptyset_{i_1} )</td>
<td>( \emptyset_{i_2} )</td>
<td>(c, B)</td>
<td>(b, NB)</td>
<td>(a, B)</td>
<td>(b, B)</td>
<td>(d, B)</td>
<td>(a, B)</td>
<td>(e, B)</td>
<td></td>
</tr>
<tr>
<td>( \emptyset_{i_3} )</td>
<td>( \emptyset_{i_4} )</td>
<td>(e, B)</td>
<td>( \emptyset_{i_5} )</td>
<td>(a, B)</td>
<td>(a, NB)</td>
<td>(d, NB)</td>
<td>(a, NB)</td>
<td>(e, NB)</td>
<td></td>
</tr>
<tr>
<td>( \emptyset_{i_6} )</td>
<td>( \emptyset_{i_7} )</td>
<td>( \emptyset_{i_8} )</td>
<td>( \emptyset_{i_9} )</td>
<td>( \emptyset_{i_6} )</td>
<td>( \emptyset_{i_7} )</td>
<td>( \emptyset_{i_8} )</td>
<td>( \emptyset_{i_9} )</td>
<td>( \emptyset_{i_7} )</td>
<td></td>
</tr>
</tbody>
</table>

First consider the following case:

\( S_a: \) Open \( \triangleright^a \) Open \( \triangleright^a \) Walk-zone \( \triangleright^a \) Walk-zone \( \triangleright^a \) Open, \( S_b: \) Walk-zone, \( S_c: \) Open, \( S_d: \) Walk-zone and \( S_e: \) Open. The outcome of cumulative offer algorithm is:

\[ \mu = \begin{pmatrix}
  i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 & i_8 & i_9 \\
  (a, NB) & (a, NB) & (a, NB) & (a, B) & (b, B) & (c, B) & (d, B) & (e, NB)
\end{pmatrix}. \]

Next, consider the case where the last open slot under \( \triangleright^a \) is replaced with a non-busing slot.
The outcome of cumulative offer algorithm for this case is:

\[ \nu = \begin{pmatrix}
  i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 & i_8 & i_9 \\
  (a, NB) & (a, NB) & (c, B) & (e, B) & (b, NB) & (a, B) & (d, NB) & (a, B) & (a, NB)
\end{pmatrix}. \]

In the first case the number of students assigned to \( a \) with busing contract is 1, whereas in the second case this number is 2. Moreover, in both cases for all schools the number of assigned students with busing contract is 4.

### B.5 Proof of Theorem 2.1

We will refer to the following corollary of Proposition B.1:

**Corollary B.1.** Take any set of contracts \( \bar{X} \subseteq X \) with the following property: if \((i^*, a^*, t) \in \bar{X} \) for some \( i^* \in I \), then \((i^*, a^*, B) \in \bar{X} \). If there are more than \(|S_{a^*}| \) acceptable contracts for all slots, then there are at least \(|\bar{X}| - |S_{a^*}| - 1 \) contracts that are rejected in the computation of both \( C_{a^*} \) and \( D_{a^*} \).

We can now proceed with the proof of Theorem 2.1.

**Proof.** (of Theorem 2.1.) Suppose that the assumptions of the theorem hold, and let \( A = \{a, b\} \). Without loss of generality, we assume the priority structure of school \( a \) has changed, i.e. \( s^* \in S_a \) is converted from an open slot to a non-busing slot. Let the matching \( \mu \) be obtained from the cumulative offer algorithm when \( s^* \) is an open slot (when \( a \) has choice function \( C \)), and let \( \nu \) be the allocation obtained from the cumulative offer algorithm when \( s^* \) is a non-busing slot (when \( a \) has choice function \( D \)).

Since the outcome of cumulative offer algorithm is independent of the proposal order, without loss of generality, consider the proposal order where for the first \(|I| \) steps of the cumulative offer algorithm a different student offers her most preferred contract. Denote the set of contracts offered to school \( a \) and \( b \) at the end of step \(|I| \) by \( X_{a}^{1,k} \) and \( X_{b}^{1,k} \) under choice function \( k \in \{C, D\} \), respectively. According to the proposal order, \( X_{a}^{1,C} = X_{a}^{1,D} \) and \( X_{b}^{1,C} = X_{b}^{1,D} \). To save notation, let \( X_{a}^{1} = X_{a}^{1,C} \) and \( X_{b}^{1} = X_{b}^{1,C} \). Note that, as a consequence of Assumption 2.1, all contracts in \( X_{a}^{1} \cup X_{b}^{1} \) are busing contracts. Let \( U^{1} \) be the set of busing contracts in \( X_{a}^{1} \cup X_{b}^{1} \) that are offered by a walk-zone student, i.e. let

\[ U^{1} = \{ x \in X_{a}^{1} \cup X_{b}^{1} : i(x) \in w(a(x)) \} \]

**Note:**

- Any busing contract offered to school \( a \) by any student in \( w(a) \) cannot be in \( C_{a}(X_{a}^{1}) \) or \( D_{a}(X_{a}^{1}) \), because these contracts are unacceptable for any slot in \( S_{a} \). Similarly, any busing contract offered to \( b \) by any student in \( w(b) \) cannot be in \( C_{b}(X_{b}^{1}) \) or \( D_{b}(X_{b}^{1}) \). Since all contracts in \( U^{1} \) are rejected, \( U^{1} \subseteq (X_{a}^{1} \cup X_{b}^{1}) \setminus (C_{a}(X_{a}^{1}) \cup C_{b}(X_{b}^{1})) \) and \( U^{1} \subseteq (X_{a}^{1} \cup X_{b}^{1}) \setminus (D_{a}(X_{a}^{1}) \cup D_{b}(X_{b}^{1})) \).
• Under both \(C_a\) and \(D_a\) any contract in \(X_a^1 \setminus U^1\) and \(X_b^1 \setminus U^1\) are acceptable for all slots in \(S_a\) and \(S_b\), respectively.

Let \(I^1\) be the set of students with a contract in \(U^1\), i.e. \(I^1 = i(U^1)\). Under both choice functions, consider the proposal order where students in \(I^1\) offer their next most preferred contracts. That is, for the next \(|U^1| = |I^1|\) steps of the cumulative offer algorithm, students in \(I^1\) offer their second choices. Denote the set of contracts offered to school \(a\) in these steps by \(\tilde{X}_a\) and the set of contracts offered to school \(b\) in these steps by \(\tilde{X}_b\). Then, the set of contracts offered to school \(a\) at the end of step \(|I| + |I^1|\) is \(X_a^1 \cup \tilde{X}_a\) under both choice functions. Similarly, the set of contracts offered to school \(a\) at the end of step \(|I| + |I^1|\) is \(X_a^1 \cup \tilde{X}_a\) under both choice functions. Moreover, because both choice functions satisfies irrelevance of rejected contracts condition, at the end of step \(|I| + |I^1|\) one can remove the rejected contracts in previous steps. Accordingly, we define

\[
\begin{align*}
X_a^2 &:= (X_a^1 \cup \tilde{X}_a) \setminus U^1 \\
X_b^2 &:= (X_b^1 \cup \tilde{X}_b) \setminus U^1
\end{align*}
\]

These are the set of contracts school \(a\) and school \(b\) consider at the end of step \(|I| + |I^1|\) under both choice functions. Note that all contracts in \(X_a^2\) and \(X_b^2\) are acceptable by all slots in \(S_a\) and \(S_b\), respectively. This is due to two reasons:

1. \(U^1\) is specifically constructed such that all the busing contracts offered by a walk-zone student are removed.

