INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2020IPPAH001

Thèse de doctorat

Essays on Game Theory, Mechanism Design, and Financial Economics

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à HEC Paris

École doctorale de l'Institut Polytechnique de Paris – (ED IP Paris) n°626
Spécialité de doctorat: Sciences de Gestion

Thèse présentée et soutenue à Jouy-en-Josas, le Date, par 04/03/2020

RIVERA Thomas

Composition du Jury :

Nicolas VIEILLE Professor, HEC Paris, E&DS	Président
Matthieu BOUVARD Professor, Toulouse School of Economics, Finance	Rapporteur
Eduardo PEREZ-RICHET Associate Professor, Sciences Po, Economics	Rapporteur
Marco SCARSINI Professor, LUISS Guido Carli, Economics and Finance	Examinateur
Tristan TOMALA Professor, HEC Paris, E&DS	Directeur de thèse
Jean-Edouard COLLIARD Associate Professor, HEC Paris, Finance	Co-Directeur de thèse

Essays on Game Theory, Mechanism Design, and Financial Economics

PhD Thesis: Advised by Tristan Tomala & Jean-Edouard Colliard

Thomas J. Rivera
HEC Paris

March 25, 2020

Abstract

This thesis develops and utilizes tools in game theory and mechanism design to study multiple applications in economics and finance. The first chapter studies the problem of implementing communication equilibria of strategic games when players communicate with an impartial mediator through a network. I characterize necessary and sufficient conditions on the network structure such that any communication equilibrium of any game can be implemented on that network. The next chapter studies a model of supply chain congestion whereby capacity constraints lead to very inefficient Nash equilibria and I show how the use of correlated equilibria can substantially resolve those inefficiencies. The final two chapters study related issues in the design of bank capital requirements. In Chapter 3, I characterize optimal bank capital requirements when banks have private information about the value of their existing assets. I show how the implementation of capital requirements can eliminate the bank's cost of raising capital by revealing their information to the market and conditions under which doing so is optimal. In Chapter 4, I show how when the bank's private information is about the riskiness of its assets instead, then any risk sensitive capital requirement will lead banks to optimally misreport their risk whenever investors are sufficiently risk averse, highlighting important robustness concerns.

Acknowledgements

First and foremost, I would like to thank my family for supporting me throughout the PhD. To my parents Israel and Catherine, for their moral and financial support and for being so understanding and never doubting my determination to achieve this goal. To my sister for inspiring me to always strive for the best and for always being someone I could talk to. To my Aunt Linda and Uncle Richie for their support, encouragement, and enlightened views. To Salomé, for your love, support, drives to the airport, and motivation. And finally to the rest of my family with whom I had spent time with at all of our events and celebrations throughout these years, thank you for all of your love and support.

Next I would like to thank all of the professors that had helped throughout the years to make this thesis what it is. To my advisor Tristan, thank you for all of your support, patience, and for all of the great conversations we had. I have learned an enormous amount about research, academia, and life in general, from our interactions. I couldn't have asked for a better nor more supportive advisor and cannot thank you enough for all of the time and effort you spent on me. To my co-advisor Jean-Edouard, I also thank you for your support and patience and for helping me to see more clearly the big picture in finance and banking. I met you at a crucial time and am extremely grateful for all of our conversations and all of your great advice. To the rest of the E&DS department and all of you that helped me mature as an academic with all of your advice I am extremely grateful to have had your support. In particular, a special thanks to Eric for your advice on the job market and writing, to Raphael for taking the time to listen to my projects and pushing me to improve them, to Tomasz for helping me see the big picture to Stefania for guiding me through our many years of teaching together, and to Marco for hosting me in Rome.

I would further like to thank all of the PhD students that I have met and shared research and PhD experiences with. In particular, to Bruno, Gaetan, and Xavier, who I looked up to and took me under their wing into the Parisian game theory community. To the younger E&DS PhD cohort, Fan, Veronica, Emilien, Wei, and Atulya I was happy to share my experiences with you, which was very cathartic for me, and I hope that I was able to successfully pass along some of the advice that had been passed down to me throughout the years. A special thanks as well to the rest of the HEC PhD cohort with whom I had spent many hours discussing PhD life and all of its ups and downs.

I would also like to thank all of my friends. To those who were there long before the PhD, Andrew, Frank, Ricky, Natalya, and Tim, our time hanging out together on my short trips home constitutes some of my fondest memories over the past 6 years. Especially to Frank for the countless drives to and from the airport every time I would come home, I can't thank you enough for being such a great friend. To those friends that I have met during my

studies, in particular to Maddie, Flora, Nik, Rudy, Emma, Andrej, for sharing and creating all of my favorite memories over nights out, dinners, and general trolling around Paris and Rome.

Finally, I would like to thank Matthieu Bouvard and Eduardo Perez-Richet for agreeing to be part of my thesis jury and for all of their questions and comments regarding my research.

Contents

0 Introduction **7**

0.1 Incentives and the Structure of Communication 12

0.2 Strategic Inventory Management in Capacity Constrained Supply Chains . . 15

0.3 Bank Regulation, Investment, and the Implementation of Capital Requirements 17

0.4 Robust Regulation of Bank Risk: Reporting and Risk Aversion 21

1 Incentives and the Structure of Communication **26**

1.1 Introduction 26

1.1.1 Related Literature 29

1.2 Preliminaries 34

1.2.1 Implementation 34

1.3 An Illustration of the Results: The 3-player Case 37

1.3.1 Sufficiency 38

1.3.2 Necessity of Secrecy and Resiliency 40

1.4 Main Results 43

1.5 Incomplete Information 48

1.6 Applications 51

1.7 Comments and Conclusion 53

1.8 Appendix 55

1.8.1 Cheap Talk Protocols 55

1.8.1.1 Correlated Equilibrium Protocols 55

1.8.1.2 Communication Equilibrium Protocols 57

1.8.2 Proof of Theorem 1: Sufficiency 57

1.8.3 Proof of Proposition 1 63

1.8.4 Proof of Corollary 1 75

1.8.5 Proof of Corollary 2 75

1.8.6 Proof of Theorem 2 76

1.8.7 Proof of Corollary 3 78

1.8.8 Proof of Corollary 4 80

1.8.9 Proof of Lemma 1 80

2 Strategic Inventory Management in Capacity Constrained Supply Chains **86**

2.1 Introduction 86

2.1.1 Related Literature 88

2.2 Model . . . 90
2.3 Simple Case . . . 94
2.3.1 Prices . . . 98
2.4 General Results . . . 99
2.4.1 Social Optimum . . . 100
2.4.2 Worst Nash Equilibrium Cost . . . 100
2.4.3 Best Nash Equilibrium Cost . . . 101
2.4.4 Premiums . . . 102
2.5 Correlated equilibria . . . 102
2.5.0.1 Correlated equilibria. . . . 102
2.5.0.2 Outcomes. . . . 103
2.6 Appendix . . . 107
2.6.1 Proofs of Section 4: . . . 107
2.6.1.1 Proof of Lemma 2.7 . . . 107
2.6.1.2 Proof of Lemma 2.8 . . . 108
2.6.1.3 Proof of Lemma 2.9 . . . 108
2.6.1.4 Proof of Theorem 2.10 . . . 109
2.6.1.5 Proof of Theorem 2.11 . . . 109
2.6.1.6 Proof of Theorem 2.12 . . . 110
2.6.1.7 Proof of Lemma 2.13 . . . 112
2.6.1.8 Proof of Lemma 2.15 . . . 112
2.6.1.9 Proof of Lemma 2.16 . . . 113
2.6.1.10 Proof of Lemma 2.17 . . . 114
2.6.1.11 Proof of Theorem 2.18 . . . 114
2.6.1.12 Proof of Theorem 2.20 . . . 118
2.6.1.13 Proof of Theorem 2.21 . . . 120

3 Bank Regulation, Investment, and the Implementation of Capital Requirements 124

3.1 Introduction . . . 124
3.2 A Model of Capital Regulation Under Asymmetric Information . . . 132
3.2.1 Baseline Model . . . 132
3.2.2 Capital Securities . . . 132
3.2.3 The Regulatory Environment . . . 133
3.2.4 Welfare . . . 136
3.2.5 The Capital Raising Game $\Gamma(\mathcal{M})$ . . . 138

3.2.6 Equilibrium Concept and Refinements . . . 140
3.3 Preliminary Results and Equilibria of the Capital Raising Game . . . 142
3.3.1 Preliminary Results . . . 142
3.3.2 Equilibria of Pooling Mechanisms . . . 144
3.3.3 Equilibria of Separating Mechanisms . . . 145
3.4 Optimal Mechanisms . . . 146
3.4.1 Optimal Pooling Mechanisms . . . 146
3.4.2 Optimal Separating Mechanisms . . . 148
3.5 Comparison of Optimal Mechanisms . . . 151
3.6 Policy Implications . . . 153
3.7 Conclusion . . . 156
3.8 Extensions . . . 159
3.8.1 Continuum of Types . . . 159
3.8.1.1 Pooling . . . 160
3.8.1.2 Separating . . . 161
3.8.1.3 Extension of Proposition 4.4 to the Continuum Case . . . 161
3.9 Appendix . . . 162
3.1.1 Proofs of Section 2 . . . 162
3.1.1.1 Proof of Lemma 3.8 . . . 162
3.1.1.2 Proof of Lemma 3.9 . . . 163
3.1.2 Proofs of Section 3 . . . 164
3.1.2.1 Proof of Lemma 3.10 . . . 164
3.1.2.2 Proof of Lemma 3.11 . . . 165
3.1.2.3 Proof of Lemma 3.12 . . . 167
3.1.3 Proofs of Section 4 . . . 167
3.1.3.1 Proof of Lemma 3.14 . . . 167
3.1.3.2 Proof of Proposition 3.13 . . . 167
3.1.3.3 Proof of Lemma 3.15 . . . 169
3.1.3.4 Proof of Proposition 3.16 . . . 169
3.1.3.5 Proof of Corollary 3.17 . . . 171
3.1.3.6 Proof of Lemma 3.18 . . . 172
3.1.3.7 Proof of Lemma 3.19 . . . 173
3.1.3.8 Proof of Proposition 3.20 . . . 175
3.1.3.9 Proof of Lemma 3.21 . . . 176
3.1.4 Proofs of Section 5 . . . 177
3.1.4.1 Proof of Proposition 4.4 . . . 177

3.1.4.2 Proof of Proposition 3.24 . . . 180
3.1.4.3 Proof of Corollary 3.25 . . . 181

4 Robust Regulation of Bank Risk: Reporting and Risk Aversion 182
4.1 Literature Review . . . 186
4.2 Model . . . 189
4.3 Main Results . . . 195
4.4 Robust Mechanisms to Changes in Risk Aversion . . . 198
4.4.1 Controlled Risk Taking . . . 199
4.5 Appendix . . . 202
4.5.1 Reformulating the bank's problem . . . 202
4.5.2 Lemma 4.8 and Lemma 4.9 . . . 203
4.5.3 Proof of Proposition 4.4 . . . 205
4.6 Proof of Proposition 4.6 . . . 207
4.7 Proof of Corollary 4.7 . . . 208

0 Introduction

The key focus of the study of game theory is in modeling strategic interactions— which we refer to as *games*— and predicting their outcomes. A *Nash equilibrium* of any game is a profile of strategies (one for each player of the game) whereby no player can profitably deviate by choosing a different strategy when keeping fixed the strategies of the other players. These Nash equilibria encompass an intrinsic stability condition so that, at the very least, we should never expect an outcome of a game to rationally lay outside of the set of Nash equilibria. Furthermore, every game (satisfying some standard conditions) is guaranteed to have at least one Nash equilibrium. It is for this reason that this solution concept is heavily utilized in economics as a way of predicting the outcomes of strategic situations, necessary for analysis in many applications. Mechanism design, sometimes referred to as *reverse/inverse game theory* on the other hand makes an attempt to alter the underlying strategic interaction by utilizing different instruments (e.g. taxes, contracts, etc.) that provide incentives to the players of the game. These incentives alter the profitability of any particular strategy and therefore the underlying set of Nash equilibria. This allows policy makers to model and study the rational reaction of strategic agents to their policies, helping to guide and understand the underlying economic environment.

This thesis is the culmination of my PhD studies focused on game theory and mechanism design and how tools from these fields can be utilized to generate insights in various

applications. One particular application of interest has been the regulation of financial institutions, particularly related to prudential regulation of privately informed banks via capital requirements. The chapters are presented in the order in which they were originated and while the last two chapters study a common theme, the first and second chapters are independent projects on relatively unrelated topics. In this introduction I will present the main models and results of each of these chapters, noting that each chapter is a self contained paper. Further, given the varying topics of study, all literature reviews are done within their respective chapters.

The first chapter of this thesis, published in the Journal of Economic Theory, studies two more general versions of the Nash equilibrium solution concept called *correlated* and *communication* equilibria defined for games of complete and incomplete information respectively. Generally speaking, the set of correlated equilibria consists of all equilibrium outcomes that could be generated when augmenting the original game with a pre-play communication phase whereby players receive correlated signals/recommendations from an impartial mediator.[1] In games of incomplete information, whereby players have private information, communication equilibria are the set of equilibrium outcomes under which first players report their private information to an impartial mediator, who then draws (potentially correlated) recommendations from a distribution contingent on the information received. The set of correlated and communication equilibria drastically expand the set of achievable outcomes when compared to their counterparts for games of complete (Nash equilibria) and incomplete information (Bayesian equilibria). Furthermore, they are much simpler to compute when compared to Nash equilibria making them ideal candidates for certain applications.

The particular application I have in mind for the first chapter of this thesis is that of an organization consisting of a principal (who acts as the mediator) and multiple agents with differing incentives (given by the underlying game). In this scenario, the multiple agents gather and report (private) information to the principal who then delegates tasks to the agents based on the collective information received. The main focus of the paper is to study the natural case whereby this type of communication of information and tasks occurs through a network (e.g. in a hierarchical organization). In this case, misaligned incentives of the agents can restrict the set of achievable (communication) equilibrium outcomes by generating incentives to miscommunicate during the communication phase. The main question I ask in this paper is: what are the conditions on the structure of the network of communication such that for any game and any correlated (communication) equilibrium of that game there exists a way

[1] In what follows when I utilize the term *equilibrium outcome* I am referring to an outcome whereby no player can profitably deviate by choosing another strategy given their information at that time, and keeping the other players strategies fixed.

to communicate on that network which results in an equilibrium outcome identical to the correlated (communication) equilibrium in question? Informally, this is equivalent to generating conditions on the communication network such that any correlated (communication) equilibrium of any (in)complete information game can be *implemented* on that network. The main result (described below) generates necessary and sufficient conditions on the network structure to achieve this goal.

In the second chapter of this thesis coauthored with Marco Scarsini and Tristan Tomala, we study an issue of inventory management in a supply chain with a single supplier and multiple retailers who must each decide when to order their inventory before a deadline. We show how when the supplier is capacity constrained and the cost of missing the deadline is large, then all Nash equilibria of the resulting inventory ordering game result in highly inefficient inventory costs with respect to the social optimum (i.e. the outcome that minimizes the sum of total expected costs). This is due to the fact that when one player deviates and orders their inventory one period later, the players affected most by this deviation are the players who order the latest (rather than the player who deviated). Hence, the incentive to minimize inventory costs (and therefore order as late as possible) imposes externalities on the players ordering at the latest times, essentially forcing an equilibrium condition whereby all players order very early.

We then show how these inefficient equilibrium can be improved upon by utilizing dynamic prices and that when the penalty cost is large, the use of the correlated equilibrium solution concept can almost entirely reduce the inefficiency. The idea here is that while prices can help to alter the Nash equilibrium outcomes, they impose a cost on the players and therefore have limited scope in reducing the sum of total costs. Correlated equilibria on the other hand can correlate the players ordering times in such a way that it can approximate the social optimum (which has players spread out their orders over the latest order times subject to never exceeding capacity). The way that this is done is to put small probability on special outcomes that have the property that any player told to order at time t under special outcome t is late with positive probability when ordering one period later (and late for sure when ordering two or more periods later). These outcomes therefore *enforce* the order time t in a sense that they provide incentives for any player to never deviate to an earlier time when told to depart at time t if there is sufficient probability on special outcome t. We then show how the optimal correlated equilibrium only randomizes over these special outcomes and the social optimum outcome. Further, as the penalty cost for missing the deadline increases, the probability that the optimal correlated equilibrium puts on the special outcomes goes to zero, and therefore the probability it puts on the social optimum goes to one. Hence, we say this correlated equilibrium approximates the social optimum, importantly without the use

of additional costs (i.e. prices) and ensuring that no player is ever late in equilibrium.

The final two chapters of this thesis study the issue of regulating a bank who has private information about the value (Chapter 3) and riskiness (Chapter 4) of their existing assets. The main form of regulation considered is capital regulation and transfers (loosely interpreted as deposit insurance premia or the cost of closer regulatory inspection). In the third chapter, my Job Market Paper, I study how the bank's private information about the value of their existing assets leads to an adverse selection problem whereby higher capital requirements lead to socially costly underinvestment. This happens due to the fact that when banks have good news about the value of their existing assets, and this news is private information, then their shares will be underpriced by the market. In particular, the (uninformed) market will optimally price the bank's shares as if its existing assets were of average quality (with respect to their prior beliefs). Once the bank's shares are underpriced in this way, then the bank faces a tradeoff when it must raise new capital in order to meet capital requirements on a new investment: the new positive net present value (NPV) investment increases the value of existing equity but raising new equity imposes a cost as it requires giving away more of the existing assets (due to the underpricing of equity) than would be the case if markets were perfectly informed. Therefore, once the cost of raising new equity— which is strictly increasing in the capital requirement — outweighs the NPV of the new project, the bank will optimally forgo investment. Furthermore, when these investments are time sensitive or relationship specific (as we assume), then underinvestment will be socially costly as the NPV will be lost providing a trade off for the regulator: higher capital requirements reduce the expected spillover costs of bank failure but induce a higher level of underinvestment.

The main insight of this chapter is to show how the implementation of capital requirements can resolve the underinvestment problem by revealing the bank's information to the market, eliminating the adverse selection problem and thus the cost of raising capital. In particular, I show how the regulator can design a menu consisting of different levels of capital requirements and transfers with the property that it is optimal for the banks with good (bad) news to choose the good (bad) menu option. Once this is the case, the market will correctly learn the bank's type after observing which menu option it has chosen and therefore will correctly price the bank's equity— eliminating any incentives to forgo investment. Important to note though, is that while the regulator resolves the underinvestment problem by utilizing such a *revealing* mechanism, they face another cost when doing so. Namely, the mechanism must provide incentives to the banks to truthfully reveal their types and therefore must pay *information rents* in the form of lower capital requirements. In line with this fact, we then show how information revelation may not always be the optimal form of regulation given the underlying parameters of the model. The main result shows how *pooling* the banks together

and providing investment incentives is optimal whenever the banking sector is strong (i.e. has many good types) whereas utilizing information revelation and providing incentives to reveal that information is preferred when the strength of the banking sector is intermediate. Finally, when the banking sector is very weak, the regulator optimally pools the banks with a single capital requirement which is then set very high leading to a recapitalization of the many weak banks in exchange for underinvestment by the strong banks.

In the last chapter of this thesis, a working project, I study the case whereby a bank has private information about the riskiness of its existing assets and what regulators can do to induce the bank to reveal that information in order to charge them the appropriate risk sensitive capital requirements. As explained in Chapter 3, information revelation has the benefit reducing underinvestment which I take as given in this chapter (i.e. rather than modeling it endogenously). The question I ask is how to induce bank's to reveal their true level of risk, understanding the tradeoff faced when the level of risk aversion of investors (and shareholders) may vary over time. The key trade-off comes from the fact that, from societies perspective, riskier banks should optimally finance themselves with larger amounts of equity. On the other hand, when banks are financed with insured deposits, they privately prefer to raise as little equity as possible. I then show how revealing to the market that a bank is riskier than expected leads to two opposing effects on the bank's cost of raising capital. First, assuming riskier bank's have the same expected returns but a higher variance on those returns, then revealing this information to the market will actually decrease the bank's cost of raising equity when the market is risk neutral. This comes from the fact that depositors do not price this risk and therefore increased variance leads bank shareholders to absorb the large gains (which are now more likely) but avoid the large losses which fall on the deposit insurance fund. The second effect only exists when the market is risk averse, in which case when the bank reveals that they are riskier, this will lead to an increase in the bank's cost of raising equity as the market will charge them a higher risk premium. Naturally, when the level of risk aversion is high this latter effect offsets the former and riskier bank's no longer wish to reveal that information to the market. Note that this does not necessarily imply that the regulator cannot induce the bank to reveal this information, only that the benefit of doing so shifts from the more risky banks to the less risky banks.

The main result of the paper is to show how this incentive for information revelation, and its ability to diminish with the level of market risk aversion, generates an important robustness issue for regulators. Namely, I show that for any *risk sensitive* capital requirement that maps more risk to higher capital requirements, that capital requirement can be made incentive compatible only if the level of risk aversion is sufficiently high or sufficiently low. This is an important insight as it has become very clear to researchers that the level of

investor risk aversion does vary over time (see e.g. Cochrane 2017) and typically is counter cyclical: investors are less risk averse during booms and more risk averse during recessions. An important note to make here is that while researchers understand that risk aversion is time varying, there is not a clear answer as to what drives this variation. This can create substantial issues whereby an incentive compatible risk sensitive capital requirement properly functions and maximizes welfare when investors share a low aversion to risk (e.g. in booms) but then has a break down of incentives whereby bank's hold lower levels of capital than they should when the level of risk aversion increases (e.g. after some economic shock) and vice-versa. For this reason, I then study the issue of providing incentives that are robust to small perturbations in the regulators estimated level of risk aversion. Namely, given an initial measure of risk attitude, under what conditions can the regulator can design a mechanism that remains incentive compatible for small perturbations of risk aversion (i.e. slightly more or less risk aversion) around that initial value. In this case, and for connected reasoning to the first results, I show how this is not possible for any level of perturbation unless the initial level of risk aversion is either sufficiently low or sufficiently high. Finally, I show how the results remain to be true when bank's can alter their level of risk, which will become optimal whenever the level of risk aversion is large.