2. As a consequence of Assumption 2.1, a busing contract by a non-walk-zone student is offered before a non-busing contract. Therefore, \(X_a^2 \cup X_b^2\) does not contain any non-busing contracts offered by a non-walk-zone student.

Consider the following cases:

**Case 1:** \(|X_a^2| \leq |S_a|\) and \(|X_b^2| \leq |S_b|\)

Under both choice functions, all the contracts in \(X_a^2\) and \(X_b^2\) will be accepted by \(a\) and \(b\), respectively. We conclude that \(\nu(a) = \mu(a)\) and \(\nu(b) = \mu(b)\), hence the number of non-busing contracts selected under \(\nu\) and \(\mu\) are equal to each other.

**Case 2:** \(|X_a^2| < |S_a|\) and \(|X_b^2| > |S_b|\)

Because all the contracts in \(X_a^2\) are acceptable by all slots in \(S_a\) under both \(C_a\) and \(D_a\), \(C_a(X_a^2) = D_a(X_a^2) = X_a^2\). Because \(C_b = D_b\), \(C_b(X_b^2) = D_b(X_b^2)\). Moreover, because all the contracts in \(X_b^2\) are acceptable by all slots in \(S_b\), \(|C_b(X_b^2)| = |D_b(X_b^2)| = |S_b|\).

In the following steps of cumulative offer algorithm, consider the students who do not have a tentatively accepted contract. Under Assumption 2.1, in any step \(k > |I| + |I^1|\), if a student offers a contract to \(a\) then it is either an acceptable contract by all slots in \(S_a\) under both \(C_a\) and \(D_a\), or it is an unacceptable contract by all slots in \(S_a\) under both \(C_a\) and \(D_a\). If it is an acceptable contract, it will be chosen under both \(C_a\) and \(D_a\) without causing the rejection of any contract in \(X_a^2\) because \(|S_a| \geq |I| - |S_b|\). Therefore, in any step \(k > |I| + |I^1|\) the set of contracts offered to \(b\) will be the same under both choice functions and the set of rejected
and accepted contracts will be the same since $C_b = D_b$. Hence, the number of non-busing contracts under $\nu$ and $\mu$ are equal to each other.

**Case 3:** $|X^a_2| > |S_a|$ and $|X^b_2| > |S_b|$  

This case is not possible since we are assuming $|I| \leq |S_a| + |S_b|$.

**Case 4:** $|X^a_2| > |S_a|$ and $|X^b_2| < |S_b|$  

First, consider the case where $|X^a_2| > |S_a| + 1$. By Corollary B.1, there exists at least $|X^a_2| - |S_a| - 1$ contracts in $(X^a_2 \setminus C_a(X^a_2)) \cap (X^a_2 \setminus D_a(X^a_2))$. Let $x_1 \in (X^a_2 \setminus C_a(X^a_2)) \cap (X^a_2 \setminus D_a(X^a_2))$, and let $x_3 = X^a_2 \setminus x_1$. In step $|I| + |I_1| + 1$, $i(x_1)$ offers her next most preferred contract. If that contract is offered to $a$, then add that contract to $X^a_3$. If that contract is offered to $b$ and $i(x_1) \notin w(b)$, then it will be accepted and will not cause the rejection of any contract in $X^b_2$ (because $|X^b_2| < |S_b|$). Otherwise, consider her next best contract and add it to $X^a_3$ if her next best is offered to $a$. That is, either $|X^a_2| = |X^a_3| - 1$ or $|X^a_2| = |X^b_2|$. In general, in step $|I| + |I_1| + k$, if $|X^{k+1}_a| > |S_a| + 1$ then continue with a student associated with a contract in $(X^{k+1}_a \setminus C_a(X^{k+1}_a)) \cap (X^{k+1}_a \setminus D_a(X^{k+1}_a))$. Construct $X^{k+2}_a$ by removing that contract from $X^{k+1}_a$, then allowing this student to offer her next most preferred contract. If that contract has term $B$ and if that student is in $w(b)$, consider the following best contract of her.

Due to the finite number of students, at some $T > 3$ we have $|X^a_T| = |S_a| + 1$. At this stage, if $C_a(X^a_T) = D_a(X^a_T)$, consider the contract in $X^a_T \setminus C_a(X^a_T)$ and continue in the same manner. If $C_a(X^a_T) \neq D_a(X^a_T)$, by Proposition B.1, there exist two contracts such that $x \in [D_a(X^a_T) \setminus C_a(X^a_T)]$ and $x' \in [C_a(X^a_T) \setminus D_a(X^a_T)]$. There are two possible cases:

1. $x \in X(I \setminus w(a), NB)$ and $x' \in X(I \setminus w(a), B)$.
2. $x \in X(w(a), NB)$ and $x' \in X(I \setminus w(a), B)$.

Let $i = i(x)$ and $i' = i(x')$. Let $n^C_a$, $n^C_b$, $n^D_a$ and $n^D_b$ be the number of contracts with NB term in $C_a(X^a_T)$, $C_b(X^b_T)$, $D_a(X^a_T)$ and $D_b(X^b_T)$, respectively. Note that $n^b_C = n^D_b$ and $n^a_C = n^D_a + 1$.

Under choice function $D_a$, in step $T + 1$, $i'$ will offer her next best contract. Since $(i', a, B)$ is rejected, she will offer $\emptyset$, $(i', a, NB)$, $(i', b, B)$ or $(i', b, NB)$. Note that any contract in $D_b(X^b_T)$ will not be rejected in any case, because $|D_b(X^b_T)| < |S_b|$ and all contracts in $D_b(X^b_T)$ are acceptable by all slots in $S_b$. Therefore, the number of contracts with NB term tentatively accepted in step $T + 1$ will be greater than or equal to $n^a_D + n^b_D$. In general, in any further step $m \geq T + 1$, there will be at most one student whose contract is not tentatively accepted. This student will offer her next best contract, and any contract in $D_b(X^b_T)$ will not be rejected in any further step. Following the same reasoning, in any step $m \geq T + 1$ (and consequently, under $\nu$) the number of contracts with NB term tentatively accepted will be greater than or equal to $n^a_D + n^b_D = n^C_a + n^C_b + 1$.

For choice function $C$, we investigate the cases below.

**Case 4.1:** $x \in X(I \setminus w(a), NB)$ and $x' \in X(I \setminus w(a), B)$: Under $C_a$, $i$ will apply either to remaining unassigned option or to $b$. If $i$ applies to $\emptyset$, then the number of contracts with NB term in $\mu$ is $n^C_a + n^C_b$. Suppose $i$ applies to $b$. If $i \in w(b)$, then she applies with a non-busing
contract and will be accepted.\(^3\) Otherwise, she applies with a busing contract and will be accepted. Hence the number of contracts with term NB in \(\mu\) is at most \(n_a^C + n_b^C + 1\).

**Case 4.2:** \(x \in X(w(a), NB)\) and \(x' \in X(I \setminus w(a), B)\): Under \(C_a\), \(i\) will apply either to remaining unassigned option or to \(b\). If \(i\) applies to \(\emptyset\), then the number of contracts with NB term in \(\mu\) is \(n_a^C + n_b^C\). Otherwise, \(i\) will apply to \(b\) with a busing contract and will be accepted. Hence the number of contracts with term NB in \(\mu\) is \(n_a^C + n_b^C\).

This completes the analysis of all possible cases, and the proof follows. \(\square\)

### B.6 Proof of Proposition 2.5

**Proof.** In the proof, we will refer to the notation and objects constructed in the proof of Theorem 2.1.

Begin by noting that in the proof of Theorem 2.1, the only case where \(\mu \neq \nu\) is possible is Case 4. Therefore, for any student to have a different allocation under \(\nu\) than \(\mu\), we must be in the case:

\[|X_a^2| > |S_a| \text{ and } |X_b^2| < |S_b|\]

As argued in the proof of Theorem 2.1, in this case, any acceptable contract that is offered to \(b\) is accepted without causing the rejection of any other student. Therefore, for any \(i \notin w(a)\) with \((i, a, NB) P_i \emptyset_i\),

- If \((i, b, NB) \ P_i (i, a, NB)\), then \(i\) will receive a contract which is at least as preferable as \((i, b, NB)\) under \(\nu\). That is, \(\nu(i) \ R_i (i, b, NB) \ P_i (i, a, NB)\).
- If \((i, b, B) \ P_i (i, a, NB)\) and \(i \notin w(b)\), then student \(i\) will receive a contract which is at least as preferable as \((i, b, B)\) under \(\nu\). That is, \(\nu(i) \ R_i (i, b, B) \ P_i (i, a, NB)\).

Therefore, the only remaining cases we need to consider are:

1. \((i, a, NB) \ P_i (i, b, B)\), and,
2. \((i, b, B) \ P_i (i, a, NB) \ P_i (i, b, NB)\) and \(i \in w(b)\).

In either case, suppose \(\mu(i) \neq \nu(i)\) and, towards a contradiction, assume that \(\mu(i) \ R(i) (i, a, NB)\) but \((i, a, NB) \ P_i \nu(i)\).

In the first case, \(\mu(i) \ R(i) (i, a, NB)\) implies that \(a(\mu(i)) = a\). Therefore, \(i \in i(C_a(X_a^T))\). Moreover, \((i, a, NB) \ P_i \nu(i)\) implies that \(a(\mu(i)) \neq a\). Therefore, \(i \notin i(D_a(X_a^T))\) even though \((i, a, NB) \in X_a^T\). This is a contradiction to Proposition 2.4.

In the second case, \((i, b, B)\) is not an acceptable contract for any slot under \(D_a\) and \(D_b\), and therefore \(i\) is never admitted to \(b\) under a busing contract. Consequently, \(\mu(i) \ R(i) (i, a, NB)\)

---

\(^3\)In this case, \(i\) first applies to \(b\) with a busing contract and she will be rejected. Then, we need to consider her next best choice, which is either \(\emptyset\) or \((i, b, NB)\).
implies that \( a(\mu(i)) = a \). Moreover, \((i, a, NB) P_i \nu(i)\) implies that \( a(\mu(i)) \neq a \). Therefore, \( i \notin i(D_a(X_a^T)) \) even though \((i, a, NB) \in X_a^T\). Once again, this is a contradiction to Proposition 2.4.

In either case we have a contradiction, and the result follows. \( \square \)
Appendix C

Appendix to Chapter 3

C.1 Proof of Lemma 3.1

We will begin by considering a more general class of mechanisms than the one we consider in the main text, which are persuasion mechanisms as denoted by Kolotilin et al. (2017). In a persuasion mechanism, the politician is able to commit to a distinct signal distribution for each citizen. To this end, the politician asks each citizen to report her prior belief, and sends an action recommendation \( m \in \{0, 1\} \) to the citizen\(^1\) conditional on her report and the realized state.

A persuasion mechanism, therefore, is:

\[
\psi := \{(\psi_0^i, \psi_1^i)\}_{i \in I}
\]

where

\[
\psi_0^i = \Pr\{m = 1 | \theta = 0\} \\
\psi_1^i = \Pr\{m = 1 | \theta = 1\}
\]

are the probabilities of "personalized" action recommendations conditional on the reported type \( p_i \) and the state \( \theta \). The requirement to report the types truthfully and follow the action recommendation imposes an incentive compatibility requirement, which is formulated next.

Given a persuasion mechanism \( \psi \), the subjective utility attained by citizen \( i \in I \) if she reports type \( p_j \) and takes actions \( a_0 \in \{0, 1\} \) following \( m = 0 \), \( a_1 \in \{0, 1\} \) following \( m = 1 \) is:

\[
U_\psi(p_i, p_j, a_0, a_1) := (1 - p_i)(a_0(1 - \psi_0^j) + a_1 \psi_0^j)(-c) + p_i(a_0(1 - \psi_1^j) + a_1 \psi_1^j)(1 - c)
\]

\(^1\)Due to the linearity of payoffs in actions, the citizen would prefer to have \( a_i \in \{0, 1\} \) unless she has a knife-edge case of beliefs. In case of indifference, we will assume that the citizen will take \( a = 1 \), as the politician can always induce a slightly higher belief to break the indifference.
**Definition C.1.** A persuasion mechanism \( \psi \) is **incentive compatible** if, for each \( i \in I \):

\[
U_\psi(p_i) := U_\psi(p_i, p_i, 0, 1) \geq U_\psi(p_i, p_j, a_0, a_1) \quad \forall j \in I, a_0 \in \{0, 1\}, a_1 \in \{0, 1\}
\]

The following result clarifies why the class of persuasion mechanisms are more general than the class of mechanisms considered in the main text.

**Lemma C.1.** Let \( S \) be a set of signals, and assume the politician commit to some publicly observable signal distribution

\[
\sigma = (\sigma_0, \sigma_1) \in \Delta(S) \times \Delta(S)
\]

The outcome of \( \sigma \) can be implemented via a incentive compatible persuasion mechanism.

**Proof.** Fix \( \sigma \). For any citizen \( i \in I \), \( \sigma \) generates two conditional distributions of posteriors at two states \( \theta \in \{0, 1\} \). Let the cdf of posterior distribution under \( \theta = 0 \) be \( H_{\sigma,0}(\cdot) \), and the cdf of posterior distribution under \( \theta = 1 \) be \( H_{\sigma,1}(\cdot) \). Then, set:

\[
\psi_0^i = 1 - H_{\sigma,0}(c) \\
\psi_1^i = 1 - H_{\sigma,1}(c)
\]

It is trivial to check that the persuasion mechanism \( \{(\psi_0^i, \psi_1^i)\}_{i \in I} \) is incentive compatible. \( \square \)

Given Lemma C.1, we restrict attention to incentive compatible persuasion mechanisms. We will demonstrate that if \( f(.) \) is single-peaked or single-dipped, the optimal strategy can be implemented by only two signals.

The following Lemma follows from usual arguments in mechanism design, and echoes Lemma 1 of Kolotilin et al. (2017).