The remainder of this introduction will provide further details regarding the modelling and results of the four chapters of this thesis.

0.1 Incentives and the Structure of Communication

As mentioned above, this paper studies the *implementation* of correlated/communication equilibria on a communication network. In this introductory section we will only discuss the issue of implementing correlated equilibria.[2] For any game $\Gamma = (I, (S_i)_{i\in I}, (u_i)_{i\in I})$ with a finite number of players $i \in I$ and strategy sets S_i, a correlated equilibrium Q of Γ is a distribution over strategy profiles $S := \Pi_{i\in I} S_i$ such that

$$\sum_{s_{-i}\in S_{-i}} u_i(s_i, s_{-i})Q(s_{-i}|s_i) \geq \sum_{s_{-i}\in S_{-i}} u_i(s_i', s_{-i})Q(s_{-i}|s_i)$$

for all $s_i, s_i' \in S_i$ and all $i \in I$. In other words, Q is a correlated equilibrium if conditional on being told to play strategy s_i, Player i finds it optimal to play that strategy, knowing that the other players' strategies are distributed according to $Q(s_{-i}|s_i) := \frac{Q(s_i,s_{-i})}{\sum_{s_{-i}} Q(s_i,s_{-i})}$.

To illustrate this solution concept, consider the 2-player game of Figure 1. There is a

[2]The result for communication equilibria is a simple extension of the correlated equilibria result so that not much is lost in focusing on this case.

	A	B
a	6, 6	2, 7
b	7, 2	0, 0

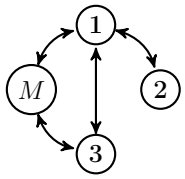


Figure 1: The game Γ_1 and the network $\mathcal{N}_1$.

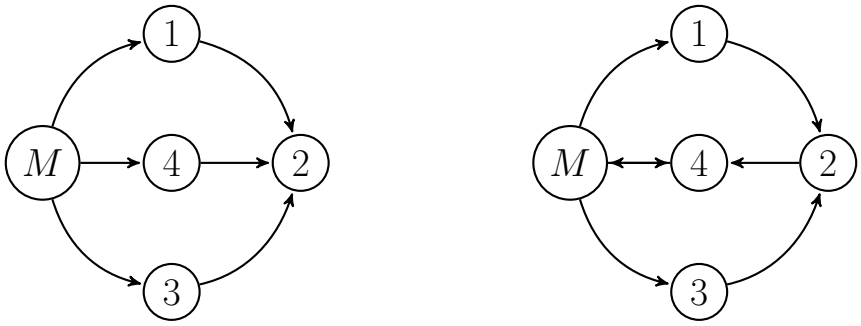
well known correlated equilibrium Q_1 of this game whereby players play the action profiles (a, A), (b, A), and (a, B) each with equal probability $\frac{1}{3}$. The *canonical* way to implement this correlated equilibrium is to have an impartial mediator draw a strategy profile according to Q_1 and then to report to each player $i \in \{1, 2\}$ their component of the realized strategy profile s_i. It is easy to check that under such a *communication protocol* no player can profit by deviating to some strategy s_i' different from the strategy suggested to them. For example, whenever Player 1 is suggested to play b, then it is always optimal for them to play b. This is due to the fact that in this case, Player 1 knows that the strategy profile drawn was (b, A) given that $Q_1(A|b) = 1$ and this strategy profile is in fact a pure Nash equilibrium of this game. Similarly, if Player 1 is suggested to play a, then that player only knows that Player 2 will play A with probability $\frac{1}{2}$ and B with probability $\frac{1}{2}$: $Q_1(A|a) = Q_1(B|a) = \frac{1}{2}$ which makes playing a more profitable than b. It can be checked that, for similar reasons, Player 2 always has the incentive to follow her suggested strategy as well.

Note that in implementing Q_1 it is assumed that the impartial mediator can send recommendations directly to both players 1 and 2. In order to illustrate the aim of this paper, suppose that we wished instead to *implement* the same correlated equilibrium Q_1 on the network $\mathcal{N}_1$ of Figure 1 whereby the mediator can communicate directly with Player 1 and an auxiliary Player 3 but not directly with Player 2. Namely, consider the game $(\Gamma, \mathcal{N}_1)$ which consists of the game Γ extended by an arbitrarily long but finite cheap talk phase restricted to the network $\mathcal{N}_1$ (i.e. players only communicate with their neighbors). Then this implementation question translates to: does there exist a *communication strategy* ρ whereby players only communicate to their neighbors on network $\mathcal{N}_1$ for a finite period of time and then play some strategy $\hat{\sigma}$ dependent on the history of communication satisfying the condition that (1) $(\rho, \hat{\sigma})$ is a perfect Bayesian equilibrium of $(\Gamma, \mathcal{N}_1)$ and (2) $\mathbb{P}_\rho(\hat{\sigma} = s) = Q_1(s)$ for all $s \in S$.

Denote by $B(\Gamma, \mathcal{N})$ the set of all perfect Bayesian equilibria $(\rho, \hat{\sigma})$ of the game $(\Gamma, \mathcal{N})$ and by $C(\Gamma)$ the set of all correlated equilibria of Γ. Then more generally the question this paper looks to answer is: what are the necessary and sufficient conditions on the network $\mathcal{N}$ such that $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for all games Γ. Before describing these conditions, note that network

$\mathcal{N}_1$ does not satisfy our conditions as Q_1 cannot be implemented on $\mathcal{N}_1$: for any $(\rho, \hat{\sigma})$ such that $\mathbb{P}_\rho(\hat{\sigma} = s) = Q_1(s)$ for all $s \in S$, $(\rho, \hat{\sigma}) \notin B(\Gamma, \mathcal{N})$. The reasoning highlights one of the necessary conditions that any network must satisfy to guarantee $B(\Gamma, \mathcal{N}) = C(\Gamma)$: for any player that the mediator cannot directly send messages to in the network $\mathcal{N}$, there must be a way for the mediator to send messages *secretly* to that player. The reason this condition is necessary is illustrated by the Q_1 and $\mathcal{N}_1$ example. Namely, no matter how sophisticated the communication strategy $(\rho, \hat{\sigma})$ is, all messages received by Player 2 are sent through Player 1. Further, Player 1 knows $(\rho_2, \hat{\sigma}_2)$ (i.e. has correct equilibrium beliefs) and therefore can back out Player 2's strategy suggestion given any history of communication. Hence, Player 1 will have an optimal deviation to play b whenever he is told to play a and Player 2 is told to play A (which he learns with probability 1 whenever Player 2 correctly learns their suggested strategy) and thus Q_1 cannot be implemented on $\mathcal{N}_1$.

The additional necessary condition for $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for all games Γ is that for any Player $i \in I$ that the mediator cannot directly send messages to, the network must be such that the mediator can send messages to Player i that are correctly received with probability 1 even if any other player *deviates* by playing any communication strategy during the communication phase. Section 3.2 of Chapter 1 illustrates why this condition is necessary.



(a) *A network satisfying Condition* (1). (b) *A network satisfying Condition* (2).

Figure 2: *4-player networks satisfying Theorem 1.*

The main result of the paper states that $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for all games Γ if and only if for all players $i \in I$ that the mediator (M) cannot directly send messages to, either (1) $\mathcal{N}$ is strongly 3-connected from M to i or (2) $\mathcal{N}$ is strongly 2-connected from M to i and strongly 1-connected from i to M with all three connecting paths disjoint. Note that a directed network is strongly k-connected from node v_1 to node v_2 iff there exists k (vertex) disjoint paths from v_1 to v_2. Conditions (1) and (2) are illustrated in Figure 2 in the 4-player case and an illustration of a communication protocol that can implement any correlated equilibrium on any network satisfying these conditions is presented in Section 3.1 of Chapter 1. As can be seen in this figure, the only player that the mediator cannot directly communicate to

in either network is Player 2. Further, in Figure 2 (a) there are 3 disjoint paths from the mediator to player 2 while in Figure 2 (b) there are 2 disjoint paths from the mediator to player 2 and one path from player 2 to the mediator, with all three being disjoint. Further, this illustration could include any number of players on the paths connecting the mediator to Player 2 provided that the network satisfy one of these two conditions for each of those players as well.

The paper then goes on to characterize similar conditions for the implementation of all communication equilibria of all Bayesian games on a network. Finally, the paper concludes with some potential applications of the results, one of which is the communication of bank risks between local regulators and a central supranational regulator.

0.2 Strategic Inventory Management in Capacity Constrained Supply Chains

In this paper we study a problem whereby multiple retailers must source their inventory before a deadline from the same wholesaler. In a discrete context, retailers choose what integer time period to order their inventory, facing an inventory cost g per period for holding the inventory before the deadline $t^\star$ and a large penalty cost C if their inventory is not received before the deadline (e.g. lost revenue). In such a scenario, if the wholesaler had unlimited capacity and the delivery time was β, then all retailers would order their inventory exactly β periods before the deadline at time $t^\star - \beta$. If instead the wholesaler could only serve γ retailers, then this would creates strategic tensions between the retailers: ordering later saves on the inventory cost, but if all retailers order late then there will be congestion and some orders will not arrive on time imposing the large penalty cost. In particular, denoting by I the number of retailers, then a capacity $\gamma < I$ implies that if all retailers ordered β periods before the deadline, then only $\lfloor \frac{I}{\gamma} \rfloor$ retailers will receive their orders in time with the remaining $I - \lfloor \frac{I}{\gamma} \rfloor$ receiving them at least one period late. In order to clarify the priority structure, we assume that whenever $k > \gamma$ retailers order at the same time and there is no backlog, then the wholesaler chooses uniformly at random $\lfloor \frac{k}{\gamma} \rfloor$ of the k retailers to serve and forms a backlog with the remaining $k - \lfloor \frac{k}{\gamma} \rfloor$ retailers who then receive priority over any orders made at any later time. Similarly, if there is a backlog at the time when the k retailers order then those with priority are served first (first come first served) and then the remaining capacity is filled using the uniform random assignment.

In order to present the results of the paper, we will look at the (notationally) simple case whereby $\gamma = \beta = g = 1$. Our first result is to show that whenever the penalty cost $C > I^2$, then the worst (i.e. highest cost) Nash equilibrium of this game generates the worst

individually rational payoff for each player.[3] Namely, in this Nash equilibrium each player pays the same cost as if they ordered at the *safe time* $t^\star - I$: given that the wholesaler can serve one retailer per period, $t^\star - I$ is the latest time period such that any retailer ordering at this time can guarantee delivery independent of the strategies of the remaining players. In this sense, no rational retailer will ever pay more than $g \cdot I = I$ in expectation as otherwise they could optimally order at time $t^\star - I$ and pay exactly I. The first result states that in the worst Nash equilibrium, all players pay I in expectation leading to a social cost of $\mathsf{WorstEq} = I^2$. The social optimum on the other hand has one retailer order at each time period from $t^\star - I$ to $t^\star - 1$ generating a social cost of $\mathsf{Opt} = \frac{I(I+1)}{2}$ and therefore $\mathsf{WorstEq} = \mathsf{Opt} + \frac{I(I-1)}{2}$. The social optimum is not a Nash equilibrium. One example of a profitable deviation is for the the retailer ordering at time $t^\star - I$ to depart one period later. In this case, that retailer always receives their order before the deadline and therefore the deviation is profitable (they pay one less period of inventory cost). Note that when making this deviation, the player ordering at time $t^\star - 1$ receives their order after the deadline with probability 1. This highlights how congestion created by early deviations imposes costs on retailers ordering later as opposed to those who made the deviation. We then look at the best Nash equilibrium and show that it does not perform much better than the worst Nash equilibrium. Namely, in this equilibrium one player pays the highest individually rational cost I and the remaining players pay in expectation $I - 1$ generating a social cost of $\mathsf{BestEq} = I + (I-1)^2 = \mathsf{Opt} + \frac{(I-1)(I-2)}{2}$.

Next we look at the case whereby the wholesaler can charge a premium (over the cost of the inventory) depending on the time the retailer orders. In this case we show how the optimal premium structure is zero at time $t^\star - I$ and increasing by 1 up to some time $t^\star - k$ and then set to zero after. Any Nash equilibrium of the game augmented by this premium schedule is such that one player departs at each time $t^\star - I$ to time $t^\star - k$, all paying exactly I (after premiums) and then the remaining players play the worst Nash equilibrium of the remaining k player game. Note that such a premium schedule does not improve over the worst Nash equilibrium for the first $I - k$ players but significantly reduces the cost of the remaining k players. Still, this schedule is limited in improving the sum of total costs (which include the premiums) which is given by $\mathsf{Premium} = \frac{3}{4}I^2 = \mathsf{Opt} + \frac{I(I-2)}{4}$.

Finally, we turn to the case of the optimal correlated equilibrium. We characterize the support and necessary and sufficient conditions on the probabilities over this support (resulting in fewer inequalities than the basic definition). As mentioned in the introduction, this correlated equilibrium mixes over the social optimum and special *enforcing outcomes-t*

[3]The bound on C is a function of the number of players given that the number of players (and capacity) dictates the earliest time than any rational retailer would order.

for each time period $t = t^\star - I, ..., t^\star - 1$. Any enforcing outcome-$t$ is an outcome whereby one retailer is told to order at time $t^\star - t$ and $t - 1$ retailers are told to depart at time $t^\star - (t-1)$. In this case, whenever a retaler is told to order at time $t^\star - t$ and the outcome drawn by the correlated equilibrium is an enforcing outcome-t then they know they will be late with positive probability if they order instead at time $t^\star - (t-1)$ and late for sure if they depart any time later than that. Hence, whenever the correlated equilibrium puts sufficient probability on these outcomes then no player will deviate by ordering at an earlier time, even if the correlated equilibrium puts a very high probability on the social optimum (under which such a deviation is profitable). While the best correlated equilibrium is very easy to calculate, there is no closed form solution over the probabilities. Therefore, we utilize a simple *heuristic* correlated equilibrium to obtain our bounds on the cost of the optimum correlated equilibrium. Namely, we show that whenever $C > 2(I+1)I$ then the best correlated equilibrium cost $\mathsf{BestCE} < (1-\alpha(C)) \cdot \mathsf{Opt} + \alpha(C) \cdot \mathsf{BestEq} = \mathsf{Opt} + \alpha(C)\frac{(I-1)(I-2)}{2}$ where $\alpha(C) = \frac{(I+1)I}{C} < \frac{1}{2}$. Hence, even at the weak bound $2(I+1)I$, the best correlated equilibrium does better than the optimal premium schedule and approximates the social optimal as C increases given that $\alpha(C) \to 0$ as $C \to +\infty$.

0.3 Bank Regulation, Investment, and the Implementation of Capital Requirements

This paper studies the role of the implementation of capital requirements and their effect on the cost of raising capital. We start by assuming the bank starts with existing assets (financed with existing equity) with binary value of high (a_h) or low (a_ℓ) known only to the bank (with extensions of the main results to a continuum in the appendix). The prior belief of the market is that the asset value is high with probability p and low with probability $(1-p)$. Note that we assume $a_\ell \geq 0$ so that all banks are solvent at the start of the model. Hence, we study the regulation of banks in *normal* times, excluding issues of zombie banks and gambling for resurrection, problems that deserve their own attention but are undoubtedly related to the regulation of banks in normal times.

After learning its type (h or ℓ) the bank then receives a new investment opportunity with fixed cost I and NPV $b > 0$ (independent of the bank's type). The bank must seek financing for this new investment and can utilize insured (cheap) deposits subject to meeting the regulator's capital requirement. This requirement states that the bank must raise $K \leq I$ of the funds for the new project through the sale of some allowed *capital security* which we take in this introduction to be exclusively equity. Given that deposits are cheap, this capital requirement will always bind conditional on making the investment. Keeping this in mind,

the bank may not always decide to undertake the investment given the capital requirement. In particular, when the market cannot discern which bank is which, then it will be costly for the h type bank to raise equity. On the other hand, the ℓ type bank will receive a subsidy for raising equity conditional on it being optimal for the h type bank to undertake the investment.[4] In order to understand these incentives, denote by $b_\theta(K)$ the intrinsic value of the new project to the existing shareholders conditional on raising equity K. In general, $b_\theta(K) > b$ as it includes the value of the *deposit insurance put option*. Now, the decision of the bank to invest or not depends on the adverse selection problem characterized by (a_h, a_ℓ, p). Namely, the break even condition for the uninformed market is that in exchange for K funds, they receive a share $\alpha(K)$ of the bank's future cash flows. Under a competitive markets assumption, investors break even and therefore $\alpha(K)$ satisfies $\alpha(K)\hat{V} = K$ where $\hat{V} := pV_h + (1-p)V_\ell$ is the uninformed market value of the bank's post investment equity.[5] Now, given the market demands $\alpha(K)$, then the h-type bank will invest if and only if

$$V_h - \frac{K}{\hat{V}} V_h \geq a_h$$

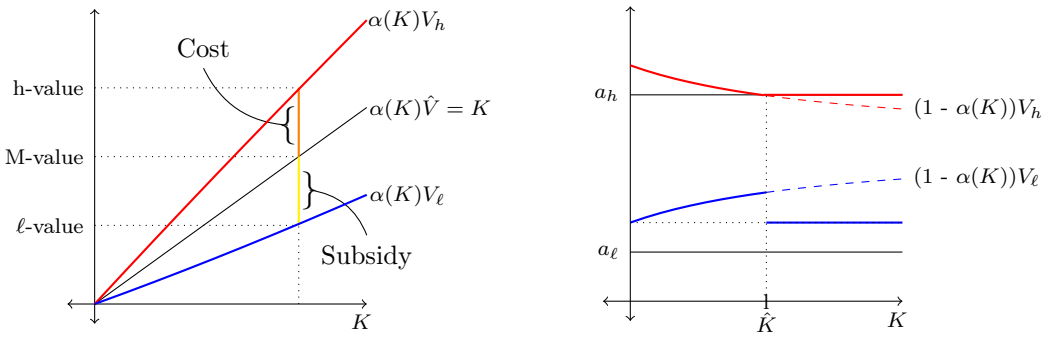
noting here that if $V_h = \hat{V}$ then equity is correctly priced and therefore this inequality always holds as $V_h = a_h + b_h(K) + K$. Whenever $p \in (0,1)$ and $a_h \neq a_\ell$ though, then $V_h > \hat{V}$ and therefore the h-type bank will always *over pay* for its equity by an amount $\frac{V_h}{\hat{V}}K - K > 0$. In more general terms, when the bank issues a security s, then mispricing of that security is $\mathbb{E}_h[s] - \mathbb{E}_p[s]$: the difference in the informed market value of that security and the uninformed market value. Therefore the bank invests whenever

$$b_h(K) \geq \mathbb{E}_h[s] - \mathbb{E}_p[s] = (1-p)(\mathbb{E}_h[s] - \mathbb{E}_\ell[s]) \tag{0.1}$$

Further, the fact that $b_h(K)$ is decreasing in K and $\mathbb{E}_h[s] - \mathbb{E}_\ell[s]$ is increasing in K implies that there exists some value of $\hat{K}$ such that this inequality is violated whenever $K > \hat{K}$. This is illustrated in Figure 3. As can be see in Figure 3 (a), the cost (subsidy) of raising equity for the h (ℓ) type is strictly increasing in the capital requirement K. This translates to Figure 3 (b) which plots the post investment value of inside equity conditional on meeting the capital requirement K. Naturally, the value of h (ℓ) type inside equity is decreasing (increasing) in the capital requirement. Yet, given that the h type outside option is to forgo the investment and obtain a_h, we can see that once $K > \hat{K}$ the h-type optimally forgoes. Further, once the h-type forgoes, the ℓ-type continues to invest, only its shares are now

[4] We show how under the optimal capital requirements there is no way for the h type to signal its type to the market via a larger equity issuance.

[5] Note that V_θ is a function of K but we drop dependence for notational convenience



(a) Bank v.s. Market Value of Shares.

(b) Bank profit post investment.

Figure 3: Illustration of Cost/Subsidy of Capital

correctly priced (the market knows no h type invests) and therefore it receives no subsidy (hence the discontinuity).

Now, the regulator's job is to design the capital requirement in order to maximize welfare which varies depending on the investment decision:

$$W_\theta(invest|K) = a_\theta + b - \lambda L_\theta(K) \qquad W_\theta(forgo) = a_\theta$$

Namely, when the bank invests, welfare increases by the surplus (NPV) generated by that investment b but decreases by the increase in the expected loss to the deposit insurance fund which we model as proportional to the expected loss.[6] Hence, the regulator's trade off is to set K as high as possible subject to inducing investment (assuming $L_\theta(K)$ is small with respect to b). What we then show is that the regulator has another option: design a menu of capital requirements and transfers $\{(K_h, T_h), (K_\ell, T_\ell)\}$ such that it is optimal (i.e. incentive compatible) for the h type to choose the option (K_h, T_h) and the ℓ type to choose (K_ℓ, T_ℓ). If this decision is observed (and still optimal) then the market will price each type's equity correctly once observing the menu option they have chosen. We call such a mechanism screening (Screen) and any mechanism that does not reveal any information ($K_h = K_\ell$ and $T_h = T_\ell$) pooling (Pool). Finally, we note that it may be optimal for the regulator to induce underinvestment. Namely, in the case where there are many ℓ-type banks, it will be optimal for the regulator to set very high capital requirements whereby they recapitalize the large proportion of the ℓ-type banks in exchange for underinvestment by the small proportion of the h type banks. We call this the underinvestment (Und) mechanism.

The main results are illustrated in Figure 4 which plots the social cost (first best welfare

[6]One can think of this as the deadweight loss caused by taxation in order to raise the necessary funds to reimburse depositors. Equivalently, one could assume a bank's failure leads to spillover effects to other financial institutions and the real economy.