**Lemma C.2.** For an incentive compatible persuasion mechanism \( \psi \), the following conditions must be satisfied:

1. \( U_\psi(p_i) = \psi_0^i c + \psi_1^i (1 - c) \) for each \( p_i \),
2. \( \psi_0^i c + \psi_1^i (1 - c) \) is increasing in \( p_i \), and,
3. \( U_\psi(0) = 0 \) and \( U_\psi(1) = 1 - c \).

**Proof.** The subjective payoff of citizen \( i \in I \) if she reports truthfully and obeys the action recommendation is:

\[
U_\psi(p_i) := (1 - p_i)\psi_0^i (-c) + p_i \psi_1^i (1 - c) \\
= -\psi_0^i c + (\psi_0^i c + \psi_1^i (1 - c))p_i
\]

One can then use the standard envelope arguments to derive that the first two conditions are necessary for \( \psi \) to be incentive compatible.
By construction, $U_\psi(p_i)$ is bounded above by the payoff of "full disclosure" solution:

$$\bar{U}(p_i) = (1-c)p_i$$

and bounded below by the payoff of "no disclosure" solution:

$$U(p_i) = \max\{p_i - c, 0\}$$

Consequently, one must have: $U_\psi(0) = 0$ and $U_\psi(1) = 1 - c$. \hfill \Box

The figure below illustrates the main idea of Lemma C.2: incentive compatibility of $\psi$ restricts $U_\psi(p_i)$ to be an increasing and convex function between $U(p_i)$ and $\bar{U}(p_i)$.

![Figure C-1: Illustration of Lemma C.2.](image)

**Lemma C.3.** If $f(.)$ is single-peaked, optimal incentive compatible persuasion mechanism $\psi^*$ induces a piecewise linear $U_{\psi^*}(p_i)$ as follows:

$$U_{\psi^*}(p_i) = \begin{cases} 
0, & \text{if } p_i < p \\
\tau(p_i - p), & \text{if } p_i \geq p 
\end{cases}$$

where $\tau \in [1-c, 1]$ and $p = 1 - \frac{1-c}{\tau}$.

**Proof.** The subjective payoff of politician under an incentive compatible persuasion mechanism $\psi$ is:

$$V(\psi) := \int_0^1 \left(((1-p^*)\psi_0^* + p^*\psi_1^*) f(p_i)dp_i \right.$$  

Substituting $p^* = (1-c)$ and using the first part of Lemma C.2 yields:

$$V(\psi) = \int_0^1 \left(c\psi_0^* + (1-c)\psi_1^* \right) f(p_i)dp_i = \int_0^1 U_{\psi^*}(p_i)f(p_i)dp_i$$

133
Integration by parts and using the last part of Lemma C.2 gives:

\[ V(\psi) = (1 - c)f(1) - \int_0^1 U_\psi(p_i)f'(p_i)dp_i \]

Therefore, the optimal incentive compatible persuasion mechanism must be the solution to:

\[ \max_{\psi} - \int_0^1 U_\psi(p_i)f'(p_i)dp_i \]

subject to:

- \( U_\psi(p_i) \) is increasing and convex,
- \( U(p_i) \leq U_\psi(p_i) \leq \bar{U}(p_i) \)

Since \( f(.) \) is single-peaked and continuously differentiable, \( f'(.) \) is continuous and crosses zero at most once from above. Let \( \bar{p} \) be the point where \( f'(\bar{p}) = 0 \). Then, politician wants to minimize \( U_\psi(p_i) \) for \( p_i < \bar{p} \) and maximize \( U_\psi(p_i) \) for \( p_i \geq \bar{p} \). One can then replicate the argument in Theorem 2 of Kolotilin et al. (2017) to prove that a piecewise linear payoff schedule improves upon any other increasing and convex payoff schedule. Below is an illustration of the idea:

![Figure C-2: Optimality of a piecewise linear function.](image)

Here, the blue curve is the increasing and convex payoff schedule, and the red curve is the piecewise linear function that improves upon it.

**Lemma C.4.** If \( f(.) \) is single-peaked, the payoff schedule of optimal incentive compatible persuasion mechanism \( U_{\psi^*}(p_i) \) can be implemented by a publicly observable signal distribution with only two signals.

**Proof.** By Lemma C.3, \( U_{\psi^*}(p_i) \) is piecewise and linear with a kink at \( p \). One can then take

\[ \text{If } f'(p) > 0 \text{ for all } p \in [0, 1], \text{ set } \bar{p} = 0. \text{ If } f'(p) < 0 \text{ for all } p \in [0, 1], \text{ set } \bar{p} = 1. \]
a publicly observable signal distribution

\[ \sigma = (\sigma_0, \sigma_1) \in [0, 1] \times [0, 1] \]

and set \( \sigma_1 = 1 \) and \( \sigma_0 = \frac{\tau - (1-c)}{c} \). It is trivial to check that this signal distribution yields the same payoff schedule. \( \square \)

For a single-dipped \( f(.) \), one can use the symmetric arguments to replicate Lemma C.3 and C.4.

### C.2 Simplification of Politician’s Problem

This section involves more detailed steps for the simplification of politician’s problem in Section 3.3.1.

Substituting in \( p^* = 1 - c \) into the objective function of politician and rearranging gives:

\[
\left( F\left( \frac{c(1-\sigma_0)}{c(1-\sigma_0) + (1-c)(1-\sigma_1)} \right) - F\left( \frac{c\sigma_0}{c\sigma_0 + (1-c)\sigma_1} \right) \right) (p^*\sigma_1 + (1-p^*)\sigma_0) + \left( 1 - F\left( \frac{c(1-\sigma_0)}{c(1-\sigma_0) + (1-c)(1-\sigma_1)} \right) \right) \\
= \left( F\left( \frac{c(1-\sigma_0)}{c(1-\sigma_0) + (1-c)(1-\sigma_1)} \right) - F\left( \frac{c\sigma_0}{c\sigma_0 + (1-c)\sigma_1} \right) \right) ((1-c)\sigma_1 + \sigma_0) + \left( 1 - F\left( \frac{c(1-\sigma_0)}{c(1-\sigma_0) + (1-c)(1-\sigma_1)} \right) \right) \\
= F\left( \frac{c(1-\sigma_0)}{c(1-\sigma_0) + (1-c)(1-\sigma_1)} \right) (c(-1 + \sigma_0) + (1-c)(-1 + \sigma_1)) - F\left( \frac{c\sigma_0}{c\sigma_0 + (1-c)\sigma_1} \right) (c\sigma_0 + (1-c)\sigma_1) + 1
\]

Once again rearranging, removing the constant term which does not affect the optimal value of the decision variable, and turning this into a minimization instead of a maximization problem, one can write the politician’s problem as the one in Equation 3.1.