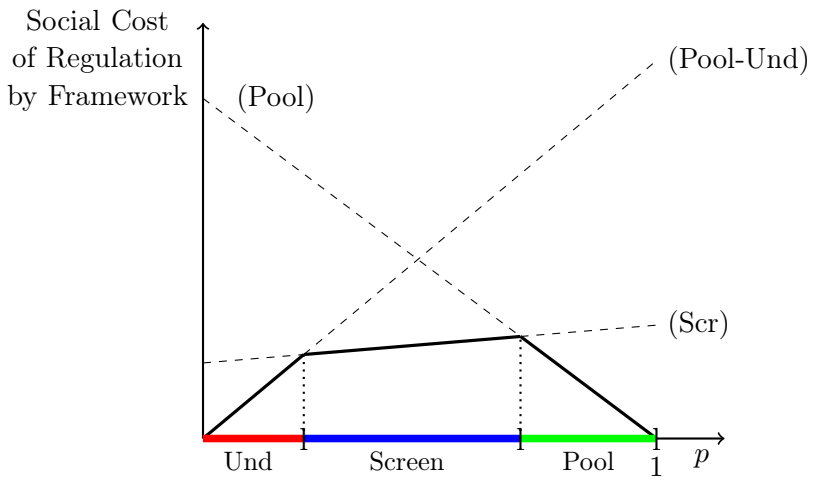


Figure 4: The Cost of Restricting to Each Framework

minus welfare of the respective framework) of utilizing each of the aforementioned frameworks (Sep, Pool and Und) conditional on optimally designing each framework. As can be seen, the pooling framework is optimal when the proportion of h-type banks (a measure of the strength of the banking sector) is large but becomes more and more expensive as that proportion decreases. This can be seen in equation (0.1) whereby when $p = 1$ the bank will invest given any capital requirement as $b_h(K) > b > 0$ for all K. Hence, when p is large, the cost of raising equity for the h type is small and therefore the regulator can still set a large capital requirement while inducing investment. On the other hand, when p is small the regulator must set a very small capital requirement in order to induce investment. Now as we can see for reasons mentioned above the underinvestment mechanism is optimal whenever p is small as the opportunity cost of underinvestment is small in that case. Finally, for intermediate values of p the optimal mechanism is screening. Note that the capital requirements of the screening mechanism are pinned down by the incentive compatibility conditions and independent of p (variation in the social cost comes from the increased probability of the bank being the h type). Hence, screening is only optimal when the cost of pooling and underinvestment is too high which is for intermediate values of p.

The remainder of the paper characterizes the optimal screening, pooling, and underinvestment mechanisms and deals with concerns of multiplicity of equilibria and signaling outside of the mechanism. I then discuss the features of optimal securities for the regulation of banks, showing how in a pooling mechanism contingent convertible bonds are optimal while in a screening mechanism the regulator should have the ℓ type banks raise capital by selling existing assets and the h type banks raise capital by issuing contingent convertible bonds. Finally, I also discuss the relevance of these results for policy implications, contrast-

ing them to the way that banks are currently regulated and relating them to new policy proposals such as the Counter Cyclic Capital Buffer of Basel III.

0.4 Robust Regulation of Bank Risk: Reporting and Risk Aversion

In this chapter of the thesis I analyze the ability of a regulator to induce bank's to reveal their private information regarding the riskiness of their assets. I assume that bank's must raise external funds to pay I for the new investment consisting of insured (i.e. cheap) deposits and newly raised equity K. The bank has access to unlimited cheap deposits, subject to meeting the regulator's capital requirement $K \geq \rho \cdot I$. Further, I assume that the market and regulator only know the *observable* distribution of returns f while the bank has private information regarding the *true* distribution of returns f_θ. For simplicity I parametrize the private information into a single variable $\theta \in [0, \overline{\theta}]$ and assume that the only difference between bank types is in the spread of their returns measured by a single parameter $\sigma_\theta = \sigma + \theta$ for some $\sigma \geq 0$: for any $\theta > \theta'$, f_θ second order stochastically dominates $f_{\theta'}$ in a strict sense so that $\int_{-\infty}^{x} F_\theta(t) - F_{\theta'}(t)dt > 0$ for almost all $x \in \mathbb{R}$. Further, in order to disentangle the incentives to report risk v.s. reporting returns I assume that all distributions f_θ generate the same expected return and therefore only vary in their spread. Under some standard assumptions, a larger spread will increase the risk neutral value of the bank's equity and the bank's existing equity holders will always prefer to raise as little equity as possible in the complete information case.

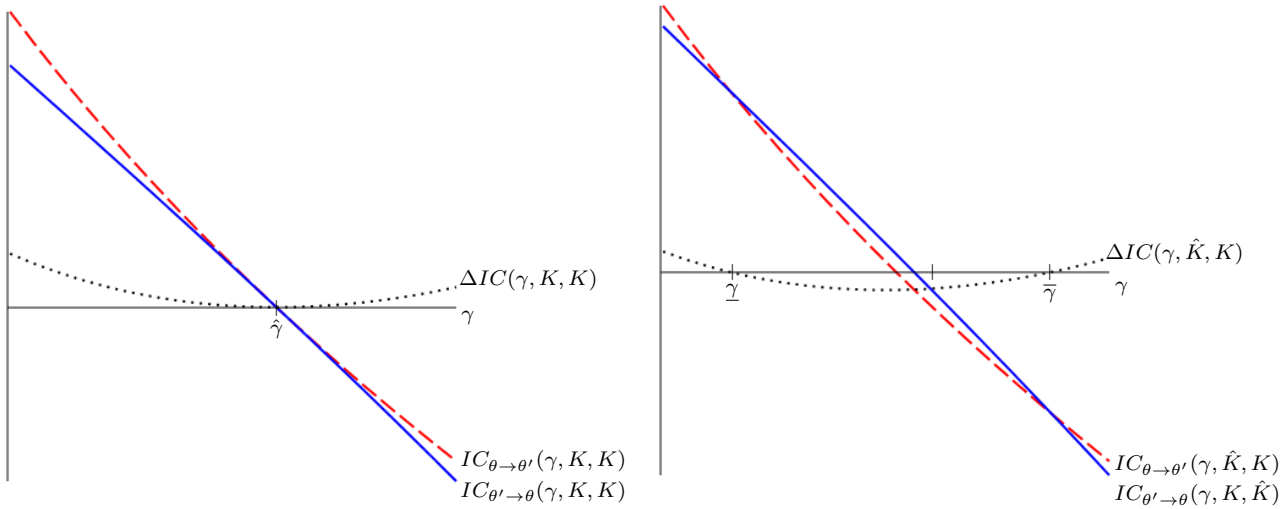
Denoting by X the random variable representing the net return on the bank's assets, then the regulators objective is to maximize social welfare:

$$W(f, K) := \mathbb{E}_f[X] - \lambda \cdot L(f, K) - c(K)$$

where $L(f, K)$ is the expected loss of the bank (and therefore the deposit insurance fund) and λ is a parameter representing the deadweight loss of bank failure, where it is assumed that this loss is proportional to the size of the bank's losses. Further, we assume that $c(K)$ is the social cost of raising capital which we take to be exogenous and note that $c(K)$ can be interpreted as the cost of underinvestment as described in Chapter 3. The only importance this cost plays is to provide us with an objective whereby the regulator does not wish to optimally set capital requirements to 100%.

The main innovation of the paper is to then introduce the possibility that investors are risk averse. Namely, we assume that all investors share a common utility function u_γ where γ measures the level of risk aversion, $\gamma = 0$ representing the risk neutral case and $\gamma = 1$ representing the infinitely risk averse case. Note that given that the bank's risk is based on

a single parameter θ, then it is without loss to parameterize γ to a single variable. Now, the risk adjusted value of the bank's equity is given by $V_\gamma(f_\theta, K)$, decreasing in γ and increasing in θ. We further make some natural assumptions regarding the behaviour of V_γ in order to ensure that it satisfies the natural features of risk aversion.



(a) Incentive constraints when $K_\theta = K_{\theta'} = K$

(b) Incentive constraints when $K_\theta = \hat{K} > K$

Figure 5: Incentive constraints when f_θ is normally distributed as $\mathcal{N}(1, \sigma_\theta)$ with $\sigma_{\theta'} = 1$, $\sigma_\theta = 3$, $K = 1$, and $\hat{K} = 1.1$.

In order to incentivize the bank to reveal its private information about risk the regulator must set capital requirements and transfers so that for all θ and θ' it is the case that

$$V_\gamma(f_\theta, K_\theta) - K_\theta - T_\theta \geq V_\gamma(f_\theta, K_{\theta'}) - K_{\theta'} \frac{V_\gamma(f_\theta, K_{\theta'}) - T_{\theta'}}{V_\gamma(f_{\theta'}, K_{\theta'}) - T_{\theta'}} \tag{0.2}$$

The term on the LHS represents the payoff to the bank's existing shareholders after truthfully reporting their type is θ as opposed to the payoff on the RHS from lying and reporting type θ'. As we can see, when the bank reports truthfully, under a competitive markets assumption (and normalizing the risk free rate to zero) they simply need to repay the investors the capital that they raised. On the other hand, when the type θ bank lies and reports it is type θ', then it faces a different capital requirement $K_{\theta'}$ and also a different cost of financing that capital $\frac{V_\gamma(f_\theta, K_{\theta'}) - T_{\theta'}}{V_\gamma(f_{\theta'}, K_{\theta'}) - T_{\theta'}} K_{\theta'}$. Similar to Chapter 3, when $\theta > \theta'$ then for small values of γ it is the case that $V_\gamma(f_\theta, K) > V_\gamma(f_{\theta'}, K)$ and therefore the type θ bank overpays for the equity financing $K_{\theta'}$ when it pretends to be type $\theta' < \theta$.

Now, in a risk neutral world there always exists a capital requirement $K_\theta > K_{\theta'}$ and a transfers T_θ and $T_{\theta'}$ such that 0.2 is satisfied. This comes from the fact that in a risk neutral world, the bank's value of revealing they are riskier is higher for riskier banks. Therefore,

riskier banks are willing to pay more (i.e. though a higher capital requirement) to reveal they are risker than less risky banks. In fact, after rearranging Equation 0.2 and the equation that ensures the type θ' bank prefers to report truthfully as opposed reporting it is type θ, incentive compatibility between θ and θ' requires

$$IC_{\theta\to\theta'}(\gamma, K_\theta, K_{\theta'}) \geq T_\theta - T_{\theta'} \geq IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K_\theta) \tag{0.3}$$

where

$$IC_{x\to y}(\gamma, K_x, K_y) = V_\gamma(f_x, K_x) - V_\gamma(f_x, K_y) + K_y \frac{V_\gamma(f_x, K_y) - T_y}{V_\gamma(f_y, K_y) - T_y} - K_x$$

In fact, when $K_\theta = K_{\theta'} = K$, then $IC_{\theta\to\theta'}(\gamma, K_\theta, K_{\theta'}) - IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K_\theta)$ represents exactly the extra amount over the type θ' bank that the type θ bank is willing to pay to have the market believe it is type θ as opposed to θ'. I then show that this difference shrinks to zero as γ goes from 0 to some value $\hat{\gamma}$ (illustrated in Figure 5 (a)) and then increases as γ goes from $\hat{\gamma}$ to 1. In fact, the value $\hat{\gamma}$ corresponds precisely to the value at which $V_{\hat{\gamma}}(f_\theta, K) = V_{\hat{\gamma}}(f_{\theta'}, K)$. Namely, while in a risk neutral setting taking more risk leads to a higher value of equity, this benefit is strictly decreasing as the level of investor risk aversion increases and eventually becomes negative. Now, as illustrated in Figure 5 (b), if the bank is indifferent between the level of risk θ and θ' then that bank will always choose the regulatory option that gives them the lowest capital requirement as in this case there is no mispricing of equity. Hence, when $\gamma = \hat{\gamma}$ and $K_\theta = \hat{K} > K = K_{\theta'}$ then the mechanism cannot be incentive compatible, regardless of the transfers. I further prove that, as illustrated in Figure 5 (b), for any $\hat{K} > K$ there exists an interval $(\underline{\gamma}, \overline{\gamma})$ such that that any mechanism that sets $K_\theta = \hat{K}$ and $K_{\theta'} = K$ for any $\theta > \theta'$ is not incentive compatible whenever $\gamma \in (\underline{\gamma}, \overline{\gamma})$.

Given this last result, I then move to study when the regulator can provide (γ_0, ϵ)-*robust* incentives whereby the mechanism is incentive compatible for an initial value γ_0 and remains incentive compatible for all $\gamma \in [\gamma_0 - \epsilon, \gamma_0 + \epsilon]$. The main result then states that for any $\epsilon > 0$ a (γ_0, ϵ)-robust mechanism exists only if γ_0 is sufficiently large or sufficiently small. The reasoning for this is illustrated in Figure 6, which shows the incentive constraints when $K_\theta = K_{\theta'} = K$. In this case, the regulator would only like to induce the bank's to reveal their private information as opposed to also adjusting their capital requirement. The robustness issue arises precisely when γ approaches $\hat{\gamma}$ as given that incentive compatibility requires 0.3 to hold, then (γ_0, ϵ)-incentive compatibility requires 0.3 to hold for all $\gamma \in [\gamma_0 - \epsilon, \gamma_0 + \epsilon]$. I then show that for a fixed ϵ not to large, this can be achievable for small and large values of γ but not for values around $\hat{\gamma}$. This is illustrated in Figure 6 whereby $\Delta T = T_\theta - T_{\theta'}$ is

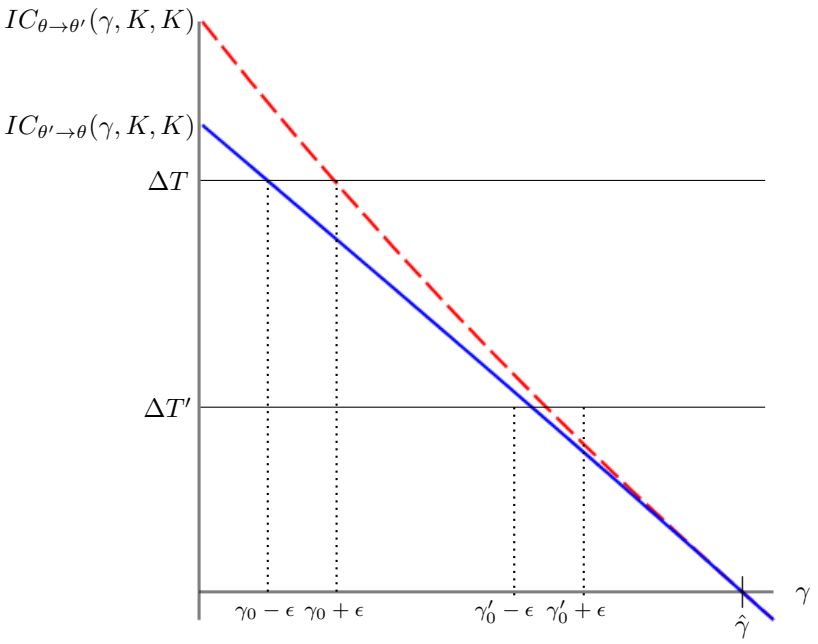


Figure 6: Incentive constraints when $K_\theta = K_{\theta'} = K$

(γ_0, ϵ)-robust, but not (γ_0', ϵ)-robust. This comes from the fact that for the initial value of risk aversion γ_0' the difference in transfers ΔT will be too large when $\gamma = \gamma_0' + \epsilon$ but too small when $\gamma = \gamma_0' - \epsilon$. Finally, I show how this result applies for all values of $\epsilon > 0$ and is even more pronounced when capital requirements are report specific (i.e. $K_\theta > K_{\theta'}$).

As a final remark, I should note that the fact that robustness is incentive compatible for large values of γ_0 comes from the fact that once $\gamma > \hat{\gamma}$, then the θ' types are willing to pay more than the θ types to convince the market that they are type θ' as opposed to type θ. This is generated by a symmetric argument from before, noting that whenever $\gamma > \hat{\gamma}$ then revealing more risk above θ' strictly decreases the market's valuation of your equity in which case the θ' types are willing to pay more to reveal this information. Hence, by setting the right transfers the regulator can still achieve incentive compatibility when γ is large. Yet, by similar arguments, such incentive compatibility will break down as γ approaches some interior value $\hat{\gamma}$ as will the ability to design robust incentives. Finally, I extend these results to the case whereby a bank of type θ can choose any level of risk in $[0, \theta]$. In this case, a similar result holds only that the upper bound goes from $\overline{\gamma} > \hat{\gamma}$ to $\overline{\gamma} = \hat{\gamma}$ as the incentives that break down between reporting type θ and θ' disappear whenever $\gamma > \hat{\gamma}$ as in that case choosing a level of risk $\theta > \theta'$ is strictly suboptimal and therefore no type θ will exist.

The remainder of this thesis consists of the four aforementioned chapters and their self-contained literature reviews.

References

[1] Cochrane, J. H. (2017): “Macro-Finance." *Review of Finance*, 21(3), 945-985.

1 Incentives and the Structure of Communication

Abstract

This paper analyzes the issue of implementing correlated and communication equilibria when pre-play communication is restricted to a particular network (*e.g.,* a hierarchy). When communication between the mediator and the players is not direct and private, as assumed when invoking the revelation principle, there may be incentives for other players in the communication network to misbehave while players report their private information to the mediator and the mediator sends suggested actions to the players. To remedy this issue, we provide necessary and sufficient conditions on the topology of the network of communication such that restricting communication between the mediator and the players to a particular network does not restrict the set of (communication equilibrium) outcomes that could otherwise be achieved. We show that for any underlying game and any equilibrium outcome available when communication is direct, there exists a communication scheme restricted to a particular network that implements all such outcomes (*i.e.,* does not induce players to deviate in the communication phase) if and only if that network satisfies our conditions.[7]

1.1 Introduction

It is well known that the set of equilibrium outcomes can be largely expanded when players have the ability to communicate with an impartial third party prior to playing a game of complete or incomplete information. The revelation principle is a powerful tool that characterizes exactly what outcomes can be achieved in this context. It states that any equilibrium of an arbitrary communication mechanism between the players and an impartial third party, or mediator, can be replicated with a *direct mechanism*. Under such a mechanism, players report their private information to the mediator who then draws an outcome from a distribution contingent on the reported information and suggests each player an action to play. This implies that the set of outcomes induced by direct mechanisms that are *incentive compatible* (*i.e.,* players have the incentives to report their types truthfully and to play

[7]I would like to thank my PhD advisor Tristan Tomala, Charlène Cosandier, Françoise Forges, Frédéric Koessler, Eric Mengus, Ludovic Renou, Joel Sobel, Péter Vida, Marie Vigeral, and seminar participants at: the Theory, Organizations, and Markets seminar at the Paris School of Economics, the HEC Economics and Decision Sciences departmental seminar, and the LUISS Economics and Finance departmental seminar for helpful discussions and comments. Finally, I would like to thank the Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047) for supporting this research.

the action suggested to them) characterize all outcomes that can be achieved with *pre-play communication*.

An important assumption of the revelation principle is that the mediator has the ability to communicate *directly and privately* with each player of the game when collecting their reported information and suggesting actions to them. Due to the fact that in many contexts this assumption may be strong, or costly to maintain, the purpose of this paper is to investigate the conditions under which restricting communication between the players and the mediator to a particular network does not restrict the set of outcomes that they could otherwise achieve. In this sense, we are interested in the class of communication networks that guarantee, for any achievable outcome with direct communication, the existence of a communication scheme restricted to each such network that induces the same outcome. We characterize robust conditions such that no matter how the game nor its achievable (equilibrium) outcomes may change, such a communication scheme always exists. This amounts to characterizing conditions on the network such that for any game one can implement the entire set of *correlated equilibria* (games of complete information, Aumann (1974)) and *communication equilibria* (games of incomplete information, Forges (1986) and Myerson (1986)) of that game.

Thus, our main results provide necessary and sufficient conditions on the communication network topology that guarantee any correlated/communication equilibrium of any game can be implemented as a perfect Bayesian equilibrium of the extended game where communication is restricted to such a network.

The key insight developed is that for any game, any correlated/communication equilibrium outcome can be implemented when preplay communication is restricted to some network only if this network guarantees that the mediator can send messages to each player in a *perfectly secure* fashion. This requires that the network be sufficiently *connected* so that there exists a communication protocol for suggesting tasks to each player satisfying a *secrecy* condition — that no player learns any information about the action suggested to any other player, and a *resiliency* condition — that each player receives their correct suggested action with probability 1 under any unilateral deviation from the protocol.[8] Hence, our main results state that the implementation of the entire set of correlated/communication equilibria of any game requires that the mediator be sufficiently *connected* to any player that he cannot directly communicate with.

The requirement of perfectly secure communication as a sufficient condition is straightforward: if the mediator can communicate in a perfectly secure fashion with each player,

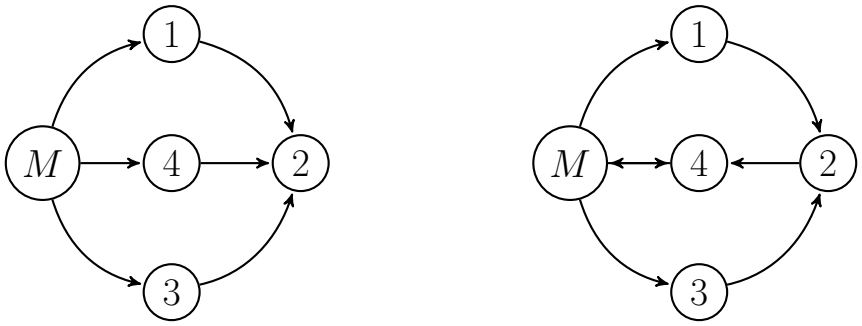
[8]The terms secrecy and resiliency come from the computer science literature where the study of secure communication originates.

then this implies that there exists a way for him to communicate with each player *as if* the communication were direct and private.[9] The main contribution of this paper is in characterizing the necessary and sufficient conditions on the topology of the communication network that ensure perfectly secure communication is achievable. Given that we do not restrict the class of cheap talk communication protocols in any way (other than that they operate over the underlying network), this requires proving that perfectly secure communication is not achievable on any network not satisfying our conditions, no matter how sophisticated the communication protocol and continuation play may be. For example, we allow for repeated communication between the mediator and players (*i.e.,* cycles in the network) which opens up the possibility for protocols equipped with subprotocols for the detection and punishment of deviations in the communication phase. Hence, the novelty of our construction is that we can show our conditions are necessary among the extremely large set of conceivable communication protocols and continuation strategies. Finally, in order to prove sufficiency of our results, we construct a communication protocol for implementing any correlated/communication equilibrium over any network satisfying our conditions and prove that it is perfectly secure.

The main result of this paper for games of complete information (Theorem 1) is that the mediator (M) can implement any correlated equilibrium outcome of any game with communication restricted to the network $\mathcal{N}$ if and only if for every Player i that the mediator cannot directly send messages to, the network $\mathcal{N}$ satisfies the condition that either (1) there are 3 disjoint directed paths from M to i, or (2) there are 2 disjoint directed paths from M to i and one additional directed path from i to M (all three being disjoint). Figure 7 illustrates these necessary and sufficient conditions in a 4-player network where the mediator can directly send messages to all players except for Player 2. The intuition is that with only two disjoint directed paths between the mediator and the players we prove that one cannot guarantee simultaneously privacy and resiliency of communication. We then show that the lack of privacy and/or resiliency equates to the existence of a profitable deviation in *any* underlying communication protocol whenever the network does not have three disjoint paths. Therefore, after showing that two disjoint paths *from* the mediator *to* any Player i that he cannot directly communicate with are necessary but not sufficient (as illustrated in the example of Figure 5 in Section 3.2), we conclude that there must be an additional disjoint path connecting the mediator and Player i: if the path is from M to i, then this is Condition (1) of Theorem 1, whereas if the path is from i to M, then this is Condition (2)

[9] In this context, one simple definition of direct and private communication is a mode of communication through which the mediator can send a message to a player in such a way that no other player learns any information about the message sent (private) nor can prevent the message from being received (direct). This is equivalent to perfectly secure communication.

of Theorem 1. For further discussion on how these conditions are derived, see Section 3.2 below where we illustrate our results for the 3-player case.



(a) *A network satisfying Condition* (1). (b) *A network satisfying Condition* (2).

Figure 7: *4-player networks satisfying Theorem 1.*

Next, we look at the case of implementing communication equilibrium. Theorem 2 provides necessary and sufficient conditions for implementation in Bayesian games of any environment; for every player that cannot communicate directly to M, $\mathcal{N}$ must have 2 disjoint directed paths from i to M and satisfy the conditions of Theorem 1. Finally, as special cases of the general results, Corollary 3 characterizes the networks that allow for the implementation of all communication equilibria of Bayesian games satisfying the private values and common independent beliefs assumptions and Corollary 4 provides necessary and sufficient conditions for implementation on undirected networks.

1.1.1 Related Literature

There has been a long line of literature characterizing when players can achieve the set of correlated/communication equilibria outcomes when an impartial mediator is not available to collect information and report suggested actions. Bárány (1992) and Forges (1990) construct protocols for games of 4 or more players that allow the players of the game to achieve the set of correlated and communication equilibrium outcomes (resp.) without a mediator as the set of Nash equilibria of the game extended by a preplay communication phase. Ben-Porath (2003) constructs a similar protocol for games of 3 or more players whenever the solution concept is sequential equilibrium of the *extended game* under the assumption that there exists a (Bayesian) Nash equilibrium outcome that makes all players worse off than any correlated (communication) equilibrium outcome. Gerardi (2004) constructs a protocol for games of 5 or more players that allows players to replace the mediator and achieve the set of correlated/communication equilibria via sequential equilibrium of the extended game without making any assumptions on the underlying game. For a more detailed survey of

implementation of correlated and communication equilibria without a mediator we refer the reader to Forges (2009). It is worth noting that similar problems have also been studied in the computer science literature to which we refer the reader to Halpern (2007). The main similarity between this literature and the current paper is in understanding when the set of correlated and communication equilibrium outcomes can still be achieved while departing from the standard framework with a trustworthy mediator who communicates directly and privately with each player.

There are many additional references in computer science that study the possibility of secure communication between two nodes in a network given that some subset of nodes on the paths connecting them wish to learn the message or prevent the receiver from learning the message. For further references on this topic we refer the readers to Renault, Renou, and Tomala (2014) for the problem from a game theoretic perspective and Dolev et. al. (1993) for the problem approached from the computer science literature. In particular, Dolev et. al. (1993) consider a network between a sender and receiver consisting of k undirected vertex disjoint paths and ask when there exists a communication protocol for sending a message m between the sender and receiver in a way such that any agent listening to σ of the paths *and* with the ability to alter the communication on ρ of the paths cannot learn anything about m *nor* interrupt the transmission of m. They show (among other results) that when the listening paths are a subset of the disrupting paths or vice-versa, such a communication protocol exists if and only if there are $k \geq \sigma + 2\rho + 1$ disjoint paths when communication flows one-way from sender to receiver and $k \geq max\{\sigma + \rho + 1, 2\rho + 1\}$ disjoint paths when communication flows two-way between sender and receiver. The main differences between this paper and Dolev et. al. (1993) is that (1) we consider general undirected networks (similar to Renault, Renou, and Tomala (2014)), and (2) while there is a single sender (the mediator), there are n receivers: the mediator must be able to send messages to each player without revealing information to any of the other players, nor inducing them to alter the messages they are required to send in the process. As will be seen below, the protocols that we construct to prove sufficiency of our conditions on the network topology heavily rely on *encryption techniques* from computer science.

The paper most closely related to this one is Renou and Tomala (2012) who study a similar problem in a mechanism design context. They characterize conditions on the topology of communication networks for the implementation of any social choice function when players are required to report their private information to the designer and communication is restricted to a network. This paper extends the results of Renou and Tomala (2012) by generalizing this work to the case where players take actions after reporting their types and to the case where the mediator has the ability to make suggestions to the players.

Our conditions are independent and stronger than the conditions provided by Renou and Tomala (2012): the conditions for the mediator to be able to send suggested actions to each player without inducing deviations imply the conditions for the players to be able to send their private information to the mediator. In fact, Renou and Tomala (2012) show that when solving their problem one can assume, without loss, that the mediator does not communicate. Therefore, they only need to consider acyclic communication in the network which greatly simplifies the problem. The main difficulty in proving our results comes from necessarily relaxing this assumption. For example, in contrast to Renou and Tomala (2012), in our construction we must allow for the players to report detected deviations to the mediator in order to attempt to deter them with punishment strategies (see the example of Figure 5 in Section 3.2 for an illustration of this point). In this sense, in proving necessity of our conditions we must consider a much larger class of communication protocols and subprotocols that allow for the mediator to repeatedly communicate with the players. Finally, the full characterization of Renou and Tomala (2012) requires that the environment satisfy either a private values and common independent belief assumption or a worst case outcome assumption.[10] In this paper we obtain a full characterization without making any assumptions on the underlying environment (Theorem 2).

As mentioned above, our necessary conditions for the implementation of correlated equilibria require that the mediator have access to a perfectly secure communication protocol that allows him to suggest strategies to each player.[11] The notion of a secure communication protocol satisfying the conditions of secrecy and resiliency (as introduced above) has been characterized in Gossner (1998) in a game theoretic context. Gossner defines a protocol of a communication mechanism that induces a particular information structure as an interpretation of the signals that players receive before playing the game and an associated action based on that interpretation. A secure protocol is one such that for any game, any Nash equilibrium induced by that game coupled with some exogenous information structure is also a Nash equilibrium of the same game augmented by a communication mechanism that induces the same information structure – interpreting signals and taking the associated action as specified by the protocol is optimal. He then shows that a protocol is secure if and only if it satisfies secrecy and resiliency. This paper extends this characterization for games of complete information to the case where the communication mechanism is restricted to

[10] They obtain a partial characterization for general Bayesian games.

[11] Note that the term *perfectly secure* has also been used in the cryptographic literature to refer to cryptosystems or protocols that are *information-theoretically secure* in the sense that they satisfy some definition of security even against adversaries with unlimited computational power. While in this paper we assume that players have unlimited computational power, the notion of *perfectly secure communication* as defined in this paper is not intended to match any specific cryptographic definition.

a particular network and the solution concept is perfect Bayesian equilibrium (we use the term perfectly secure instead of secure to clarify this distinction). Namely, a protocol of a communication mechanism restricted to the network $\mathcal{N}$ is secure (in Gossner's sense) under the perfect Bayesian equilibrium concept if and only if $\mathcal{N}$ guarantees secrecy and resiliency of communication from the mediator to every player. In the computer science context, Ben-Or, Goldwasser and Wigderson (1988) and Chaum, Crépeau, and Damgård (1988) construct perfectly secure protocols that are strictly stronger than the protocols considered in Gossner (1998). They consider the case where n parties with private information $x_1, ..., x_n$ wish to compute any function $f(x_1, ..., x_n) = (y_1, ..., y_n)$ in such a way that no party $i = 1, 2, ..., n$ learns more than their input x_i and output y_i. They construct a protocol such that for any coalition of players of size less than $\frac{n}{3}$, any joint deviation by the coalition produces no additional information and does not disrupt the messages received by the remaining players. They further prove that these bounds are tight in the sense that no such communication protocol exists when the coalition is greater than $\frac{n}{3}$: in this case there exists a function of the inputs that cannot be computed in a fashion that is resilient to deviations by the coalition without releasing additional information to one of the parties.

The results of this paper can be applied to the study of optimal communication in organizations as it falls within the intersection of the *incentives* approach to organizations (*i.e.,* principal agent problems) and the *team theoretic* approach to organizational design (see *e.g.,* Marschak and Radner (1972)). In fact, one could imagine a principal-agent setting whereby the principal would like to achieve his most preferred outcome while minimizing his own cost of communication — direct communication with each agent being the most costly for the principal. In such a setting our characterization would guarantee that no matter how the underlying incentives of the organization changed, any communication network satisfying our conditions would allow the principal to implement his most preferred outcome. Any further analysis past this point would necessarily have to weigh the benefits of choosing a *cheaper* communication network, not satisfying our conditions, against the costs of settling for a less preferred outcome. In this sense, we see our results as being as general and robust as one can obtain, providing a baseline for more stylized work in understanding the conditions under which the tradeoff between incentivizing agents to behave and minimizing the cost of communication is relevant. An interesting implication of the conditions we characterize is that no matter what the underlying incentives of the agents are, the principal can always achieve his most preferred outcome by utilizing a communication network whereby he *only* communicates directly with three agents *regardless of the size of the organization.*[12]

To the author's knowledge, the vast majority of the literature that studies both incentives

[12]Each agent need only communicate with three other agents as well.

and communication act as generalizations to the seminal paper of Crawford and Sobel (1982) on cheap talk communication (one exception being Renou and Tomala (2012) discussed above).[13] Dewatripont (2006) combines a model of imperfect communication introduced by Dessein and Santos (2003), with the moral hazard approach to communication as studied in Dewatripont and Tirole (2005). The key insight obtained is the coexistence of cheap talk and costly communication within the organization. Hagenbach and Koessler (2010) and Galeotti *et. al.* (2013) analyze equilibrium networks of truthful information transmission that arise in the Crawford and Sobel setting when biased agents can communicate with each other prior to choosing actions. Another relevant issue in the design of organizations is delegation and communication. Dessein (2002) studies a sender/receiver problem à la Crawford and Sobel (1982) and shows that an uninformed principal prefers to delegate the decision rights of a particular task to the informed agent whenever there is a large amount of uncertainty regarding the underlying environment. Similarly, Alonso *et. al.* (2008) and Rantakari (2008) study whether the principal prefers to communicate or delegate decision rights when the organization consists of multiple divisions, each of which aims to maximize their own division's profit (dependent on their private information), but also benefit from coordination (*e.g.,* due to economies of scale in joint production). The relation to this paper is that we analyze a case where the size of the firm is very large so tasks must be delegated and analyze the incentives to truthfully communicate post delegation. In Section 6 we discuss an application where the delegation of an international mechanism (the Single Supervisory Mechanism) generates a network of communication via reporting of bank risks' to national bank supervisors who then report this information to the European Central Bank. An interesting future line of research would be to understand how our results impact the delegation decisions themselves. Important to note here is that while these papers combine costly communication and incentives, they are independent from this paper which specifically analyzes how a fixed communication structure can create incentives to misbehave. Further, these papers make use of relevant, but stylized, models with standard quadratic loss functions to obtain their results, unlike the current paper which makes no assumptions on payoffs.

The rest of the paper is organized as follows. Section 2 presents important preliminary definitions and notation including our notion of implementation. Section 3 illustrates in detail why our conditions are necessary and sufficient in the context of the 3-player case. Sections 4 and 5 present the main results of the paper for the implementation of correlated equilibrium and communication equilibrium respectively. In Section 6 we analyze two applications to which the insights of this paper can prove useful, the organization of a large multinational

[13]For a survey highlighting the lack of research that addresses both incentives and communication within the organization and further motivation for models incorporating both features see Mookherjee (2006).

enterprise and communication of international bank risks in the European Union. Section 7 concludes and provides some comments regarding the extension of our results to the case where players have access only to finite message spaces. Proofs are relegated to the appendix in Section 8.

1.2 Preliminaries

This section will introduce some important definitions and notation that will be used repeatedly throughout the rest of the paper. Most importantly, we will introduce our concept of implementation of the set of correlated equilibrium outcomes as perfect Bayesian equilibrium outcomes of the extended game with preplay communication restricted to some network $\mathcal{N}$.

We consider finite normal form games, each represented as $\Gamma = (I, (S_i)_{i\in I}, (u_i)_{i\in I})$ where I represents the set of players, S_i the set of pure actions for Player $i \in I$, and $u_i : \times_{i\in I} S_i \to \mathbb{R}$ the payoffs of Player i dependent on the outcome $s \in S := \times_{i\in I} S_i$. We add a communication constraint by restricting preplay communication between the mediator, M, and the players of the game, I, to directed communication networks of the form $\mathcal{N} = (V, A(\mathcal{N}))$ with vertex set $V = I \cup \{M\}$. We let the set of directed edges $A(\mathcal{N}) \subset \{ij : i \in V, j \in V\}$ of the network $\mathcal{N}$ represent the available private communication channels such that Player $i \in V$ can send messages to Player $j \in V$ if and only if $ij \in A(\mathcal{N})$. For any network $\mathcal{N}$ and any $i \in I \cup \{M\}$ we will denote $Y_{\mathcal{N}}(i) := \{j \in I \cup \{M\} : ij \in A(\mathcal{N})\}$ as the set of *successors* of Player i in the network $\mathcal{N}$ and $X_{\mathcal{N}}(i) := \{j \in I \cup \{M\} : ji \in A(\mathcal{N})\}$ as the set of *predecessors* of Player i in the network $\mathcal{N}$. When the network is clear from context we will drop the subscript on $X_{\mathcal{N}}(i)$ and $Y_{\mathcal{N}}(i)$. Throughout we assume that players have access to some *universal* message space and, as a simplifying assumption, we will suppose that if $ij \in A(\mathcal{N})$ then at each time t in the preplay phase, Player i can send any k-vector $m \in \mathcal{M}^k = [0, 1)^k$ to Player j for any finite $k \in \mathbb{N}$. This assumption is without loss as long as the message space is infinite. For a discussion on the extension of our results to finite message spaces we refer the reader to Section 7.

1.2.1 Implementation

The object of interest is the extensive form game $(\Gamma, \mathcal{N})$; the game Γ augmented by an arbitrarily long but finite preplay cheap talk communication phase *restricted* to the network $\mathcal{N}$. Denoting by $B(\Gamma, \mathcal{N})$ the set of distributions over outcomes induced by perfect Bayesian equilibrium (PBE) of the game $(\Gamma, \mathcal{N})$, the question this paper looks to answer is what is the relationship between $B(\Gamma, \mathcal{N})$ and $C(\Gamma)$ the set of correlated equilibria distributions of Γ. Recall, that $Q \in \Delta(S)$ is a correlated equilibrium of the game Γ if and only if it satisfies

the conditions that

$$\sum_{s\in S} u_i(s_i, s_{-i})Q(s_i, s_{-i}) \geq \sum_{s\in S} u_i(\delta_i(s_i), s_{-i})Q(s_i, s_{-i})$$

for all $i \in I$ and $\delta_i : S_i \to S_i$.

For the correlated implementation problem we consider a communication protocol as a vector $\mathcal{P}(\mathcal{N}) = (T, \rho, \hat{\sigma})$ where T is the length of communication[14], ρ is a *communication strategy* that maps for any time $t = 0, 1, ..., T-1$ the history of communication of each player up to time t to a vector of messages to be sent by each player at time $t+1$ to their successors in the network $\mathcal{N}$. Finally, $\hat{\sigma}$ is an *action strategy* that maps any time T communication history induced by ρ to an outcome in S. Namely, under the action strategy $\hat{\sigma}$ Player $i \in I$ with time T communication history h_i^T plays the action $\hat{\sigma}_i(h_i^T) \in S_i$ in the play phase of the extended game. These communication protocols are defined precisely in Section 8.1 of the appendix. We can now define our concept of correlated equilibrium implementation on $\mathcal{N}$.

Definition 1.1. 2.1 (Implementation of Correlated Equilibrium) Let $\Gamma = (I, (S_i)_{i\in I}, (u_i)_{i\in I})$ be a finite game and $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ a communication network. Then, we say that the set of perfect Bayesian equilibria of the game $(\Gamma, \mathcal{N})$ is equal to the set of correlated equilibria of Γ if for every correlated equilibrium $Q \in C(\Gamma)$ there exists a perfect Bayesian equilibrium $(\rho, \hat{\sigma})$ of the game $(\Gamma, \mathcal{N})$ that induces the same distribution over outcomes as Q such that $\mathbb{P}_\rho(\hat{\sigma} = s) = Q(s)$ for all $s \in S$.

Comment 1.2. At this point we should mention that our definition of implementation and our results do not consider implementation without a mediator as in Gerardi (2004). In this paper we have in mind a principal-agent problem whereby the principal (acting as mediator) chooses an equilibrium that best meets his or the organization's (*e.g.,* the shareholders') objectives à la Myerson (1982). In such a setting, it is important to understand what the principal can achieve given the structure of communication within the organization to which our characterization lends insight. That being said, there have been advances in the computer science literature in addressing the problem of unmediated implementation on a network. While Gerardi (2004) makes heavy use of public communication in his protocol, it has been shown by Lamport, Shostak, and Pease (1982) that public communication can be replicated with private communication (assuming a private communication channel between

[14]Note that we only require that the communication mechanism involve weakly more periods of communication than T since we can always take beliefs of the players to be that they will not receive any informative messages after time T. Given these beliefs, it is optimal for them to not communicate after time T and therefore any protocol that is an equilibrium when the mechanism has T periods of communication is also an equilibrium when the mechanism has $T' > T$ periods of communication.

every pair of parties) if and only if the number of misbehaving parties is less than $\frac{n}{3}$. Dolev (1982) then generalizes this work to communication over private networks and shows that public communication can be replicated on a private communication network if the number of faulty parties is less than $\frac{n}{3}$ *and* the number of faulty parties is less than $\frac{C}{2}$ where C is the connectivity of the communication network. Finally, as remarked in Dolev et. al. (1993), by combining their result with the results of Ben-Or, Goldwasser, and Wigderson (1988), one can achieve secure computation (see the description of the results of Dolev et. al. (1993) in Section 1.1 for the meaning of secure computation) whenever the number of faulty parties is less than $\frac{n}{3}$ and the number of faulty parties is less than $\frac{C}{2}$ where C is the connectivity of the communication network. Therefore, given that general secure computation is sufficient to address unmediated correlated equilibrium implementation, (although beyond the scope of this paper) one can start with these results to obtain a potential characterization for unmediated implementation of correlated equilibrium on a directed network.[15]

In what follows, for any CE Q we denote by $\tilde{s}$ the random variable distributed according Q. We can now introduce our notion of perfectly secure communication.

Definition 1.3. 2.2 A correlated equilibrium communication protocol $(T, \rho, \hat{\sigma})$ is *perfectly secure* if it satisfies the following two conditions:
(1) *Secrecy:* No Player $i \in I$ learns any information about the strategy suggested to Player $j \in I \backslash \{i\}$: $\mathbb{P}_{\rho,Q}(\hat{\sigma}_j = s_j | h_i^T) = \mathbb{P}_{\rho,Q}(\hat{\sigma}_j = s_j)$ for all $h_i^T \in supp_i(\rho)$.
(2) *Resiliency:* Player $i \in I$ learns their suggested strategy with probability 1 under any unilateral deviation from the protocol: $\mathbb{P}_{\rho'_j,\rho_{-j}}(\hat{\sigma}_i = s_i | \tilde{s}_i = s_i) = 1$ for all $j \in I \backslash \{i\}$ and all communication strategies ρ'_j.

Now, if we let $\mathcal{N}^\star$ be the *complete network* where all players can communicate privately with the mediator and each other, then it is easy to show that $B(\Gamma, \mathcal{N}^\star) = C(\Gamma)$. Namely, Q is a correlated equilibrium if and only if it can be implemented as a PBE of the game $(\Gamma, \mathcal{N}^\star)$ through the *canonical communication protocol* (see Section 3.1 for an illustration). Hence, $C(\Gamma) \subset B(\Gamma, \mathcal{N}^\star)$. Further, if we denote by $NE(\Gamma, \mathcal{N}^\star)$ the set of distributions induced by Nash equilibria of $(\Gamma, \mathcal{N}^\star)$ then by the revelation principle we know that any distribution in $NE(\Gamma, \mathcal{N}^\star)$ can be implemented via the canonical protocol (given the mediator can communicate directly with each player in $\mathcal{N}^\star$) and therefore must be an element of $C(\Gamma)$. Therefore, $B(\Gamma, \mathcal{N}^\star) \subset NE(\Gamma, \mathcal{N}^\star) = C(\Gamma)$. Next, it is easy to see that any PBE of the game $(\Gamma, \mathcal{N})$ is also a PBE of the game $(\Gamma, \mathcal{N}^\star)$ where any messages sent on each edge $ij \notin$

[15] It is not clear if these results can directly be applied when utilizing solution concepts other than Nash equilibrium, such as perfect Bayesian equilibrium (as used in this paper) or sequential equilibrium (as used in Gerardi (2004)).