### C.3 Proof of Lemma 3.3

Proof. Assume, to get a contradiction, that there is an interior solution \((\sigma_0^*, \sigma_1^*) \in int(\Delta)\), which implies:

\[
\frac{\partial \Pi(F, (\sigma_0^*, \sigma_1^*))}{\partial \sigma_0} = 0 \\
\frac{\partial \Pi(F, (\sigma_0^*, \sigma_1^*))}{\partial \sigma_1} = 0
\]
Taking derivatives of $\Pi$ yields:

$$\frac{\partial \Pi(F,(\sigma_0,\sigma_1))}{\partial \sigma_0} = c \left[ f \left( \frac{c \sigma_0}{c \sigma_0 + (1-c) \sigma_1} \right) - f \left( \frac{c (1-\sigma_0)}{c (1-\sigma_0) + (1-c) (1-\sigma_1)} \right) \right]$$

$$- \left[ F \left( \frac{c (1-\sigma_0) + (1-c) (1-\sigma_1)}{c (1-\sigma_0) + (1-c) (1-\sigma_1)} \right) - F \left( \frac{c \sigma_0}{c \sigma_0 + (1-c) \sigma_1} \right) \right]$$

$$\frac{\partial \Pi(F,(\sigma_0,\sigma_1))}{\partial \sigma_1} = \left( 1-c \right) \left[ f \left( \frac{c (1-\sigma_0) + (1-c) (1-\sigma_1)}{c (1-\sigma_0) + (1-c) (1-\sigma_1)} \right) - f \left( \frac{c \sigma_0}{c \sigma_0 + (1-c) \sigma_1} \right) \right]$$

$$- \left( 1-c \right) \left[ F \left( \frac{c (1-\sigma_0) + (1-c) (1-\sigma_1)}{c (1-\sigma_0) + (1-c) (1-\sigma_1)} \right) - F \left( \frac{c \sigma_0}{c \sigma_0 + (1-c) \sigma_1} \right) \right]$$

Substituting $p$ and $\bar{p}$ from Lemma 3.2, this simplifies to:

$$\frac{\partial \Pi(F,.)}{\partial \sigma_0} = c \left[ f \left( p \right) (1-p) - f \left( \bar{p} \right) (1-\bar{p}) - (F(\bar{p}) - F(p)) \right]$$

$$\frac{\partial \Pi(F,.)}{\partial \sigma_1} = \left( 1-c \right) \left[ f \left( \bar{p} \right) \bar{p} - f \left( p \right) p - (F(\bar{p}) - F(p)) \right]$$

Since $c \in (0,1)$, in the optimal solution, one has:

$$F(\bar{p}) - F(p) = f(p)(1-p) - f(\bar{p})(1-\bar{p})$$

$$F(\bar{p}) - F(p) = f(p)\bar{p} - f(\bar{p})p$$

These equalities can be simultaneously satisfied under two possible scenarios: (i) $p = \bar{p}$, or (ii) $f(p) = f(\bar{p})$. Case (i) corresponds to the $\sigma_0 = \sigma_1$ case, which contradicts our initial assumption that the solution is interior. Therefore, case (ii) must be true. But then the following must hold in the optimal solution:

$$(F(\bar{p}) - F(p)) = f(\bar{p})(\bar{p} - p)$$

which can be rewritten as:

$$\int_p^{\bar{p}} f(s) ds = \int_p^{\bar{p}} f(\bar{p}) ds$$

thus

$$\int_p^{\bar{p}} f(s) ds = \int_p^{\bar{p}} f(\bar{p}) ds$$

where $f(\bar{p}) = f(p)$. It's easy to see that, for any single-peaked or single-dipped $f(.)$, this solution cannot be satisfied for $p \neq \bar{p}$. This is because for a single-peaked $f(.)$, we have: $f(s) > f(\bar{p})$ for all $s \in (p, \bar{p})$. Similarly, for a single-dipped $f(.)$, $f(s) < f(\bar{p})$ for all $s \in (p, \bar{p})$. The contradiction follows.

**C.4 Checking the Boundaries of $\Delta$**

Lemma 3.3 demonstrates that the optimal solution must be on the boundaries. The remaining question is: which boundary? We apply a further change of variables into the problem which

---

3Just drawing $f(.)$ would provide a sufficient visual interpretation.
will allow us to redefine the problem as an optimization problem with one variable only.

1. Begin with the \( \sigma_0 = \sigma_1 \) boundary. The value of objective function in Equation (3.1) in this boundary is \( F(c) \) for \( \sigma_0 \in [0, 1] \).

2. Now, consider the \( \sigma_0 = 0 \) boundary. The value of objective function in this boundary is:

\[
\Pi(F, (0, \sigma_1)) = F \left( \frac{c}{c + (1 - c)(1 - \sigma_1)} \right) (c + (1 - c)\sigma_1)
\]

for \( \sigma_1 \in [0, 1] \). Adopting the transformation \( z := \frac{c}{c + (1 - c)(1 - \sigma_1)} \) suggests that this can be written as: \( \frac{F(z)}{c} \) where \( z \) varies between \( c \) and 1. Note that Equation 3.1 is a minimization problem, so essentially we will calculate

\[
\min_{z \in [c, 1]} \frac{F(z)}{c}
\]

and compare it to the minimum values obtained in other boundaries.

3. Finally, consider the \( \sigma_1 = 1 \) boundary. The value of objective function in this boundary is:

\[
\Pi(F, (\sigma_0, 1)) = c(1 - \sigma_0) + F \left( \frac{c\sigma_0}{c\sigma_0 + 1 - c} \right) (c\sigma_0 + 1 - c)
\]

for \( \sigma_0 \in [0, 1] \). Adopting the transformation \( z := \frac{c\sigma_0}{c\sigma_0 + 1 - c} \), this can be written as:

\[
1 - \frac{1 - F(z)}{1 - z} (1 - c)
\]

where \( z \) varies between 0 and \( c \). Therefore, in this boundary, we will calculate

\[
\min_{z \in [0, c]} 1 - \frac{1 - F(z)}{1 - z} (1 - c)
\]

and compare it to the minimum values obtained in other boundaries.

All in all, the politician’s problem can be reduced to:

\[
\min \left\{ F(c), \min_{z \in [c, 1]} \frac{F(z)}{z} c, \min_{z \in [0, c]} 1 - \frac{1 - F(z)}{1 - z} (1 - c) \right\} \quad (C.1)
\]

Finally, one can see that \( F(c) \) is indeed the “corner case” for either of the functions in the minimization problems (plugging \( z = c \) into either function yields \( F(c) \)), hence it is redundant. Therefore, the problem we will consider is:

\[
\min \left\{ \min_{z \in [c, 1]} \frac{F(z)}{z} c, \min_{z \in [0, c]} 1 - \frac{1 - F(z)}{1 - z} (1 - c) \right\}
\]

or, equivalently,

\[
\min_{z \in [0, 1]} G(z) \quad (C.2)
\]

where

\[
G(z) = \begin{cases} 
1 - \frac{1 - F(z)}{1 - z} (1 - c) & \text{if } z \leq c \\
\frac{F(z)}{z} c & \text{if } z \geq c
\end{cases}
\]
This is a straightforward, one-dimensional optimization problem. The interpretation of solution and its mapping to the politician’s original problem is as follows.

- If the minimizer of this problem is \( z = c \), the solution has \( \sigma_0 = \sigma_1 \), i.e. it is fully uninformative.
- If the minimizer of this problem is some \( z \in (c, 1) \), then the solution has \( \sigma_0 = 0 \), i.e. the “good” message is fully informative.
- If the minimizer of this problem is some \( z \in (0, c) \), then the solution has \( \sigma_1 = 1 \), i.e. the “bad” message is fully informative.
- As an extreme case, if some \( z \in \{0, 1\} \) is the minimizer,\(^4\) then the solution has \( \sigma_0 = 0 \) and \( \sigma_1 = 1 \), i.e. the messages are fully informative.