$A(\mathcal{N}^\star) \cap A(\mathcal{N})$ are treated as meaningless by Player $j \in I\backslash\{i\}$.[16] This argument shows that $B(\Gamma, \mathcal{N})$ is weakly monotone increasing in the number of edges of $\mathcal{N}$ and therefore in general we obtain that $B(\Gamma, \mathcal{N}) \subseteq C(\Gamma)$. The first main result of this paper (Theorem 1) provides necessary and sufficient conditions on the topology of the network $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ such that $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for every game Γ with $|I|$ players.

We will now introduce two important definitions regarding the *connectivity* of the network $\mathcal{N}$. For a further understanding of these graph theoretic concepts we refer the reader to Bang-Jensen and Gutin (2002).

Definition 1.4. 2.3 (Strong Connectivity) The network $\mathcal{N} = (V, A(\mathcal{N}))$ is *strongly k-connected from i to j* if there exists k vertex disjoint directed paths *from i to j*.

Definition 1.5. 2.4 (Disjoint Connecting Paths) The network $\mathcal{N}$ is *strongly k-connected from vertex i to vertex j and strongly l-connected from vertex j to vertex i with all $k + l$ connecting paths disjoint* if there are k disjoint paths from i to j, l disjoint paths from j to i, and each of these $k + l$ paths are vertex disjoint.

Finally, for any $x \in [0, 1)$ and $y \in [0, 1)$, we will define addition modulus 1 ($\oplus$) and subtraction modulus 1 ($\ominus$) as:

$$x \oplus y := \begin{cases} x + y & \text{if } x + y < 1 \\ x + y - 1 & \text{if } x + y \geq 1 \end{cases} \qquad x \ominus y := \begin{cases} x - y & \text{if } x - y \geq 0 \\ x - y + 1 & \text{if } x - y < 0 \end{cases}$$

1.3 An Illustration of the Results: The 3-player Case

In this section we will highlight the main results for the 3-player case and illustrate why perfectly secure communication between the mediator and every player, guaranteed by the conditions of Theorem 1, is necessary and sufficient for the implementation of all correlated equilibria of any underlying game. The content of this section solely serves to introduce and illustrate the features of our main results. The reader can feel free to skip directly to Section 4 where the main results are stated.

Figure 8 provides some 3-player networks and shows the relationship between the set of correlated equilibrium outcomes $C(\Gamma)$ of all 3-player games Γ and the set of the perfect Bayesian equilibrium outcomes $B(\Gamma, \mathcal{N})$ of each game Γ augmented by a finite preplay communication phase restricted to the network $\mathcal{N}$. The necessary and sufficient conditions on the network $\mathcal{N}$ such that $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for any 3-player game Γ then precisely translates

[16] This equates to a *partially babbling* equilibrium of the game $(\Gamma, \mathcal{N}^\star)$ where Player $i \in I\backslash\{j\}$ is indifferent between sending any message on the edge $ij \notin A(\mathcal{N}^\star) \cap A(\mathcal{N})$ and therefore in equilibrium optimally does not communicate on those edges.

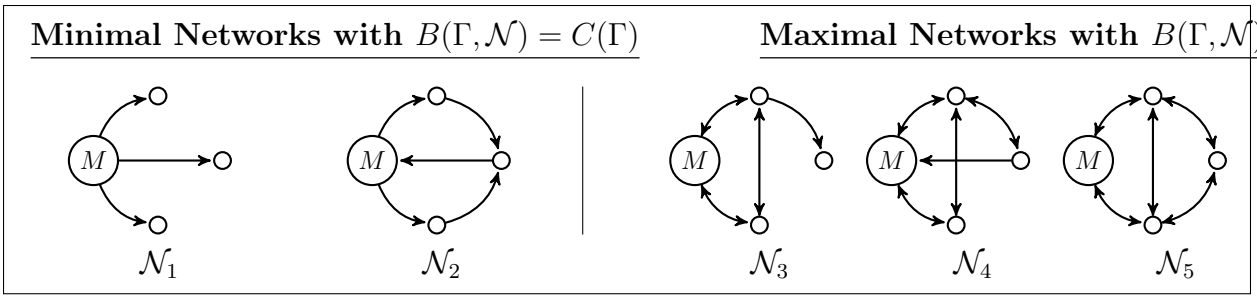


Figure 8: *An illustration of the main results for the 3-player case.*

to the condition that either $\mathcal{N}_1 \subset \mathcal{N}$ or $\mathcal{N}_2 \subset \mathcal{N}$ up to a permutation of the players labels where $\mathcal{N}_1$ and $\mathcal{N}_2$ are the networks in Figure 8.[17]

In the right portion of Figure 8 we have three networks such that there exists a class of games with correlated equilibria that cannot be implemented as a perfect Bayesian equilibrium of any of these games augmented by a finite cheap talk phase restricted to the networks $\mathcal{N}_3$, $\mathcal{N}_4$, and $\mathcal{N}_5$ respectively. In the following subsections we will show why the aforementioned conditions for implementation are sufficient and give two examples of correlated equilibria that cannot be implemented on the network $\mathcal{N}_3$, which does not guarantee secrecy, and the network $\mathcal{N}_4$, which does not guarantee resiliency. In Section 4 we show that the proof of necessity of our main results can be reduced to showing that there exists a 3-player game with a correlated equilibrium that cannot be implemented when communication is restricted to the network $\mathcal{N}_5$.

1.3.1 Sufficiency

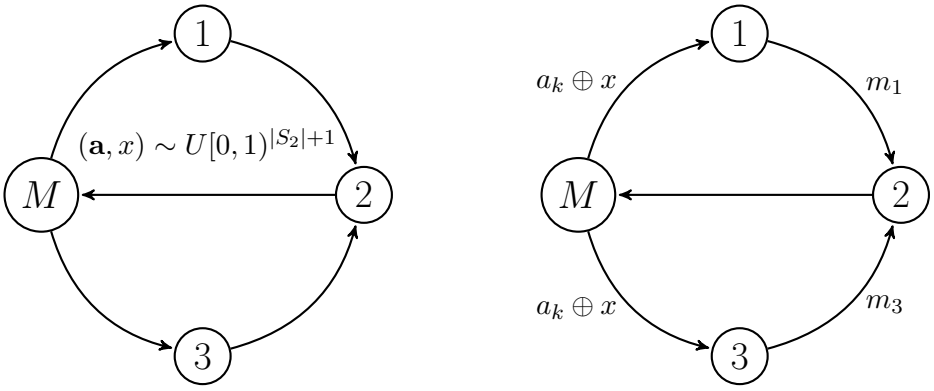
Sufficiency of the condition that $\mathcal{N}_1 \subset \mathcal{N}$ is straightforward as in this case the mediator can implement any correlated equilibrium using the *canonical protocol*: the mediator draws an action profile $\tilde{s} = (\tilde{s}_1, \tilde{s}_2, \tilde{s}_3)$ from the correlated equilibrium distribution Q and sends to each Player $i \in \{1, 2, 3\}$ their component $\tilde{s}_i$ (see the next subsection for a concrete example). Then, the fact that Q is a correlated equilibrium means that it is a best response for all players to play the action $\tilde{s}_i$ that is suggested to them.

Figure 9 illustrates the protocol for implementing any correlated equilibrium of any 3-player game whenever $\mathcal{N}_2 \subset \mathcal{N}$.[18] First, note that $\mathcal{N}_2 \subset \mathcal{N}$ implies that there is only one

[17]For any two networks $\mathcal{N} = (V, A(\mathcal{N}))$ and $\mathcal{N}' = (V', A(\mathcal{N}'))$ with vertex sets V and V' resp., and edge sets $A(\mathcal{N})$ and $A(\mathcal{N}')$ resp., we say that $\mathcal{N} \subset \mathcal{N}'$ if and only if $V \subset V'$ and $A(\mathcal{N}) \subset A(\mathcal{N}')$.

[18]It is worth noting that there exists a much simpler protocol for implementation in 3-player games that we do not present here and which does not rely on *encoding* and *decoding* messages. We chose to present this version of the protocol to provide the reader with an introductory example regarding the techniques that will be used for the general case.

player that the mediator cannot directly communicate with – we suppose that this is Player 2. For the remaining players the mediator can directly send suggested actions as before, but for Player 2 he must utilize a slightly more sophisticated protocol. The protocol proceeds in two steps. **Step 1**: Player 2 encodes her pure actions, $s_2^1, s_2^2, s_2^3, ...$, into the $[0,1)$ interval by drawing a vector $\mathbf{a} := (a_1, ..., a_{|S_2|}) \sim_{i.i.d.} U[0,1)^{|S_2|}$ where S_2 is the set of pure actions of Player 2, $|S_2|$ the cardinality of S_2, and $U[0,1)^{|S_2|}$ the uniform distribution over $[0,1)^{|S_2|}$. In the same step, Player 2 also draws a key $x \sim U[0,1)$ and then sends the vector $(\mathbf{a}, x)$ to the mediator via the edge $2P$ (see Figure 9). **Step 2**: the mediator draws an action profile $\tilde{s}$ from the correlated equilibrium distribution. Whenever $\tilde{s}_2 = s_2^k$ (*i.e.,* the mediator must suggest Player 2 to play her k^{th} pure action) the mediator sends $a_k \oplus x$ to both Player 1 and Player 3 where $\oplus$ is addition modulus 1. The protocol then instructs Player 1 and Player



(a) *Step 1: encode strategies and generate a key.*

(b) *Step 2: Sending the suggested action* s_2^k.

Figure 9: *An illustration of the protocol when* $\mathcal{N}_2 \subset \mathcal{N}$.

3 to forward the message they received from the mediator to Player 2. Denoting m_1 and m_3 as the messages *actually* sent by players 1 and 3 respectively in this previous (cheap talk) step, Player 2 then computes $m_1 \ominus x$ and plays the action s_2^k if $m_1 \ominus x = a_k$ for some $k \in \{1, 2, ..., |S_2|\}$.[19] Otherwise, she computes $m_3 \ominus x$ and plays the action s_2^k if $m_3 \ominus x = a_k$ for some $k \in \{1, 2, ..., |S_2|\}$. Finally, if Player 2 does not *decode* a message corresponding to a suggested action in the previous steps then she randomizes uniformly over her pure actions.

It should be clear that perfectly secure communication from the mediator to every player of the game is a sufficient condition for the implementation of any correlated equilibrium. Namely, any protocol guaranteeing this type of communication leaves the players with the same information as the canonical protocol regarding the strategy profile of the other players,

[19]Note that, based on the encryption techniques utilized in this protocol, we have that $a_k \neq a_l$ for any $k, l \in \{1, 2, ..., |S_2|\}$ such that $k \neq l$ with probability 1. For completeness we could specify that if $a_j = a_k$ for some $j \neq k$ that Player 2 randomize uniformly over a_j and a_k.

and does so even under all unilateral deviations. In light of this, we can guarantee that the protocol illustrated in Figure 9, when coupled with the correct off path belief system, allows us to implement any correlated equilibrium by checking that it satisfies secrecy and resiliency of message transfer from the mediator to Player 2.

The protocol guarantees secrecy due the fact that if $x \in [0,1)$ and $Y \sim U[0,1)$ then $x \oplus Y \sim U[0,1)$ and $x \ominus Y \sim U[0,1)$ (see Lemma 0 in the appendix). Hence, all messages received by players 1 and 3 are uniformly distributed on $[0,1)$ with respect to their information and therefore completely uninformative about the action suggested to Player 2. To show that this protocol also satisfies resiliency, we note that if at Step 2 Player $j \in \{1,3\}$ sends some message $\tilde{m}_j \neq a_k \oplus x$ then given that $x \sim U[0,1)$ implies that $\tilde{m}_j \ominus x \sim U[0,1)$. This implies that whenever Player $j \in \{1,3\}$ deviates, Player 2 decodes $\tilde{m}_j$ and receives a message that, with probability 1, is not an element of the set $\{a_1, ..., a_{|S_2|}\}$. In this case, Player 2 proceeds to decode m_{-j} and receives the true message a_k due to the fact that we only consider unilateral deviations. Finally, we note that whenever Player 2, upon detecting a deviation[20], has the beliefs that all other players will play their suggested actions, then it is sequentially rational for Player 2 to play her suggested action given that Q is a correlated equilibrium distribution. Further, these beliefs are consistent as whenever Player 2 plays her suggested action (which she receives with probability 1 under any unilateral deviation) it is always a best response for Player $j \in \{1,3\}$ to play his suggested action, again due to the fact that the mediator draws this action profile from a correlated equilibrium distribution. Therefore, the resulting strategy induced by this protocol, when coupled with this belief system, constitutes a perfect Bayesian equilibrium of the game Γ augmented by a preplay communication phase restricted to any network $\mathcal{N}$ such that $\mathcal{N}_2 \subset \mathcal{N}$.

1.3.2 Necessity of Secrecy and Resiliency

In this subsection we will illustrate the necessity of secrecy and resiliency with two simple examples.

	A	B
a	6, 6	2, 7
b	7, 2	0, 0

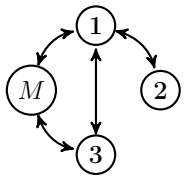


Figure 10: The game Γ_1 and the network $\mathcal{N}_3$.

[20] In this protocol this implies that $m_j \ominus x \notin \{a_1, ..., a_{|S_2|}\}$ for some $j \in \{1,3\}$.

To illustrate the necessity of secrecy, let us first look at the game Γ_1 in Figure 10. There is a well known correlated equilibrium distribution Q_1 of this game where the players play the action profiles (a, A), (b, A) and (a, B) each with equal probability $\frac{1}{3}$. When communication is not restricted, the mediator can implement this correlated equilibrium with the *canonical protocol* where first an action profile s is drawn from the correlated equilibrium distribution, Player 1 is informed about the first component, and Player 2 the second component. For example, if the mediator draws the action profile (b, A) from the correlated equilibrium distribution, the canonical protocol would have him privately suggest to Player 1 to play their action b and privately suggest to Player 2 to play their action A. Q_1 is a correlated equilibrium if it is always optimal for players to play their suggested action in this mechanism. In this example, playing b is optimal for Player 1 as he knows Player 2 will play A with probability 1 whenever he is suggested to play b. Additionally, upon being suggested action A, Player 2 has the beliefs that Player 1 will play a with probability $\frac{1}{2}$ and b with probability $\frac{1}{2}$ and this makes playing A a best response given these posterior beliefs. One can check that the same logic applies to the remaining suggestions and therefore Q_1 is a correlated equilibrium.

What we claim is that *any* communication protocol restricted to $\mathcal{N}_3$ that induces the same outcome as Q_1 has a profitable deviation for Player 1. Namely, looking at the network $\mathcal{N}_3$ we can see that any message sent or received by Player 2 must be sent through Player 1. Hence, for any communication protocol (ρ, σ) and any realization of ρ, the history of Player 2, h_2, is learned by Player 1 with probability 1. Thus, given that the protocol (ρ, σ) is common knowledge (or equivalently players have correct equilibrium beliefs) implies that Player 1 can always use σ_2 and Player 2's history to learn Player 2's suggested action: the network $\mathcal{N}_3$ does not guarantee secrecy of communication between the mediator and Player 2. Therefore, whenever Player 1 is suggested to play his action a and also learns that Player 2 has been suggested to play A, he will have a profitable deviation to play b: Q_1 cannot be implemented on the network $\mathcal{N}_3$.

	A	B
a	3 1 0	0 0 10
b	0 0 10	1 3 0

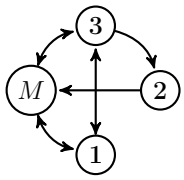


Figure 11: The game Γ_2 and the network $\mathcal{N}_4$.

To illustrate the necessity of resiliency, let us now consider the game Γ_2 in Figure 11. In this game players 1 and 2 want to coordinate perfectly on either (a, A) or (b, B) and Player 3 (who has no actions) wants players 1 and 2 to mis-coordinate on either (a, B) or (b, A). Now consider the simple correlated equilibrium Q_2 of this game where players 1 and 2 perfectly coordinate their actions by playing (a, A) and (b, B) with equal probability. The problem that arises when restricting communication to $\mathcal{N}_4$ is that Player 3 controls all of the information received by Player 2. Therefore, Player 3 can choose to not communicate any information to Player 2 whenever the protocol requires him to[21] so that all messages received by Player 2 are uninformative with regards to the outcome that players 1 and 2 are coordinating on. When Player 3 makes this deviation, there is positive probability that players 1 and 2 mis-coordinate once the communication phase has ended, which makes it profitable for Player 3 *whenever it cannot be deterred.*

Given that $2P$ is an edge in the network $\mathcal{N}_4$ we should check whether or not one could use this channel to deter the afformentioned deviation of Player 3. One solution would have Player 2 randomly draw the outcome (a, A) or (b, B) for the mediator and then send the realization to the mediator who then forwards it to Player 1. It is easy to see that this is not incentive compatible as in this situation Player 2 would always have a profitable deviation to report that the realization was (b, B) as she gets a higher payoff in this case. Instead, we could devise a subprotocol where Player 2 reports that a deviation was made by Player 3 whenever it is detected[22]. Then this subprotocol would have to specify a punishment for Player 3 and this punishment must be deterministic with respect to Player 2's action. Otherwise, Player 3 could make the same deviation as before, again preventing Player 2 from learning her suggested action, and so on and so forth. Therefore, the only options for punishments are to play one of the two pure Nash Equilibria upon a report of a deviation (if Player 2 plays any mixed strategy this gives Player 3 a strictly higher payoff than Q_2). Namely, if Player 2 reports that Player 3 has deviated, the subprotocol must specify that players 1 and 2 play either (a, A) for sure or (b, B) for sure.[23]

To see that neither of these options suffice, note that if the protocol specifies to play (b, B) whenever Player 2 reports a deviation, then Player 2 will always report that a deviation has occurred even when it has not; she gets a strictly higher payoff in this case and the mediator has no way to know that Player 3 did not deviate. On the other hand, if the protocol specifies to play (a, A) whenever Player 2 reports a deviation made by Player 3, then it would never be optimal for Player 2 to make this report, even if Player 3 does deviate. To see why this

[21] Equivalently, whenever instructed to forward some message to Player 2, Player 3 can uniformly draw a message from the message space and send that message instead.

[22] Note that Player 3's deviation is profitable whenever it is not detectable with *any* positive probability.

[23] Player 1 need not even be made aware that the deviation was reported.

is true, suppose that when Player 2 detects a deviation made by Player 3 she reports that no deviation had occurred and plays B. It can be easily checked that when following this strategy, Player 2 obtains an expected payoff of $\frac{3}{2}$ as opposed to the sure payoff of 1 when she reports that the deviation has occurred. Therefore, it will never be optimal for Player 2 to report the deviation in this subprotocol and the deviation of Player 3 will remain profitable. In summary, any communication protocol that gives Player 2 the ability to report a deviation made by Player 3 *and* deters the aforementioned deviation of Player 3 creates a profitable deviation for Player 2: the correlated equilibrium Q_2 cannot be implemented on the network $\mathcal{N}_4$.

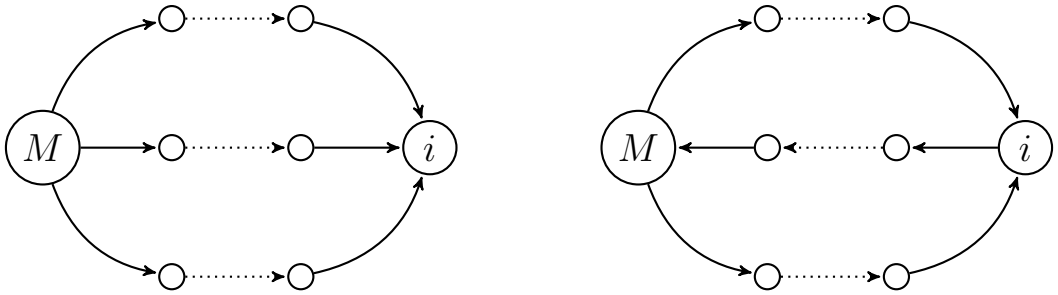
This example illustrates two important complications that arise in the correlated equilibrium implementation problem with restricted communication. First, the network $\mathcal{N}_4$ does not guarantee resiliency of message transfer from the mediator to Player 2. Therefore, Player 3 can always alter the messages in such a way so that Player 2 never receives any information about her suggested action. Second, when considering subprotocols that try to deter this deviation of Player 3, these subprotocols must not create incentives for other players to deviate as they do for Player 2 in the previous example. Our main results show that for similar reasons there exists a game and correlated equilibrium that cannot be implemented on the network $\mathcal{N}_5$. Therefore, any network of the three player game must require an additional edge between M and Player 2 or an edge between Player 2 and M and in either case these conditions are sufficient for implementation of correlated equilibrium and generalize to conditions (1) and (2) respectively of Theorem 1.

1.4 Main Results

We are now ready to present our main results for games of complete information; necessary and sufficient conditions on the network $\mathcal{N}$ such that for any game, any correlated equilibrium of that game can be implemented when communication is restricted to the network $\mathcal{N}$.

Theorem 1.6. *1 Let $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ be a communication network over M and the player set $I = \{1, 2, ..., n\}$. Then, $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for all n-player games Γ if and only if for all $i \in I \backslash Y(M)$, one of the following two conditions holds:*
(1) $\mathcal{N}$ is strongly 3-connected from M to i.
(2) $\mathcal{N}$ is strongly 2-connected from M to i and strongly 1-connected from i to M with all 3 connecting paths disjoint.

These results are illustrated in Figure 12 and follow a similar reasoning as described in Section 3 for the 3-player case. Note here that the dotted arrows of Figure 12 illustrate the



(a) *A network satisfying Condition* (1) *for Player* i.

(b) *A network satisfying Condition* (2) *for Player* i.

Figure 12: *An illustration of conditions* (1) *and* (2) *of Theorem 1.*

fact that if the network satisfies Condition (1) or (2) of Theorem 1, then it does not matter how many players are on each path connecting the mediator to Player i or vice versa.