### C.5 Proof of Proposition 3.1

**Proof.** Consider the first case in the proposition. Note that if \( f(.) \) is increasing, than \( F(.) \), being its integral, must be convex in \([0, 1]\). This, in particular, implies that: (i) \( \frac{F(z)}{z} \) is increasing everywhere, and (ii) \( \frac{1-F(z)}{1-z} \) is increasing everywhere. Therefore, we have:

\[
\arg \min_{z \in [c, 1]} \frac{F(z)}{z} = c
\]

and

\[
\arg \min_{z \in [0, c]} \frac{1-F(z)}{1-z} = c
\]

hence \( \arg \min_{z \in [0, 1]} G(z) = c \), which corresponds to the fully uninformative solution.

Now, consider the second case in the proposition. Note that if \( f(.) \) is decreasing, than \( F(.) \) must be concave in \([0, 1]\). This, in particular, implies that: (i) \( \frac{F(z)}{z} \) is decreasing everywhere, and (ii) \( \frac{1-F(z)}{1-z} \) is decreasing everywhere. Therefore, we have:

\[
\arg \min_{z \in [c, 1]} \frac{F(z)}{z} = 1
\]

and

\[
\arg \min_{z \in [0, c]} \frac{1-F(z)}{1-z} = 0
\]

hence \( \arg \min_{z \in [0, 1]} G(z) = \{0, 1\} \), which corresponds to the fully informative solution. \(\square\)

---

\(^4\)One can check that \( G(0) = G(1) \), so if one is a minimizer, the other one is a minimizer too.
C.6 Proof of Proposition 3.2

Proof. Remember the program introduced given by Equation (C.2), and denote the minimizer by $z^*$:

$$z^* := \arg\min_{z \in [0,1]} G(z)$$

where

$$G(z) = \begin{cases} 
1 - \frac{1 - F(z)}{1 - c} (1 - c) & z \leq c \\
\frac{F(z)}{z} & z \geq c 
\end{cases}$$

Recall, for reference, that $z^* = c$ corresponds to the fully uninformative solution, and $z^* \in \{0, 1\}$ corresponds to the fully informative solution. Most of the argument would include ruling out $c$ or $\{0, 1\}$ as potential values of $z^*$.

1. Begin with the first part of Proposition, where $F(c) > c$. We begin by showing that in this case the optimal solution cannot be fully uninformative, i.e. we cannot have $z^* = c$. This is rather easy to see: note that $F(1) = 1$ by construction,

$$\frac{F(1)}{1} = \frac{c}{1} < F(c)$$

That is, $G(1) < G(c)$, so one cannot have $z^* = c$. Now we continue by showing that we must have $\sigma_1 = 1$ in the optimal solution. This corresponds to ruling out $z^* \in (c, 1)$, established by the following Lemma.

Lemma C.5. Suppose $f(.)$ is continuous and single-peaked, and $F(c) > c$. Then, for any $z \in (c, 1)$, $F(z) > z$.

Proof. Suppose, to get a contradiction, that $F(z^*) - z^* \leq 0$ for some $z^* \in (c, 1)$. Combined with the fact that $F(c) - c > 0$, one can invoke the Mean Value Theorem to conclude:

There exists some $z' \in [c, z^*]$ such that $f(z') - 1 < 0$.

Similarly, combining $F(z^*) - z^* \leq 0$ with the fact that $F(1) = 1$, one can invoke the Mean Value Theorem once more to conclude:

There exists some $z'' \in [z^*, 1]$ such that $f(z'') - 1 \geq 0$.

Note that $f(.)$ is single-peaked. Given that $f(z'') > f(z')$, we conclude that the peak of $f(.)$ must lie to the right of $z''$. In particular, this implies that $f(.)$ will be increasing in $[0, z'']$, and consequently $F(.)$ must be convex and $\frac{F(z)}{z}$ must be increasing in this region. But note that we have $0 < c < z^* \leq z''$ while:
Lemma C.6 establishes that \( \frac{F(c)}{c} > \frac{F(z^*)}{z^*} \) for all \( z \in [c, 1) \), and therefore one cannot have \( z^* \in [c, 1) \) in the optimal solution. As a result, one cannot have a solution with \( c_1 < 1 \). We conclude that if the conditions in the first part of Proposition 3.2 are satisfied, one must have: \( z^* \in (0, c) \cup \{1\} \), which corresponds to: \( c_1 = 1 \) in the optimal solution. Tighter characterization of \( \sigma_0 \) is not possible, as for any \( x \in [0, c] \), one can come up with a single-peaked distribution which has \( \sigma_0 = x \) in the optimal solution. In particular, one can take \( F(.) \) as the step function, where the step is at \( q = \frac{\pi}{1+\pi} \in [0, c] \). Based on Kamenica and Gentzkow (2011), we know that the optimal solution must have \( \sigma_0 = \frac{q}{1-q} = x \).

2. Now, consider the second part of Proposition, where \( F(c) < c \). We begin by showing that in this case the optimal solution cannot be fully informative, i.e. we cannot have \( z^* \in \{0, 1\} \). This is also easy to see: because

\[
\frac{F(c)}{c} < 1 = \frac{F(1)}{1}
\]

one has \( G(c) < G(1) = G(0) \), so one cannot have \( z^* \in \{0, 1\} \). We continue by showing that one must have \( c_1 = 1 \) in the optimal solution, which again corresponds to ruling out \( z^* \in (c, 1) \). The following Lemma proves this.

\textbf{Lemma C.6.} Suppose \( f(.) \) is continuous and single-peaked, and \( F(c) < c \). Then, for any \( z \in (c, 1) \), \( \frac{F(z^*)}{z^*} > \frac{F(c)}{c} \).

\textbf{Proof.} Suppose, to get a contradiction, that \( \frac{F(z^*)}{z^*} < \frac{F(c)}{c} \) for some \( z^* \in (c, 1) \). Invoking the Mean Value Theorem three times, we have:

- There exists a \( z' \in [0, c] \) such that \( f(z') = \frac{F(c)}{c} \).
- There exists a \( z'' \in [c, z^*] \) such that \( f(z'') = \frac{F(z^*) - F(c)}{z^* - c} \).
- There exists a \( z''' \in [z^*, 1] \) such that \( f(z''') = \frac{1 - F(z^*)}{1 - z^*} \).

But note that, since \( \frac{F(z^*)}{z^*} < \frac{F(c)}{c} \):

\[
\frac{F(z^*) - F(c)}{z^* - c} < \frac{F(c)}{c}
\]

Similarly, since \( \frac{F(z^*)}{z^*} < \frac{F(c)}{c} < 1 \):

\footnote{To preserve the continuity of \( F(.) \), one can consider an arbitrarily close approximation to the step function.}
Putting everything together:

\[
\frac{1 - F(z^*)}{1 - z^*} > \frac{F(c)}{c}
\]

wheras \( z' < z'' < z''' \), which contradicts \( f(.) \) being single-peaked. \( \square \)

By Lemma C.6, one cannot have \( z^* \in (\frac{1}{2}, 1] \) in the optimal solution; therefore, one cannot have a solution with \( \sigma_1 < 1 \) (unless \( \sigma_0 = \sigma_1 \), in which case the politician is indifferent among any value of \( \sigma_0 \)).