It is worth noting that our results translate to $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for all 2-player games Γ if and only if the mediator can directly communicate with every player of the game.[24] Similarly, the necessary and sufficient conditions for 3-player games implies that $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for all 3-player games Γ if and only if for any Player i, the network $\mathcal{N}$ be such that either the mediator can directly send messages to Player i or Player i can directly send messages to the mediator. One consequence of this property is that if we are only interested in undirected (*i.e.,* 2-way) communication channels then necessary and sufficient conditions to achieve the set of correlated equilibria for any 3-player game is that the mediator be able to directly communicate with all players of the game. Therefore our results for two and three player games are negative; in organizations with two or three players, the mediator will most likely prefer direct communication. We will now sketch the proofs of Theorem 1 which can be found in the appendix.

Sketch of proof. ($\Rightarrow$) The grand protocol is constructed, in Section 8.2 of the appendix, by constructing two separate subprotocols, $(\rho^{(1)}, \hat{\sigma}^{(1)})$ and $(\rho^{(2)}, \hat{\sigma}^{(2)})$ that allow the mediator to send the suggested action to any Player $i \in I$ in a perfectly secure manner whenever that player satisfies Condition (1) or Condition (2) of Theorem 1 respectively. Although they are too cumbersome to sketch here, the subprotocols $(\rho^{(1)}, \hat{\sigma}^{(1)})$ and $(\rho^{(2)}, \hat{\sigma}^{(2)})$ utilize similar techniques as the protocol constructed for the 3-player case in Section 3.1.

Once we prove that each subprotocol does in fact allow the mediator to communicate with each Player $i \in I$ in a perfectly secure manner, we then construct the grand protocol

[24] This comes from the fact that if the mediator cannot directly communicate with some Player i in a 2-player game Γ, then a necessary condition for $B(\Gamma, \mathcal{N}) = C(\Gamma)$ is that the network be strongly 2-connected from M to i. But given there is only one other player, this says that the mediator must necessarily be able to directly send messages to Player i in the network $\mathcal{N}$.

by having the mediator draw an action profile s from the correlated equilibrium distribution in question and then send the suggested action s_i to each player $i = 1, 2, ..., n$ in sequence utilizing the correct subprotocol for each player. Given that each player learns their strategy in a perfectly secure manner, it is easy to see that if there are no profitable deviations from the subprotocols $(\rho^{(1)}, \hat{\sigma}^{(1)})$ and $(\rho^{(2)}, \hat{\sigma}^{(2)})$ then there must not be any profitable deviations from the grand protocol. Finally, we provide an off-path equilibrium belief system that, when coupled with the grand protocol, constitutes a perfect Bayesian equilibrium of the game $(\Gamma, \mathcal{N})$. □

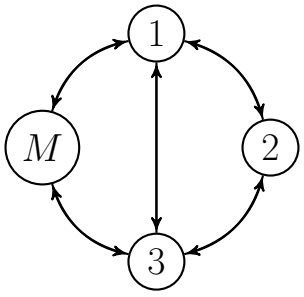
Sketch of Proof. ($\Leftarrow$) The proof of this result comes as corollary of Proposition 1, Corollary 1, and Corollary 2 presented below. Proposition 1, which we will now state and sketch the proof of, provides necessary and sufficient conditions for $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for all 3-player games Γ.

Proposition 1.7. *1 Let $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ be a communication network over M and the set $I = \{1, 2, 3\}$. Then, $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for all 3-player games Γ if and only if for all $i \in I \backslash Y(M)$, $\mathcal{N}$ is strongly 2-connected from M to i and $M \in Y(i)$.*

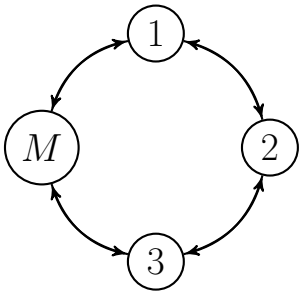
The proof of sufficiency and a network satisfying the conditions of Proposition 1 are illustrated in Section 3.1. We will now give a sketch of the proof of necessity of Proposition 1.

Sketch of proof. ($\Leftarrow$) The objective is to construct a game Γ with CE Q such that there exists no finite protocol that implements Q on the network $\mathcal{N}_0$ of Figure 13(a). The proof the proceeds in three steps.

Step 1: A reduction to implementation on $\mathcal{N}_6$



(a) Network $\mathcal{N}_0$



(b) Network $\mathcal{N}_6$

Figure 13: Implementation on $\mathcal{N}_0$ is equivalent to implementation on $\mathcal{N}_6$

The first step consists of proving an important lemma that allows us to restrict our attention to a subclass of protocols. Lemma 2 in the appendix is proven by construction,

and states that if there exists a protocol that implements some correlated equilibrium on the network $\mathcal{N}_0$ then we can construct another protocol that 1.) also implements the same correlated equilibrium on the network $\mathcal{N}_6$ of Figure 13(b) and 2.) only requires players 1 and 3 to *forward* messages.[25] Therefore, by contraposition, if there exists a CE that cannot be implemented on $\mathcal{N}_6$ then that CE also cannot be implemented on $\mathcal{N}_0$. The next two steps show that it is not possible to implement the correlated equilibrium Q_0 (Figure 15) of the game Γ_0 (Figure 14) on $\mathcal{N}_6$.

35 80 70	80 60 35	80 20 0
72 0 70	78 -10 65	90 -10 20
140 40 55	90 0 65	810 140 40
72 40 55	16 100 65	54 140 15

Figure 14: The 3-player game Γ_0.

0	$\frac{1}{4}$	0
$\frac{1}{2}$	0	0

0	0	0
0	$\frac{1}{4}$	0

Figure 15: The correlated equilibrium Q_0.

Step 2: Any protocol that implements Q_0 on $\mathcal{N}_6$ must adhere to deviations reported by Player 2.

In the game Γ_0 of Figure 14, Player 1 is the row player, Player 2 the column player, and Player 3 the matrix player. In the second step we construct a simple deviation for Player 1 consisting of sending some fixed message m_0 at every stage where he is required to forward a message to the mediator or to Player 2. We then show that whenever Player 1 makes this deviation, any protocol that implements Q_0 on $\mathcal{N}_6$ must come equipped with a subprotocol that allows Player 2 to *report* that Player 1 has made this deviation whenever she detects it (if Player 2 cannot detect this deviation with probability 1, than it is profitable for Player 1.).

[25] The protocol has players 1 and 3 use the deterministic communication strategy: forward any message received from M to Player 2 and forward any message received from Player 2 to M, otherwise do not communicate.

The reason why the protocol must allow Player 2 to *make a report* to the mediator when she detects a deviation follows from similar arguments as given in the examples of Section 3.2. Namely, if whenever Player 1 makes this deviation Player 2 still receives her correct suggested action, then Player 3 must necessarily learn Player 2's suggested action, giving him a profitable deviation whenever the realization of Q_0 is (s_1^1, s_2^2, s_3^1).[26] Hence, if the protocol does not allow Player 2 to report that such a deviation was made then either Player 2 does not correctly learn her suggested action (making the deviation profitable for Player 1), or Player 2 does correctly learn her suggested action but so does Player 3 (creating a profitable deviation for Player 3).

Step 3: There is no way to implement Q_0 on $\mathcal{N}_6$ whenever the protocol allows Player 2 to report deviati

The final step of the proof consists in showing (again similar to the example of Section 3.2) that by allowing Player 2 to report that a deviation has been made, this creates a profitable deviation for all possible continuations of the protocol. Namely, whenever the mediator receives a report from Player 2 that a deviation has been made, he should proceed to implement a new equilibrium distribution $\tilde{Q}$ that punishes Player 1, therefore preventing the deviation from being made. In this case, the distribution $\tilde{Q}$ must be incentive compatible for Player 2 and satisfy the condition that Player 2 receives a weakly lower payoff from $\tilde{Q}$ than from Q_0; otherwise, she could profitably report that a deviation has occurred even if it hasn't. Finally, we show that the unique distribution $\tilde{Q}$ that satisfies all of these conditions is the distribution Q_0 itself. What this means is that, given we restrict protocols to end in finite time, either the deviation of Player 1 will be profitable, or reporting this deviation when it has not occurred will be profitable for Player 2. We would like to note here that there are many subtle details that are not presented in this sketch and we encourage the reader to see the proof in the appendix for a full understanding. □

We will now present the two aforementioned corollaries that, coupled with Proposition 1, allow us to prove the necessity of Theorem 1.

Corollary 1.8. *1 Let $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ be a communication network over M and the set of players $I = \{1, 2, ..., n\}$. If $B(\Gamma, \mathcal{N}) = C(\Gamma)$ for all n-player games Γ then for all $i \in I \backslash Y(M)$, $\mathcal{N}$ is strongly 2-connected from M to i.*

Proof. See appendix. □

Before proceeding we will now introduce another definition regarding the connectivity of the graph $\mathcal{N}$.

[26] This action profile means Player 1 plays the action associated with the first row, Player 2 plays the action associated with the second column, and Player 3 plays the action associated with the first matrix.

Definition 1.9. 4.1 The directed graph $\mathcal{N} = (V, A)$ is *weakly k-connected* between two vertices u and v if there exists k disjoint paths connecting u and v in the underlying undirected graph.

From Corollary 2 we know that strong 2-connectivity from M to $i \in I \backslash Y(M)$ is a necessary condition, but from the proof of Proposition 1 we also know that it is not a sufficient condition (the network $\mathcal{N}_6$ is strongly 2-connected from M to $i \in I \backslash Y(M)$). What the following corollary states is that for any network $\mathcal{N}$, strong 2-connectivity between M and $i \in I \backslash Y(M)$, and weak k-connectivity between M and $i \in I \backslash Y(M)$ is still not sufficient for the implementation of any CE Q on $\mathcal{N}$ for all $k \leq n-1$.

Corollary 1.10. *2 There exists an n-player game Γ with CE Q that is not implementable on any network $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ that is strongly 2-connected from M to $i \in I \backslash Y(M)$, strongly 2-connected from $i \in I \backslash Y(M)$ to M, and weakly k-connected between M and $i \in I \backslash Y(M)$ for any $k \leq n-1$.*

Proof. See appendix. □

To conclude our proof of necessity of Theorem 1 we note that if there exists a correlated equilibrium of an n-player game that cannot be implemented on a network that is strongly 2-connected from M to $i \in I$, strongly 2-connected from $i \in I \backslash Y(M)$, and weakly $n-1$-connected between M and $i \in I \backslash Y(M)$, then in order to implement this correlated equilibrium it must be the case that there is an additional disjoint path from M to i or from i to M. Finally, we note that if the network $\mathcal{N}$ satisfies this additional strong connectivity property, then $\mathcal{N}$ must necessarily satisfy Condition (1) or (2) of Theorem 1 respectively.

1.5 Incomplete Information

In this section, we generalize the results of Theorem 1 to Bayesian games of incomplete information and the set of communication equilibria of these games. Let $G = (I, p, (S_i, \Theta_i, u_i)_{i \in I})$ be a Bayesian game where Θ_i is the set of private types of Player $i \in I$. Then, prior to the start of the game, G, a private type profile $\theta \in \Theta := \times_{i \in I} \Theta_i$, is drawn from the common prior p, and the normal form game with payoffs $(u_i(\cdot|\theta))_{i \in I}$ is played where the beliefs of Player $i \in I$ of type θ_i is that the other players types θ_{-i} are distributed according to the conditional probability distribution $p(\cdot|\theta_i) := \mathbb{P}_p(\cdot|\theta_i)$. We can define analogously $(G, \mathcal{N})$ as the game G extended by an arbitrarily long but finite preplay cheap talk communication phase restricted to $\mathcal{N}$, $B(G, \mathcal{N})$ as the set of distributions over outcomes induced by perfect Bayesian equilibria of $(G, \mathcal{N})$, and $CO(G)$ the set of communication equilibria of G. Again

recall that a *communication equilibrium* of the Bayesian game $G = (I, p, (S_i, \theta_i, u_i)_{i \in I})$ is a mapping $q : \Theta \to \Delta(S)$ such that

$$\sum_{\theta_{-i} \in \Theta_{-i}} \sum_{s \in S} p(\theta_{-i}|\theta_i) q(s|\theta_i, \theta_{-i}) u_i(s|\theta_i, \theta_{-i}) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{s \in S} p(\theta_{-i}|\theta_i) q(s|\tilde{\theta}_i, \theta_{-i}) u_i(\delta_i(s_i), s_{-i}|\theta_i, \theta_{-i})$$

for all $i \in I$, $\theta_i, \tilde{\theta}_i \in \Theta_i$, and $\delta_i : S_i \to S_i$. These conditions simply state that reporting truthfully and playing the action suggested to them, drawn from $q(\cdot|\theta)$, is optimal for each Player $i \in I$ given their posterior beliefs regarding the distribution of s_{-i}.

A communication protocol for the communication equilibrium implementation problem (also precisely defined in the appendix) can be seen as a vector $\mathcal{P}(\mathcal{N}) = ((T_1, \alpha, \theta^M), (T_2, \rho, \hat{\sigma}))$ where T_1 and T_2 are the lengths of the subprotocols (α, θ^M) and $(\rho, \hat{\sigma})$ respectively, $(\rho, \hat{\sigma})$ is interpreted identically as in Section 4, and (α, θ^M) is an analogous strategy for reporting types to the mediator. Namely, one can think of α as a type dependent communication strategy of each player — analogous to ρ — and θ^M as an *interpretation strategy* by the mediator that maps any history of the mediator induced by α to a state of the world $\theta \in \Theta$. Namely, under the interpretation strategy θ^M, upon receiving the time T_1 communication history, $h_M^{T_1}$, the mediator commits to believing that the *reported* state of the world is $\theta^M(h_M^{T_1}) \in \Theta$. We will now introduce our definition of implementation of communication equilibrium on a network.

Definition 1.11. 5.3 (Implementation of Communication Equilibrium): Let $G = (I, p, (S_i, \Theta_i, u_i)_{i \in I})$ be a Bayesian game and $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ a communication network. Then, we say that the set of perfect Bayesian equilibria of the game $(G, \mathcal{N})$ is equal to the set of communication equilibria of G if for every communication equilibrium $q \in CO(G)$ there exists a perfect Bayesian equilibrium $((\alpha, \theta^M), (\rho, \hat{\sigma}))$ of the game $(G, \mathcal{N})$ that induces the same distribution over outcomes as q such that $\mathbb{P}_\alpha(\theta^M = \theta|\theta) = 1$ and $\mathbb{P}_\rho(\hat{\sigma} = s|\theta) = q(s|\theta)$ for all $s \in S$ and $\theta \in \Theta$.

We can now state our main result for communication equilibrium; that strong 2-connectivity from $i \in I \backslash X(M)$ to M and the conditions of Theorem 1 are necessary and sufficient conditions for implementation of all communication equilibrium of any Bayesian game.

Theorem 1.12. *2 Let $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ be a communication network over M and the player set $I = \{1, 2, ..., n\}$. Then $B(G, \mathcal{N}) = CO(G)$ for all n-player Bayesian games G if and only if $\mathcal{N}$ is strongly 2-connected from $i \in I \backslash X(M)$ to M and satisfies the conditions of Theorem 1.*

Proof. See appendix. □

To understand the derivation of the conditions of Theorem 2, first note that if we are interested in implementing some COE q on a network $\mathcal{N}$, then if we take the set of types to be a singleton we are in the case of implementing a CE. Therefore, the conditions of Theorem 1 are still necessary. Furthermore, strong 1-connectivity from i to M for all $i \in I$ is also a necessary condition for the implementation of the set of communication equilibria on $\mathcal{N}$ as, in general, communication equilibria require each Player $i \in I$ to report their type to the mediator. In the proof of Theorem 2 we construct a simple game that shows strong 1-connectivity from $i \in I$ to M is not sufficient regardless of how many disjoint paths connect M to i. This is intuitive as if Player i has only one directed path to report his type to M, then any player on that path must learn Player i's type *and* can prevent M from learning Player i's correct type. Finally, we show that when adding an additional path from $i \in I \backslash X(M)$ to M so that the network is strongly 2-connected from $i \in I \backslash X(M)$ to M and satisfies the conditions of Theorem 1, then we can construct a communication protocol that allows for perfectly secure communication of suggested actions from M to $i \in I$ *and* perfectly secure communication of reported types from $i \in I$ to M.

We will now show that we can weaken the results of Theorem 2 in a special class of Bayesian games with common independent beliefs and private values. We will now define these two assumptions, commonly used in applications such as contract theory or auction design.

Definition 1.13. 5.1 (Common Independent Beliefs) Let $G = (I, p, (S_i, \Theta_i, u_i)_{i \in I})$ be a Bayesian game with common prior p. Then, G has *common independent beliefs* (CIB) if $p(\theta) = \times_{i \in I} p_i(\theta_i)$ for all $\theta \in \Theta$. Namely, p is the product of its marginal distributions.

Definition 1.14. 5.2 (Private Values) Let $G = (I, p, (S_i \Theta_i, u_i)_{i \in I})$ be a Bayesian game. Then, G has *private values* (PV) if $u_i(\cdot|\theta) = u_i(\cdot|\theta_i)$ for all $i \in I$. Namely, players payoffs depend only on their own type.

This brings us to the following corollary of Theorem 1: strong 1-connectivity from $i \in I$ to M and the conditions of Theorem 1 are necessary and sufficient conditions for $B(G, \mathcal{N}) = CO(G)$ for all games G with player set I satisfying PV and CIB.

Corollary 1.15. *3 Let $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ be a communication network over M and the player set $I = \{1, 2, ..., n\}$. Then $B(G, \mathcal{N}) = CO(G)$ for all n-player Bayesian games G with private values and common independent beliefs if and only if $\mathcal{N}$ is strongly 1-connected from $i \in I$ to M and satisfies the conditions of Theorem 1.*

Proof: See appendix. □

While the proof of necessity is a straightforward corollary of Theorem 1, for the proof of sufficiency we use the protocol of Renou & Tomala (2012) for sending reports from each $i \in I$ to M (*i.e.,* the subprotocol (T_1, α, θ^M) is equivalent to the protocol used for sufficiency in Renou & Tomala (2012)). Then, we use the protocol constructed in the proof of Theorem 1 to have the mediator send suggested actions to each Player $i \in I$. We then show that there are no profitable joint deviations from the combined protocol.

Finally, we extend our results to the case of undirected networks with the next corollary. Before proceeding we must introduce a definition of connectivity in undirected graphs.

Definition 1.16. 5.4 Let $\mathcal{N} = (V, A(\mathcal{N}))$ be an undirected network. Then, $v \in V$ and $u \in V$ are k-connected if there are k vertex disjoint paths connecting v and u.

This leads us to our final result.

Corollary 1.17. *4 Let $\mathcal{N} = (I \cup \{M\}, A(\mathcal{N}))$ be an undirected communication network over M and the player set $I = \{1, 2, ..., n\}$. Then, $B(G, \mathcal{N}) = CO(G)$ for all n-player Bayesian games G if and only if $\mathcal{N}$ is such that for all $i \in I \backslash Y(M)$, M and i are 3-connected.*

Proof. See appendix. □

1.6 Applications

The problem of restricted communication studied in this paper is particularly applicable to the situation of information transmission within the multinational enterprise. In such organizations it is common that local firms have private information about the tastes of the their customers within a specific region. Further, each local firm would like to tailor the future product lines to best match to these local tastes while the mediator of the enterprise would like to coordinate product designs as to benefit from economies of scale in the joint production process. Given that it is not feasible to collect data on the true preferences of these customers with regards to future product lines, the transmission of this local information to the national headquarters is best modeled by cheap talk. Further, it is natural to assume that the manager of each individual firm cares solely about its own profits, the manager of a specific region of firms cares solely about that regions profits, and the principal of the enterprise cares about the total profits generated. The insight to this problem provided by this paper is a sufficient condition for the principal of a multinational enterprise to guarantee maximal profits when receiving cheap talk reports about local tastes; that the network of communication satisfy the conditions of Theorem 2. Namely, whenever the network of communication does satisfy these conditions then the multinational enterprise maximizes the number of achievable outcomes for the firm due to the fact that in this case any communication equilibrium of the underlying

game can be implemented over the network of communication. Achieving such stability in a large and growing multinational enterprise is particularly important if the incentives of the local managers changes over time and it is costly to change the communication structure within the firm.

As a more concrete application of our results to multinational organizations we will now briefly discuss how the analysis of cheap talk communication on networks is relevant when considering the efficiency of the banking supervisory and regulatory mechanisms being implemented in response to the financial crisis of 2007-2008. One example of such a mechanism is the Single Supervisory Mechanism (SSM) implemented by the European Central Bank (ECB) whose role is to "to ensure the safety and soundness of the European banking system and to increase financial integration and stability in Europe."[27] The structure of the SSM consists of the ECB, in conjunction with each sovereign national supervisor of the participating member states, monitoring the largest credit institutions in Europe deemed as "significant" by the SSM's criteria. These 120 significant credit institutions will be directly monitored by the ECB, while the remaining banks of Europe will be monitored by their respective national supervisors, who will then report to the ECB. An important component of the SSM is the risk assessment of the eurozone credit institutions; "The SSM risk assessment system (RAS) is rooted in a combination of quantitative indicators and qualitative inputs; *it is not a mere mechanistic approach, but rather leaves room for judgement* guided by clearly defined mediators..." (ECB p.10, emphasis added). Now, while a banks balance sheet is certainly verifiable information to the ECB or its national supervisor, when analyzing a banks counter party or market risk with respect to their portfolio of assets and off balance sheet activities there is a well known asymmetry of information between the supervisor and the bank. This leaves room for misreporting by banks to the SSM and therefore this type of communication is best modeled by cheap talk.