In this case, tighter characterization of \( \sigma_0 \) is indeed possible, as \( \frac{F(c)}{c} \) constitutes a natural lower bound on the value of objective function. We will begin by showing that in equilibrium, this translates into a lower bound on \( z^* \).

**Lemma C.7.** Suppose \( f(.) \) is continuous and single-peaked, and \( F(c) < c \). Then \( z^* \in [\tilde{z}, c] \) where

\[
z = \frac{c - F(c)}{1 - F(c)}
\]

**Proof.** Lemma C.6 and the discussion preceding it already establishes that \( z^* \notin (c, 1] \). We will show that \( z^* \notin [0, \tilde{z}) \).

Take any \( z \in [0, \tilde{z}) \). All we need to show is that \( G(z) > G(c) \), i.e.

\[
\frac{1 - F(z)}{1 - z} < \frac{1 - F(c)}{1 - c}
\]

This easily follows from

\[
\frac{1 - F(z)}{1 - z} \leq \frac{1}{1 - z} < \frac{1}{1 - \tilde{z}} = \frac{1 - F(c)}{1 - c}
\]

where the first inequality holds because \( F(z) \geq 0 \), the second one holds because \( z < \tilde{z} \), and the equality holds by construction of \( \tilde{z} \). \( \square \)

Lemma C.7 imposes a lower bound on \( z^* \) in the region where \( z \in [0, c] \). Recall that this is the region where \( \sigma_1 = 1 \) and \( z = \frac{\sigma_0}{\sigma_0 + (1 - c)} \Leftrightarrow \sigma_0 = \frac{z(1 - c)}{c(1 - z)} \), so the lower bound on \( z^* \) implies a lower bound on \( \sigma^*_0 \). The lower bound turns out to be:

\[
\sigma^*_0 = \frac{z(1 - c)}{c(1 - \tilde{z})} = \frac{c - F(c)}{c}
\]
and the result follows.

\[ \square \]

### C.7 Proof of Proposition 3.3

**Proof.** The idea of this proof is quite similar to that of Proposition 3.2, so we will be brief and lay out the basics.

1. Begin with the case \( F(c) > c \). Because \( G(0) = G(1) < G(c) \), we conclude that the solution can never be fully uninformative.

Next, we rule out the case that \( z^* \in (0, c) \). Once again, repeatedly using Mean Value Theorem gives:

There exists \( z' \in [0, z^*] \), \( z'' \in [z^*, c] \), \( z''' \in [c, 1] \) such that
\[
\begin{align*}
f(z') &< f(z'') \\
&< f(z''').
\end{align*}
\]
But this contradicts \( f(.) \) being single-dipped.

Finally, the lower bound for \( \sigma_1 \) follows from combining two facts: (i) in the optimal solution \( z^* \in \{0\} \cup (c, 1] \), one needs to have \( \frac{F(z^*)}{z^*} \leq 1 \), and (ii) we must have: \( F(z^*) \geq F(c) \).

2. Now, consider the case \( F(c) < c \). Since \( G(c) < G(0) = G(1) \), one cannot have \( z^* \in (0, 1] \).

In order to rule out \( z^* \in (0, c) \), suppose the contrary, i.e. \( z^* \in (0, c) \). One can show that:

There exists \( z' \in [0, z^*] \) and \( z'' \in [z^*, c] \) such that \( f(z'') > f(z') \).

But the single-dippedness of \( f(.) \) then suggests that \( F(.) \) must be convex in \( [z^*, 1] \). But this implies that \( F(z) \) must hit the value of 1 at some \( z < 1 \), which contradicts \( f(.) \) being single-dipped.

\[ \square \]

### C.8 Proof of Proposition 3.4

**Proof. of Part (a).** By Proposition 3.2, the optimal solution will have \( \sigma_1^* = \bar{\sigma}_1^* = 1 \). So it is sufficient to consider the following program introduced in Equation (C.2):

\[
\min_{z \in [0,c]} 1 - \frac{1 - F(z)}{1 - z} (1 - c),
\]
where $z = \frac{\sigma_0}{\sigma_0 + 1 - c}$. This is equivalent to
\[
\max_{z \in [0,c]} \log(1 - F(z)) - \log(1 - z).
\]
Define
\[
\phi(z) = -\frac{f(z)}{1 - F(z)} + \frac{1}{1 - z}.
\]
The first-order condition requires that $\phi(z^*) = 0$, and the second-order condition requires that $\phi'(z^*) < 0$, so that $\phi(z^*)$ crosses 0 from above. By the MLRP, $\frac{f(z)}{1 - F(z)} < \frac{f(z)}{1 - F(z)}$, so that $\phi(z) > \tilde{\phi}(z)$ for all $z$. But since $\phi(z^*)$ and $\tilde{\phi}(z^*)$ cross 0 from above, this means that $z^* > \tilde{z}^*$ and therefore $\sigma_0^* > \tilde{\sigma}_0^*$.

**Proof of Part (b).** The argument is along similar lines as in the proof of part (a). By Proposition 3.3, at the optimum $\sigma_0^* = \delta_0^* = 0$ and so we can consider the following program:
\[
\min_{z \in [c,1]} \frac{F(z)}{z},
\]
where $z := \frac{c}{c + (1 - c)(1 - \sigma_1)}$. The program is equivalent to
\[
\min_{z \in [c,1]} \log(F(z)) - \log(z).
\]
Define
\[
\phi(z) = \frac{f(z)}{F(z)} - \frac{1}{z}.
\]
The first-order condition requires that $\phi(z^*) = 0$, and the second-order condition requires that $\phi'(z^*) > 0$, so that $\phi(z^*)$ crosses 0 from below. By the MLRP, $\frac{f(z)}{F(z)} > \frac{f(z)}{F(z)}$, so that $\phi(z) > \tilde{\phi}(z)$ for all $z$. But since $\phi(z^*)$ and $\tilde{\phi}(z^*)$ cross 0 from below, this means that $z^* < \tilde{z}^*$ and therefore $\sigma_1^* < \tilde{\sigma}_1^*$.

C.9 **Proof of Proposition 3.5**

*Proof.* By Lemma C.5, $\frac{S(z)}{z} > \frac{S(1)}{1} = 1$ for all $z \in (c,1)$. Therefore, $S(z) > z$ for all $z \in (c,1)$. This immediately implies that $F_{\alpha,\rho}(z) > z$ for all $z \in (c,1)$. So to find the optimal $\sigma_0$ and $\sigma_1$ when the distribution of the priors is given by $F_{\alpha,\rho}$ it is sufficient to consider the following program:
\[
\min_{z \in [0,c]} 1 - \frac{1 - F_{\alpha,\rho}(z)}{1 - z} (1 - c),
\]
where
\[
F_{\alpha,\rho}(z) = (\alpha S(z)^\rho + (1 - \alpha)z^{\alpha})^{\frac{1}{\rho}}.
\]
Define $T(z) = S(z)/z$. Then the program can be written as

$$\max_{z \in [0, c]} \frac{1 - z (\alpha T(z)^\rho + 1 - \alpha)^{1/\rho}}{1 - z},$$

The first-order condition is given by

$$z^* \alpha T'(z^*) T(z^*)^{\rho-1} (-1 + z^*) (\alpha T(z^*)^\rho + 1 - \alpha)^{(1-\rho)/\rho} - (\alpha T(z^*)^\rho + 1 - \alpha)^{1/\rho} + 1 = 0.$$  

This equation can be solved for $T'(z^*)$ to get

$$T'(z^*) = \frac{1 - (\alpha T(z^*)^\rho + 1 - \alpha)^{1/\rho}}{z^*(1 - z^*) \alpha T(z^*)^{\rho-1} (\alpha T(z^*)^\rho + 1 - \alpha)^{(\rho-1)/\rho}}$$

$$= \frac{\omega^{1/\rho} - \omega}{z^*(1 - z^*) \alpha T(z^*)^{\rho-1}}.$$

where $\omega = \alpha T(z^*)^\rho + 1 - \alpha \in [0, 1]$. Next note that the above program is equivalent to

$$\max_{z \in [0, 1]} \Omega(z, \alpha),$$

where $\Omega(z, \alpha) = \log(1 - z (\alpha T(z)^\rho + 1 - \alpha)^{1/\rho}) - \log(1 - z)$. Some simple algebra results in

$$\operatorname{sign} \left( \frac{\partial^2 \Omega(z^*, \alpha)}{\partial z \partial \alpha} \right) = \operatorname{sign} \left( -\alpha (z^* T'(z^*) + T(z^*)^2 + (\alpha + (z^* \omega^{1/\rho} - 1) T(z^*) + (z^* (2\alpha - 1) T(z^*) + T(z^*) (3\alpha - 1)) T(z^*)^\rho - T(z^*)^\rho (\alpha - 1)) \right).$$

Using the expression for $T'(z^*)$:

$$\operatorname{sign} \left( \frac{\partial^2 \Omega(z^*, \alpha)}{\partial z \partial \alpha} \right) = \operatorname{sign} \left( \frac{(\alpha T(z)^\rho + 1 - \alpha)^{1/\rho} \alpha T(z) - \alpha T(z)^\rho - \rho + \alpha}{\rho} \right)$$

$$= \begin{cases} 1 & \text{if } \rho < 1, \\ -1 & \text{if } \rho > 1, \\ 0 & \text{if } \rho = 1. \end{cases}$$

The first-order condition is given by $\Omega(z^*, \alpha) = 0$. By the implicit function theorem,

$$\frac{\partial z^*}{\partial \alpha} = -\frac{\partial \Omega(z^*, \alpha)}{\partial \alpha} / \frac{\partial \Omega(z^*, \alpha)}{\partial z}.$$
But the second-order condition is given by $\partial^2 \Omega(z^*, \alpha)/\partial z < 0$. Therefore, the sign of $\frac{\partial z^*}{\partial \alpha}$ is the same as the sign of $\frac{\partial^2 \Omega(z^*, \alpha)}{\partial \alpha}$. This proves that

$$
\text{sign} \left( \frac{\partial z^*}{\partial \alpha} \right) = \begin{cases} 
1 & \text{if } \rho < 1, \\
-1 & \text{if } \rho > 1, \\
0 & \text{if } \rho = 1,
\end{cases}
$$

and therefore

$$
\text{sign} \left( \frac{\partial \sigma^*_0}{\partial \alpha} \right) = \begin{cases} 
1 & \text{if } \rho < 1, \\
-1 & \text{if } \rho > 1, \\
0 & \text{if } \rho = 1.
\end{cases}
$$

\[ \Box \]

### C.10 Proof of Proposition 3.6

**Proof.** Using the same arguments as in the proof of Lemma 3.3 and Proposition 3.2, one can demonstrate that the optimal solution has: $\sigma^*_1 = 1$ and $\sigma^*_0 \in [0, 1]$. Therefore, in the optimal solution:

$$
\tilde{p}(\sigma_0, \kappa) = 1 - \frac{\kappa}{c(1 - \sigma_0)}
$$

$$
p(\sigma_0, \kappa) = \frac{c\sigma_0 + \kappa}{c\sigma_0 + (1 - c)}
$$

Given $\kappa$, the politician's problem then is reduced to the following program:

$$
\max_{\sigma_0 \in [0, 1]} \Pi(\sigma_0, \kappa)
$$

where

$$
\Pi(\sigma_0, \kappa) = -F(\tilde{p}(\sigma_0, \kappa))c(1 - \sigma_0) - F(p(\sigma_0, \kappa))(c\sigma_0 + (1 - c))
$$

The first-order condition is:

$$
\frac{\partial \Pi(\sigma_0, \kappa)}{\partial \sigma_0} = c \left( f(\tilde{p}(\sigma_0, \kappa))(1 - \tilde{p}(\sigma_0, \kappa)) - f(p(\sigma_0, \kappa))(1 - p(\sigma_0, \kappa)) + \tilde{p}(\sigma_0, \kappa) - p(\sigma_0, \kappa) \right) = 0
$$

Let

$$
\Lambda(\sigma_0, \kappa) := c \left( f(\tilde{p}(\sigma_0, \kappa))(1 - \tilde{p}(\sigma_0, \kappa)) - f(p(\sigma_0, \kappa))(1 - p(\sigma_0, \kappa)) + \tilde{p}(\sigma_0, \kappa) - p(\sigma_0, \kappa) \right)
$$

The first-order condition then is: $\Lambda(\sigma^*_0, \kappa) = 0$. By the Implicit Function Theorem,

$$
\frac{\partial \sigma^*_0}{\partial \kappa} = -\frac{\partial \Lambda(\sigma^*_0, \kappa)/\partial \kappa}{\partial \Lambda(\sigma^*_0, \kappa)/\partial \sigma_0}
$$

145
But the second-order condition is that \( \frac{\partial \Lambda(\sigma^*_0, \kappa)}{\partial \sigma_0} < 0 \). Therefore, the sign of \( \frac{\partial \sigma^*_0}{\partial \kappa} \) is the same as the sign of \( \frac{\partial \Lambda(\sigma^*_0, \kappa)}{\partial \kappa} \). Simple algebra gives:

\[
\frac{\partial \Lambda(\sigma^*_0, \kappa)}{\partial \kappa} = c \left( f'(\bar{p}(\sigma^*_0, \kappa)) \frac{\partial \bar{p}(\sigma^*_0, \kappa)}{\partial \kappa} (1 - \bar{p}(\sigma^*_0, \kappa)) - \frac{f'(p(\sigma^*_0, \kappa))}{\partial \kappa} \frac{\partial p(\sigma^*_0, \kappa)}{\partial \kappa} (1 - p(\sigma^*_0, \kappa)) \right)
\]

Note that when \( \kappa = 0 \), \( \bar{p}(\sigma^*_0, \kappa) = 1 \) and \( \frac{\partial \Lambda(\sigma^*_0, \kappa)}{\partial \kappa} < 0 \). By Berge’s maximum theorem, \( \bar{p}(\sigma^*_0, \kappa) \) is continuous in \( \kappa \). As a result, there exists some \( \kappa > 0 \) such that, for \( \kappa \in (0, \bar{\kappa}) \), \( \frac{\partial \Lambda(\sigma^*_0, \kappa)}{\partial \kappa} < 0 \). The result follows. \( \square \)
Bibliography


153