The application of the results in this paper to the SSM lies in the fact that while 120 of the largest credit institutions will be monitored directly by the ECB, the remaining 5,000+ credit institutions of Europe will be monitored indirectly by the ECB via their national supervisory institutions. Further, it is natural that the national supervisory institutions do not have the same preferences as the ECB. Namely, each national supervisory institution cares solely about the welfare of their respective jurisdiction while the ECB's focus is on the eurozone banking system in general. Therefore, in times of crisis if by reporting high risks of certain banks a national supervisory institution puts its jurisdiction under risk of credit contagion, effectively freezing its access to liquidity from other jurisdictions, this could

[27] European Central Bank, Banking Supervision. Sept. 9 2014. https://www.ecb.europa.eu/ssm/html/index.en.html

harm the solvent banks of said jurisdiction. Similarly, in other times one might imagine that a risk adverse national supervisor may want to over exaggerate risks of their banks to ensure that the ECB focuses more resources on their jurisdiction via stress testing and bank recapitalization programs. Such cases highlight the natural incentives that the the national supervisors may have to misreport the true risks of their jurisdictions credit institutions even if the ECB designs the mechanism so that the individual banks prefer to report truthfully. In light of this, the framework of the SSM is naturally modeled by a game of incomplete information with preplay cheap talk restricted to the network where the ECB (acting as the mediator) communicates directly with each of the 120 significant credit institutions, but indirectly (through the national supervisory institutions) with the remaining European credit institutions.

Although it is beyond the scope of this paper to analyze the losses suffered by the ECB by restricting its communication of risks to such a network (which does not satisfy the necessary and sufficient conditions of this paper), there is still insight to be gained. Namely, even though it may be extremely costly for the ECB to directly monitor each of the 6000 or so European credit institutions, the results of this paper shed light on a tradeoff that could otherwise go ignored. Namely, given that the aforementioned network of communication does not satisfy our conditions, one cannot preclude that the incentives of the banks and their national supervisors may induce misreporting of risks through the SSM, undermining its original intent. It is further likely that these incentives will be heightened particularly during times of crisis when the monitoring of risks are most relevant.

1.7 Comments and Conclusion

In this paper we study how the structure of communication can induce players to misbehave in the transmission of private information and suggested actions to and from the mediator. We provide necessary and sufficient conditions on the network of communication such that restricting communication to a network satisfying these conditions does not restrict the set of outcomes that could otherwise be achieved. We believe that our analysis is a necessary baseline in bridging the gap between the literature on costly communication and incentives in organizations. Namely, we highlight, in a general framework, the tradeoff that a principal might face between his own cost of communication and achieving his most preferred outcome. If there is a large cost for the principal of communicating directly and privately with each agent of the organization, then our results state that the principal should select the most cost efficient network among the set of networks satisfying our conditions.

At this point we would like to make some comments about the results of this paper.

First, throughout we assume that players have access to an infinite message space. This is precisely why our solution concept is that of perfect Bayesian equilibrium as opposed to sequential equilibrium. Given that the sequential equilibrium solution concept is not well defined for games with infinite sets of strategies (for further discussion on this issue see Myerson and Reny (2015)) we believe that the perfect Bayesian equilibrium solution concept is more appropriate in this context.

An open question is whether these conditions are necessary and sufficient when we assume that players only have access to a finite message space. It is easy to show that Condition (1) of Theorem 1 is no longer sufficient in this case. Namely, while there is an analog encryption technique for finite message spaces (see the proof of Lemma 1), this encryption can only guarantee that, under any unilateral deviation, a player satisfying Condition (1) in the network $\mathcal{N}$ can receive her suggested action with some probability arbitrarily close but strictly less than 1.[28] In such a case there exists a game and CE that cannot be implemented on a network satisfying Condition (1) for every player (see *e.g.*, the game of Figure 11). It remains an open question whether there exists a perfectly secure protocol that uses a finite message space and allows the mediator to send suggested actions to any player satisfying Condition (2) of Theorem 1. Why it is not clear is due to the fact that whenever a player satisfies Condition (2) she can engage in repeated communication with the mediator. This may allow for the existence of a sophisticated communication protocol utilizing repeated communication between each player and the mediator that still guarantees resiliency.

It can be shown that if in addition to the conditions of Theorem 1, the network $\mathcal{N}$ contains an additional disjoint path from the mediator to each Player $i \in I \backslash Y(M)$ then we can implement any correlated equilibrium of any game on $\mathcal{N}$ with a finite message space as stated formally in the following lemma:

Lemma 1.18. *1 Suppose that we restrict ourselves to communication protocols over finite but large message spaces such that* $|\mathcal{M}| = K \geq \max_{i \in I} |S_i|$. *Then,* $B(\Gamma, \mathcal{N}) = C(\Gamma)$ *for all n-player games* Γ *whenever* $\mathcal{N}$ *satisfies the conditions of Theorem 1 and there exists an additional disjoint path from* M *to each* $i \in I \backslash Y(M)$.

Proof. See appendix. □

The reason why adding an additional disjoint path proves sufficient for the implementation of correlated equilibria with finite message spaces is given by the fact that with 4

[28] Namely, for every $\epsilon > 0$ there exists an integer K and a protocol $(\rho^K, \hat{\sigma}^K)$ such that when players have access to a message space of size K then any player satisfying Condition (1) of Theorem 1 in the network $\mathcal{N}$ receives her correct suggested action with probability $1 - \epsilon$, under any unilateral deviation from the protocol $(\rho^K, \hat{\sigma}^K)$.

disjoint paths the mediator can resort to using certain majority rule decoding protocols that he couldn't otherwise. For example, in the protocol constructed for when the network satisfies Condition (2) of Theorem 1 and has an additional disjoint path from M to i, the second step of the decoding strategy has player i interpret the decoded message to be m' if at least two of the *decoded messages* he has received from the three paths from M to i are equal to m'. Importantly, the protocols constructed for Lemma 1 do not rely on the probability of Player i decoding the wrong suggested strategy under some deviation being equal to 0 (as was the case with infinite message spaces).

While our results do not extend to the case where players only have access to a finite message space, the fact that we can get arbitrarily close to resiliency in the sense described above allows us to partially extend these results; for all $\epsilon > 0$ there exists a finite message space such that the protocol of Theorem 1, using the analog finite encryption techniques, is an ϵ-approximate PBE of the game extended by a cheap talk communication phase restricted to any network satisfying the conditions of Theorem 1.

Finally, one subtle detail of our results is that in the proof of necessity of Proposition 1, our 3-player game is such that Player 2 has three strategies. It remains an open question whether there exists a 2x2x2 game with a correlated equilibrium that cannot be implemented on the network $\mathcal{N}_0$.

1.8 Appendix

We will first state a preliminary result that will be used in constructing protocols and encryption devices.

Lemma 1.19. *0 (Renou and Tomala (2012)):*
1.) For each $(x, y) \in [0, 1) \times [0, 1)$, $(x \oplus y) \ominus y = x$.
2.) Let Y *be a random variable in* $[0, 1)$ *and* $x \in [0, 1)$. *If* Y *is uniformly distributed, then so are* $x \oplus Y$ *and* $x \ominus Y$.
3.) Let X *and* Y *be independent random variables in* $[0, 1)$. *If* Y *is uniformly distributed, then so are* $Z = X \oplus Y$ *and* $W = X \ominus Y$. *Furthermore,* (X, Y, Z) *(resp.,* (X, Y, W)*) are pairwise-independent.*

1.8.1 Cheap Talk Protocols

1.8.1.1 Correlated Equilibrium Protocols

In this section we will start by defining more precisely cheap talk protocols for correlated equilibrium implementation. In what follows, we denote by $m^t_{i\to j} \in \mathcal{M}$ a generic message

sent from Player i to Player $j \in Y(i)$ at time t. Then, a CE protocol restricted to the network $\mathcal{N}$ is defined as the triple $\mathcal{P}(\mathcal{N}) := (T, (\rho_i^t)_{i\in I\cup\{M\}}^{t<T}, (\hat{\sigma}_i)_{i\in I})$ where;

• $T \in \mathbb{N}$ represents T (finite) periods of communication $0, 1, ..., T-1$ such that at each time $t < T$ all players $i \in I$ simultaneously send a single message $m_{i\to j}^t \in \mathcal{M}$, to each $j \in Y(i)$.[29]

• ρ_i^t is the history dependent *communication strategy* for Player $i \in I$ at time t: ρ_i^t is the distribution of the messages $(m_{i\to j}^t)_{j\in Y(i)}$ as a function of Player i's time t history.

• $\hat{\sigma}_i$ is the time T history dependent *action strategy* for Player $i \in I$: players communicate at times $t = 0, 1, ..., T-1$ according to ρ^t and then at time T they play the strategy $\hat{\sigma}$ dependent on their time T communication histories.

Now, given that every Player $i \in I$ sends a single message to each of their successors at each time t implies that every Player $i \in I \cup \{M\}$ also receives a $|X(i)| - vector$ of messages $(m_{j\to i}^t)_{j\in X(i)}$ at each time t. The history h_i^t of Player $i \in I$ at time $t \leq T$ is then defined as the concatenation of the vectors of messages $h_{i\to j}^t := (m_{i\to j}^0, m_{i\to j}^1, ..., m_{i\to j}^{t-1})$ sent from Player i to each of their successors $j \in Y(i)$ and the vector of messages $h_{j\to i}^t := (m_{j\to i}^0, m_{j\to i}^1, ..., m_{j\to i}^{t-1})$ received by Player i from each of their predecessors $j \in X(i)$. Thus, denoting by $\mathcal{M}_{ij}^t$ the set of all possible time t histories of messages sent from $i \in I$ to $j \in I$ up to time $t < T$, we can define the set of all possible time t histories of communication on the network $\mathcal{N}$ as $H^t := \times_{ij\in A(\mathcal{N})}\mathcal{M}_{ij}^t$ and the set of all possible histories H_i^t of Player $i \in I$ at time $t \leq T$ as the projection of H^t onto Player i's time t information.

In what follows we assume that $\mathcal{M}$ is endowed with its Borel σ-algebra and that H^t is endowed with the product topology and σ-algebra. It is further assumed throughout that Player i's strategy is measurable with respect to his information for all $i \in I$. Now, we will denote by $m_{i\to Y(i)}^t := (m_{i\to j}^t)_{j\in Y(i)}$ and $m_{X(i)\to i}^t := (m_{j\to i}^t)_{j\in X(i)}$ as the time t messages sent by Player i to his successors $Y(i)$ and received by his predecessors $X(i)$ respectively in the network $\mathcal{N}$. Then, given a communication strategy profile ρ we can define precisely the communication strategy of Player i at time t as the probability distribution $\rho_i^t : H_i^t \to \Delta(\mathcal{M}^{|Y(i)|})$ such that $\rho_i^t[h_i^t](m_{i\to Y(i)}^t) := \mathbb{P}_\rho(m_{i\to Y(i)}^t | h_i^t)$. Namely, $\rho_i^t[h_i^t](m_{i\to Y(i)}^t)$ is the probability that Player $i \in I$, when following the strategy ρ_i, sends the message composition $m_{i\to Y(i)}^t$ to his successors at time t given his history h_i^t. Additionally, the protocol specifies a communication strategy for the mediator $\rho_M^t : H_M^t \times S \to \Delta(\mathcal{M}^{|Y(M)|})$ such that whenever the mediator draws the action profile $s \in S$ from some correlated equilibrium distribution Q, he communicates at time $t < T$ according to the distribution $\rho_M^t[h_M^t, s]$ given that he has the history h_M^t.

[29] That players only send a single message to each successor at each time $t < T$ is a simplifying assumption and can be made without loss.

Given a communication strategy profile ρ, it will be useful to denote by π_i^t the probability distribution over H_i^t induced by ρ for any $t \leq T$ and $i \in I$. Namely, $\pi_i^t(\tilde{h}_i^t) := \mathbb{P}_\rho(h_i^t = \tilde{h}_i^t)$ is the ex-ante probability according to the communication strategy ρ that Player i receives the history $\tilde{h}_i^t$ at time t.

1.8.1.2 Communication Equilibrium Protocols

Communication equilibrium protocols are similar to correlated equilibrium protocols with an added phase of communication. When implementing a communication equilibrium the mediator must first receive a report of each players' type, prior to sending suggested actions. Hence, we model a COE protocol $\mathcal{P}(\mathcal{N}) := ((T_1, \alpha, R, \theta^M), (T_2, \rho, \hat{\sigma}))$ as two subprotocols. The subprotocol $(T_1, \alpha, R, \theta^M)$ is then defined in the following sense;

• $T_1 \in \mathbb{N}$ represents T_1 finite periods of communication.

• $R := (R_i)_{i \in I}$ is a *type report* for each Player $i \in I$ to the mediator; $R_i : \Theta_i \to \Theta_i$ is a mapping from the true type of Player i, θ_i, to a *reported type* of Player i, $R_i(\theta_i) \in \Theta_i$.

• $\alpha := (\alpha_i)_{i \in I}$ is a type report specific, history dependent, *reporting strategy*; conditional on the type report $R_i(\theta_i)$, the reporting strategy $\alpha_i^t(\cdot|R_i(\theta_i)) : H_i^t \to \Delta(\mathcal{M}^{|Y(i)|})$ is a distribution over $m_{i \to Y(i)}^t$ as a function of Player i's time t history h_i^t (analogous to ρ_i^t).

• $\theta^M : H_M^{T_1} \to \Theta$ is a *type decoder* for the mediator; upon receiving some history $h_M^{T_1} \in H_M^{T_1}$, the mediator deduces that the profile of reported types is $\theta^M(h_M^{T_1}) \in \Theta$.

Finally, upon receiving a report of types $R(\theta)$, the subprotocol $(T_2, \rho, \hat{\sigma})$ is equivalent to the correlated equilibrium protocol mentioned in the previous subsection with length $T = T_2$ and where the mediator draws an action profile from the distribution $q(\cdot|R(\theta))$.

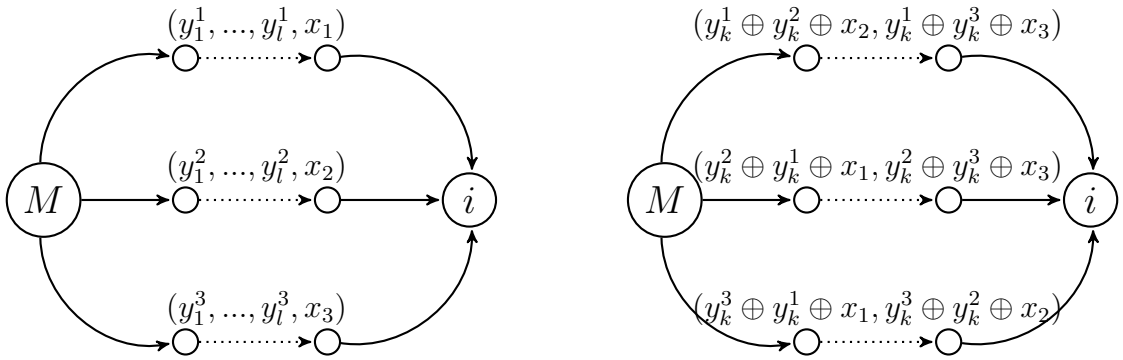
The interpretation of a communication equilibrium protocol is that players receive their private types and then use the communication strategy $\alpha_i(\cdot, R_i)$ to communicate their report to the mediator. Finally, at time T_1 the mediator deduces from θ^M the intended report profile $R(\theta)$, draws an action profile from the communication equilibrium distribution $q(\cdot|R(\theta))$, and proceeds to *suggest* this action profile using the subprotocol $(T_2, \rho, \hat{\sigma})$.

1.8.2 Proof of Theorem 1: Sufficiency

Proof. We will now construct a protocol for implementing all correlated equilibria on any network $\mathcal{N}$ satisfying the conditions of Theorem 1. We will proceed by first constructing two independent and perfectly secure subprotocols for sending the suggested action to each Player $i \in I \backslash Y(M)$ satisfying Condition (1) and Condition (2) in $\mathcal{N}$ respectively. Then, we will construct a *grand protocol* for sending suggested actions to all players and provide

a simple belief system which when coupled with the strategy profile induced by the grand protocol constitutes a PBE of the game $(\Gamma, \mathcal{N})$.

A protocol for $i \in I \backslash Y(M)$ satisfying Condition (1) in $\mathcal{N}$: Here we will construct a communication protocol that allows the mediator to send the suggested action to each Player $i \in I \backslash Y(M)$ satisfying Condition (1) of Theorem 1 in $\mathcal{N}$. Let us assume that $|S_i| = l$ and let $\tilde{s}_i = s_i^k$ be the realization of Player i's suggested action, drawn from the CE distribution in question (*i.e.,* M wants to suggest to Player i to play their k^{th} pure action). Further, denote by $p_1(M,i)$, $p_2(M,i)$, and $p_3(M,i)$ the three disjoint paths connecting M to i (guaranteed to exist in $\mathcal{N}$ by Condition (1) of Theorem 1). The protocol, illustrated in Figure 16 proceeds as follows:



(a) *Step 1 of* $\rho^{(1)}$. (b) *Step 2 of* $\rho^{(1)}$.

Figure 16: *An illustration of the communication strategy* $\rho^{(1)}$.

• Communication strategy $\rho^{(1)}$: **Step 1:** The mediator draws three vectors $\mathbf{y}^1, \mathbf{y}^2, \mathbf{y}^3 \sim U[0,1)^l$ and three keys $x_1, x_2, x_3 \sim U[0,1)$ and then sends the message $\mathbf{c}_j := (\mathbf{y}^j, x_j)$ on each Path $j = 1, 2, 3$, respectively. **Step 2:** the mediator sends the message $\mathbf{m}_1 = (m_1^2, m_1^3) := (y_k^1 \oplus y_k^2 \oplus x_2, y_k^1 \oplus y_k^3 \oplus x_3)$ on the path $p_1(M,i)$, the message $\mathbf{m}_2 = (m_2^1, m_2^3) := (y_k^2 \oplus y_k^1 \oplus x_1, y_k^2 \oplus y_k^3 \oplus x_3)$ on the path $p_2(M,i)$, and the message $\mathbf{m}_3 = (m_3^1, m_3^2) := (y_k^3 \oplus y_k^1 \oplus x_1, y_k^3 \oplus y_k^2 \oplus x_2)$ on the path $p_3(M,i)$. All players on the paths $p_1(M,i)$, $p_2(M,i)$, and $p_3(M,i)$ forward any message received from their predecessors to their successors.

• Action strategy $\hat{\sigma}^{(1)}$: Upon receiving the messages $(\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3)$ and $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)$ Player i begins decoding by computing $m_j^r \ominus x_r$ for $j = 1, 2, 3$ and $r \in \{1,2,3\} \backslash \{j\}$ in increasing sequence.[30] If at any point $m_j^r \ominus x_r = y_k^j \oplus y_k^r$ for some $k \in \{1, ..., l\}$ then Player i stops computing and plays her k^{th} pure action at the end of the communication phase. If for all $j \in \{1,2,3\}$, $r \in \{1,2,3\} \backslash \{j\}$, and $k \in \{1, ..., l\}$ it is the case that $m_j^r \ominus x_r \neq y_k^j \oplus y_k^r$ then Player i randomizes uniformly over her pure actions.

[30] For completeness, we always take x_r to be the last element of the vector c_r and if no message is sent we have Player i interpret $m_j^r \ominus x_r$ as a random draw from $U[0,1)$.

Deviation on Path:	$p_1(M,i):(\tilde{\mathbf{y}}_1,\tilde{x}_1,\tilde{m}_1^2,\tilde{m}_1^3)$	$p_2(M,i):(\tilde{\mathbf{y}}_2,\tilde{x}_2,\tilde{m}_2^1,\tilde{m}_2^3)$	$p_3(M,i):(\tilde{\mathbf{y}}_3,\tilde{x}_3,\tilde{m}_3^1,\tilde{m}_3^2)$
Decoding 1	$\tilde{m}_1^2 \ominus x_2 \sim U[0,1)$	$\tilde{m}_1^2 \ominus \tilde{x}_2 \sim U[0,1)$	$m_1^2 \ominus x_2 = y_k^1 \oplus y_k^2$
Decoding 2	$\tilde{m}_1^3 \ominus x_3 \sim U[0,1)$	$m_1^3 \ominus x_3 = y_k^1 \oplus y_k^3$	$\cdots$
Decoding 3	$\tilde{m}_2^1 \ominus \tilde{x}_1 \sim U[0,1)$	$\cdots$	$\cdots$
Decoding 4	$m_2^3 \ominus x_3 = y_k^2 \oplus y_k^3$	$\cdots$	$\cdots$
Decoding 5	$\cdots$	$\cdots$	$\cdots$
Decoding 6	$\cdots$	$\cdots$	$\cdots$

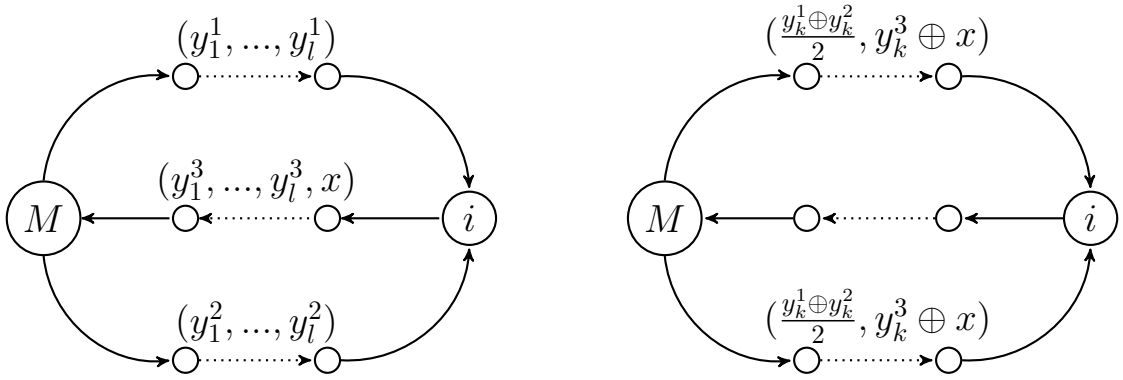
Table 1: Effects of deviations from the communication strategy $\rho^{(1)}$.

Table 1 illustrates the effects of a deviation made by any player on the path $p_1(M,i)$, $p_2(M,i)$, and $p_3(M,i)$ respectively, on the messages received by Player i in the decoding phase. For example, a deviation by some Player j on the path $p_1(M,i)$ can be represented by the vector of messages $(\tilde{\mathbf{y}}_1,\tilde{x}_1,\tilde{m}_1^2,\tilde{m}_1^3)$ *actually* sent by Player j under the communication strategy $\rho^{(1)}$. In this case, once the communication phase has ended Player i begins the first decoding step by computing $\tilde{m}_1^2 \ominus x_2$ and so on. Now, note that Player j's information regarding players 2's history is $(\mathbf{y}^1, x_1, m_1^2, m_1^3)$. Thus, given that Player j does not know $\mathbf{y^2}$, $\mathbf{y^3}$, x_2, nor x_3 implies that all other messages that Player j does not learn are distributed according to $U[0,1)$ with respect to Player j's information (see Lemma 0).[31] Hence, the protocol $\rho^{(1)}$ satisfies secrecy.

Now, if we look at the first step of the decoding by Player i after the deviation $(\tilde{\mathbf{y}}_1,\tilde{x}_1,\tilde{m}_1^2,\tilde{m}_1^3)$ made by Player j we can see that whenever $\tilde{m}_1^2 \neq m_1^2$ that $\tilde{m}_1^2 \ominus x_2 \sim U[0,1)$ given that $x_2 \sim U[0,1)$. This implies that the probability that Player i decodes the message $y_k^1 \oplus y_k^2$ for any $k = 1,...,l$ in this step is zero. Therefore, with probability 1 Player i proceeds to the decoding Step 2. Then, as seen in Table 1, Player i will receive a *meaningless* message (*i.e.,* one that is not equal to an expected message such as $y_k^1 \oplus y_k^2$) in each of the first three decoding steps. Finally, once Player i reaches the decoding Step 4 she will obtain the correct message $y_k^2 \oplus y_k^3$ given that only Player j has deviated. Thus, the decoding phase ends and Player i plays her k^{th} pure action at the end of the communication phase. As highlighted in Table 1, by symmetry, Player 2 always receives the correct suggested action in the decoding phase under any unilateral deviation by any player on any of the remaining paths $p_2(M,i)$ and $p_3(M,i)$. Therefore, we have shown that $\rho^{(1)}$ satisfies both secrecy and resiliency.

We will now proceed to define the protocol for sending suggested actions to Player $i \in I \backslash Y(M)$ satisfying Condition (2) in $\mathcal{N}$, illustrated in Figure 17.

[31] Even if Player j knows y_k^1 all messages utilizing this information are encoded with one of the keys x_2 or x_3 that Player j does not know.



(a) *Step 1 of* $\rho^{(2)}$. (b) *Step 2 of* $\rho^{(2)}$.

Figure 17: *An illustration of the communication strategy* $\rho^{(2)}$.

A protocol for $i \in I \backslash Y(M)$ satisfying Condition (2) in $\mathcal{N}$: Assume again that $|S_i| = l$, $\tilde{s}_i = s_i^k$, and let $p_1(M,i)$, $p_2(M,i)$, and $p_3(i,M)$ be the three disjoint connecting paths guaranteed to exist by Condition (2) of Theorem 1.

• Communication strategy $\rho^{(2)}$: **Step 1:** M draws two vectors $\mathbf{y}^1 := (y_1^1, ..., y_l^1) \sim U[0,1)^l$ and $\mathbf{y}^2 := (y_1^2, ..., y_l^2) \sim U[0,1)^l$ and sends them to Player i along the paths $p_1(M,i)$ and $p_2(M,i)$ respectively. At the same time, Player i draws a vector $\mathbf{y}^3 := (y_1^3, ..., y_l^3) \sim U[0,1)^l$ and a key $x \sim U[0,1)$ and sends $(\mathbf{y}^3, x)$ to M via the path $p_3(i,M)$. All players on the paths $p_1(M,i)$, $p_2(M,i)$, and $p_3(i,M)$ forward any message received from their predecessors to their successors. **Step 2:** Upon M receiving $(\mathbf{y}^3, x)$ and Player i receiving $\mathbf{y}^1$ and $\mathbf{y}^2$, M sends the message $\mathbf{m}_1 = (m_1^1, m_1^2) := (\frac{y_k^1 \oplus y_k^2}{2}, y_k^3 \oplus x)$ to Player i on the path $p_1(M,i)$ and sends the message $\mathbf{m}_2 = (m_2^1, m_2^2) := (\frac{y_k^1 \oplus y_k^2}{2}, y_k^3 \oplus x)$ to Player i on the path $p_2(M,i)$. All players on the paths $p_1(M,i)$ and $p_2(M,i)$ forward any messages received by their predecessors to their successors.

• Action strategy $\hat{\sigma}^{(2)}$: **Step 1:** Upon receiving $\mathbf{m}_1$ and $\mathbf{m}_2$, Player i plays the action s_2^k at the end of the communication phase if $m_1^1 \oplus m_2^1 = y_k^1 \oplus y_k^2$ for some $k \in \{1, 2, ..., l\}$ and otherwise proceeds to Step 2.1. **Step 2.1:** Player i computes $m_1^2 \ominus x$ and plays the action s_2^k at the end of the communication phase if $m_1^2 \ominus x = y_k^3$ for some $k \in \{1, 2, ..., l\}$ and otherwise proceeds to Step 2.2. **Step 2.2:** Player i computes $m_2^2 \ominus x$ and plays action s_2^k at the end of the communication phase if $m_2^2 \ominus x = y_k^3$ for some $k \in \{1, 2, ..., l\}$. If Player i has not learned her suggested action at the end of Step 2.2 then she randomizes uniformly over her pure strategies.

Deviation on Path:	$p_1(M,i):(\tilde{\mathbf{y}}^1,\tilde{m}_1^1,\tilde{m}_1^2)$	$p_2(M,i):(\tilde{\mathbf{y}}^2,\tilde{m}_2^1,\tilde{m}_2^2)$	$p_3(i,M):(\tilde{\mathbf{y}}^3,\tilde{x})$
Decoding 1	$\tilde{m}_1^1 \oplus m_2^1 \sim U[0,1)$	$m_1^1 \oplus \tilde{m}_2^1 \sim U[0,1)$	$m_1^1 \oplus m_2^1 = y_k^1 \oplus y_k^2$
Decoding 2.1	$\tilde{m}_1^2 \ominus x \sim U[0,1)$	$m_1^2 \ominus x = y_k^3$	$\cdots$
Decoding 2.2	$m_2^2 \ominus x = y_k^3$	$\cdots$	$\cdots$

Table 2: Effects of deviations from the communication strategy $\rho^{(2)}$.

To see that there are no profitable deviations from this subprotocol, we illustrate in Table 2 the effects on the information received by Player i resulting from a deviation on each path used during the communication phase $\rho^{(2)}$. Using the same logic as when proving secrecy of $\rho^{(1)}$, we note that Player j's information throughout the protocol is at most $(\mathbf{y}^1, x)$ and therefore the messages m_1^1 and m_1^2 are both distributed according to $U[0,1)$ with respect to Player j's information. Therefore the protocol $\rho^{(2)}$ satisfies secrecy. Next, as you can see from the table, whenever a deviation is made by some player on one of the paths utilized by $\rho^{(2)}$, Player i always decodes the correct message before or at the decoding Step 2.2. For example, suppose that a deviation is made by Player j on the path $p_1(M,i)$, represented by the profile of messages $(\tilde{\mathbf{y}}^1, \tilde{m}_1^1, \tilde{m}_1^2)$. Conditional on Player j making the deviation $(\tilde{\mathbf{y}}^1, \tilde{m}_1^1, \tilde{m}_1^2)$ we can see from Table 2 that in Step 1 of the decoding, Player i computes $\tilde{m}_1^1 \oplus m_2^1 \sim U[0,1)$ and therefore with probability 1 receives an *incorrect* message (*i.e.,* $\mathbb{P}(\tilde{m}_1^1 \oplus m_2^1 = y_{k'}^1 \oplus y_{k'}^2) = 0$ for all $k' \in \{1,...,l\}\backslash\{k\}$). By the same logic Player i receives an incorrect message with probability 1 in the decoding Step 2.1 when Player j makes this deviation. Finally, given that the remaining players play according to $\rho^{(2)}$ implies that at the decoding Step 2.2 Player i receives the message $m_2^2 \ominus x = y_k^3$ and plays the correct suggested action at the end of the communication phase. Hence, the protocol $\rho^{(2)}$ satisfies secrecy and resiliency with respect to deviations on the path $p_1(M,i)$ and by symmetry also satisfies secrecy and resiliency with respect to deviations on the path $p_2(M,i)$.

Lastly, we should check that there are no profitable deviations by some Player j on the path $p_3(i,M)$. First, given that all messages sent from Player i to M on the path $p_3(i,M)$ are distributed according to $U[0,1)$ with respect to the suggested action to be sent to Player i, the protocol $\rho^{(2)}$ trivially satisfies secrecy with respect to players on the path $p_3(i,M)$. Next, we can see from Table 2 that under any deviation $(\tilde{\mathbf{y}}^3, \tilde{x})$ of some Player j on the path $p_3(i,M)$, Player i receives the correct suggested action in the first decoding step. This comes from the fact that none of the information $(\tilde{\mathbf{y}}^3, \tilde{x})$ is used in this step. Hence, the protocol $\rho^{(2)}$ satisfies secrecy and resiliency with respect to deviations on the path $p_3(i,M)$.

We will now construct the grand protocol that combines each of these subprotocols to send suggested actions to each Player $i \in I$.

The grand protocol $\mathcal{P}^\star(\mathcal{N}) = (T^\star, \rho^\star, \hat{\sigma}^\star)$: M draws an action profile $\tilde{s}$ from the CE distribution Q in question. Then, for $i = 1, 2, ..., n$ the protocol proceeds as follows:

1.) If $i \in I \backslash Y(M)$ and i satisfies Condition (1) in $\mathcal{N}$ then M sends the suggested action $\tilde{s}_i$ to Player i via the subprotocol $(\rho^{(1)}, \hat{\sigma}^{(1)})$.

2.) If $i \in I \backslash Y(M)$ and i satisfies Condition (2) in $\mathcal{N}$ then M sends the suggested action $\tilde{s}_i$ to Player i via the subprotocol $(\rho^{(2)}, \hat{\sigma}^{(2)})$.

3.) If $i \in Y(M)$ then M sends $\tilde{s}_i$ to Player i directly via the path Pi.

All that is left now is to provide a system of beliefs which when coupled with the strategy $(\rho^\star, \hat{\sigma}^\star)$ constitutes a perfect Bayesian equilibrium of the game $(\Gamma, \mathcal{N})$ whenever $\mathcal{N}$ satisfies the conditions of Theorem 1. First, let us note that during the phase of the protocol $(\rho^\star, \hat{\sigma}^\star)$ where the mediator sends the suggested action to Player i, every message received by some Player j who is required to forward messages in this phase is on the equilibrium path; by construction the support of all messages sent is $[0, 1)$. Thus, the players $i \in I \backslash Y(M)$ are the only ones who can detect that there has been some deviation from the protocol, and they can only detect this deviation when they are decoding their suggested actions. For example, in the subprotocol $(\rho^{(2)}, \hat{\sigma}^{(2)})$ this translates to Player i computing $\tilde{m}_2^1 \oplus m_2^1 \neq y_k^1 \oplus y_k^2$ for all $k \in \{1, 2, ..., l\}$ in the first decoding step. Hence, we only need to provide a system of beliefs for each Player $i \in I \backslash Y(M)$ conditional on Player i decoding a message that is not a suggested action during each decoding step of each protocol. One simple belief system suffices for both subprotocols: upon detection of a deviation, Player $i \in I \backslash Y(M)$ believes that all players will play their suggested actions.

Now, we can easily show that this belief system is sequentially rational with respect to the strategy profile $(\rho^\star, \hat{\sigma}^\star)$ as in this case each Player $i \in I \backslash Y(M)$ has the belief that the strategy profile being played by the other players is distributed according to $Q_{-i}(\cdot|\tilde{s}_i)$. Thus, given that Q is a correlated equilibrium of the game Γ, it is sequentially rational for Player $i \in I \backslash Y(M)$ to play her suggested action $\tilde{s}_i$ when she has these beliefs. Finally, given that Player $i \in I \backslash Y(M)$ receives her correct suggested action with probability 1, even after detecting a deviation, it is easy to see that these beliefs are consistent. Namely, for any deviation of some Player j, in either subprotocol for sending the suggested action to Player i, Player j knows under this belief system that Player i will still play her correct suggested action. Hence, Player j has the equilibrium beliefs that the strategy profile of the other players is distributed according to $Q_{-j}(\cdot|\tilde{s}_j)$ and therefore it is optimal for Player j to play

his suggested action. Thus, the above belief system is consistent and sequentially rational and we have proven that $(\rho^\star, \hat{\sigma}^\star) \in B(\Gamma, \mathcal{N})$ whenever the network $\mathcal{N}$ satisfies the conditions of Theorem 1. □

The proof of necessity of Theorem 1 is outlined in Section 4 as a consequence of Proposition 1, Corollary 1, and Corollary 2 which we will now prove.

1.8.3 Proof of Proposition 1

Proof. ($\Rightarrow$) Any network $\mathcal{N}$ that satisfies the condition of Proposition 1 also satisfies Condition (2) of Theorem 1. Therefore we direct the reader to the proof of sufficiency of Theorem 1. For a sketch of a simpler protocol that can be utilized in the 3-player case see Section 3.1. □

Proof. ($\Leftarrow$) The objective here is to find a 3-player game Γ and a CE Q of Γ such that there exists no protocol that implements Q on the network $\mathcal{N}_0$ of Figure 13 satisfying the strong 2-connectivity assumption of Proposition 1, but with $M \notin Y(2)$. Namely, we will show that there exists a game Γ_0 with CE Q_0 such that any strategy $(\rho, \hat{\sigma})$ that satisfies $\mathbb{P}_\rho(\hat{\sigma} = s) = Q(s)$ for all $s \in S$ is such that $(\rho, \hat{\sigma}) \notin B(\Gamma, \mathcal{N}_0)$. The proof proceeds in four steps.

Step 1: As a first step we will prove a lemma which states that if there exists a protocol $\mathcal{P}(\mathcal{N}_0)$ that implements some correlated equilibrium Q on the network $\mathcal{N}_0$, then there exists another protocol $\mathcal{P}(\mathcal{N}_6)$ that implements Q on the network $\mathcal{N}_6$ and satisfies the condition that players 1 and 3 only forward messages between the mediator and Player 2.

Lemma 1.20. *2 If there exists a protocol $\mathcal{P}(\mathcal{N}_0)$ that implements some correlated equilibrium Q on the network $\mathcal{N}_0$, then there exists a protocol $\mathcal{P}(\mathcal{N}_6)$ that implements Q on the network $\mathcal{N}_6$ where players 1 and 3 only forward messages (i.e., for all $t < T$, $m^t_{i\to j} = m^{t-1}_{k\to i}$ for all $i \in \{1,3\}$ and $j,k \in \{2,M\}$ with $j \neq k$).*

Proof. Let $(T, \rho, \hat{\sigma})$ be a protocol that implements some CE Q on $\mathcal{N}_0$. We will now construct a protocol from $(\rho, \hat{\sigma})$ that implements Q on $\mathcal{N}_6$ and only requires players 1 and 3 to forward messages between M and Player 2. In order to do this we will show that for any $t < T$ we can construct a communication strategy $\hat{\rho}$ where $\hat{\rho}^{t'} = \rho^{t'}$ for all $t' < t$ and such that $\hat{\rho}^{t'}$ satisfies the aforementioned conditions that Player 1 and Player 3 do not communicate with each other and only forward messages between M and Player 2 for all $t' \geq t$. Further, we will prove that we can construct $\hat{\rho}$ satisfying these conditions for any $t < T$ and that there exists an action strategy $\hat{\hat{\sigma}}$ and $\hat{T}$ such that $(\hat{T}, \hat{\rho}, \hat{\hat{\sigma}})$ implements Q on $\mathcal{N}_6$.

To achieve this construction first we will add $2(T-t)+1$ additional periods to the communication protocol $\mathcal{P}(\mathcal{N}_0)$ consisting of a *pre*-time t communication period and a *post*-time t communication period. We will denote the periods of $\hat{\rho}$ by $(t-1)^+, t^-, t, t^+, ...(T-1)^-, T-1, (T-1)^+$ where k^- and k^+ are the *pre* and *post* time k communication periods respectively. For the sake of simplicity, we assume in the construction that whenever a player's communication strategy is not specified for a particular period that that they do not communicate in that period.

Construction of $\hat{\rho}$: Fix some $t < T$. Then, for any time t history h^t of the protocol $\mathcal{P}(\mathcal{N}_0)$ we construct the communication strategy $\hat{\rho}^{t'}$ for $t' \geq t$ as follows:
• At time $(t-1)^+$: Each Player $i \in \{1,3\}$ sends their history h_i^t to the mediator.
• At time t^-: For each $i \in \{1,3\}$ the mediator draws $m^t_{i\to Y(i)}$ from the distribution $\rho^t_{i\to Y(i)}[h_i^t]$ and forwards to Player i the message pair $(m^t_{i\to Y(i)}, m^t_{j\to i})$ where $\{j\} = \{1,3\}\backslash\{i\}$.
• At time t: Each Player $i \in \{1,3\}$ forwards to Player 2 the message $m^t_{i\to 2}$ sent to them by M in the previous step. Player 2 communicates according to $\rho^t_{2\to Y(i)}[h_2^t]$. M draws $m^t_{M\to Y(M)}$ according to $\rho^t_{M\to Y(M)}[h_M^t]$ and forwards the message $m^t_{M\to i}$ to each Player $i \in \{1,3\}$.
• At time t^+: Each Player $i \in \{1,3\}$ forwards to the mediator the message $m^t_{2\to i}$ that they received in the previous period from Player 2.

Now, for each $i \in \{1,3\}$ let us denote by $\hat{h}_i^{t+1}$ the *pseudo history* of Player i consisting of the concatenation of the history h_i^t and the vector of messages $(m^t_{i\to Y(i)}, m^t_{M\to i}, m^t_{2\to i})$ where $m^t_{i\to Y(i)}$ is the vector of messages that M drew at time t^-, $m^t_{M\to i}$ the respective component of the vector that M drew at time t, and $m^t_{2\to i}$ the message that Player i forwarded to M at time t^+. Similarly, denote by $\hat{h}_M^{t+1}$ the pseudo history consisting of the concatenation of the history h_M^t and the vector of messages $(m^t_{1\to M}, m^t_{3\to M}, m^t_{M\to Y(M)})$ where $m^t_{1\to M}$ and $m^t_{3\to M}$ are the messages drawn by M at time t^- and $m^t_{M\to Y(M)}$ the vector of messages drawn by M at time t. The construction then continues as follows:

• At time $(t+1)^-$: For each $i \in \{1,3\}$ the mediator draws $m^{t+1}_{i\to Y(i)}$ from the distribution $\rho^t_{i\to Y(i)}[\hat{h}_i^{t+1}]$ and forwards to Player i the message pair $(m^{t+1}_{i\to Y(i)}, m^{t+1}_{j\to i})$ where $\{j\} = \{1,3\}\backslash\{i\}$.
• At time $t+1$: Each Player $i \in \{1,3\}$ forwards to Player 2 the message $m^{t+1}_{i\to 2}$ sent to them by M in the previous step. Player 2 communicates according to $\rho^{t+1}_{2\to Y(i)}[h_2^{t+1}]$. M draws $m^t_{M\to Y(M)}$ according to $\rho^t_{M\to Y(M)}[\hat{h}_M^{t+1}]$ and forwards the message $m^t_{M\to i}$ to Player $i \in \{1,3\}$.
• At time $(t+1)^+$: Each Player $i \in \{1,3\}$ forwards to the mediator the message $m^{t+1}_{2\to i}$ that they received in the previous period from Player 2.

Now, if we analogously define $\hat{h}_i^{t+2}$ and $\hat{h}_M^{t+2}$ as the concatenation of the time $t+1$ pseudo histories and the respective vector of messages sent/recieved by each player at times $(t+1)^-$, $t+1$, and $(t+1)^+$, then by following the above three steps in the same fashion we can extend this communication until time $(T-1)^+$. Lastly, assume that at time $(T-1)^+$ the mediator sends directly to each Player $i \in \{1,3\}$ their strategy suggestion via the following encoding: if $\tilde{s}_i = s_i^k$ then at time $(T-1)^+$ have M send $m_{M\to i}^{(T-1)^+} = \frac{1}{1+k}$ to Player i. We call $\hat{\rho}$ the communication strategy resulting from this construction.

Lastly, we must define $\hat{\hat{\sigma}}$. For each Player $i \in \{1,3\}$ let $\hat{\hat{\sigma}}_i : H_i^T \to S_i$ be the mapping defined by $\hat{\hat{\sigma}}_i(h_i^T) = \hat{\hat{\sigma}}_i(m_{M\to i}^{(T-1)^+})$ where $\hat{\hat{\sigma}}_i(\frac{1}{1+k}) = s_i^k$ for all $k = 1, ..., |S_i|$. Further, define $\hat{\hat{\sigma}}_2(h_2^T)$ as the same mapping $\hat{\sigma}_2$ only that $\hat{\hat{\sigma}}_2(h_2^T)$ ignores the $2(T-t)+1$ extra periods added to the communication strategy ρ in order to construct $\hat{\rho}$. Given that each of these periods do not involve any communication involving Player 2 this construction is straightforward and no strategic considerations need to be taken.

Based on the above construction, players 1 and 3 do not communicate with each other after time $(t-1)^-$ and only forward messages between M and Player 2. To clarify, this new protocol has the mediator draw the messages players $i \in \{1,3\}$ would have sent at each time in the protocol $\mathcal{P}(\mathcal{N}_0)$, given the time t history h_i^t that each Player $i \in \{1,3\}$ sends to M at time $(t-1)^+$ and forwards the respective messages to each player. Hence, the distribution of histories does not change as every message is still drawn according to the communication strategy ρ, but the construction now has M draw these messages for each Player $i \in \{1,3\}$. Further, the mediator informs each Player $i \in \{1,3\}$ of the vector of messages $m_{i\to Y(i)}^t$ that he drew for them and the message $m_{j\to i}^t$ that Player $j = \{1,3\}\backslash\{i\}$ would have sent them directly. Therefore, each Player $i \in \{1,3\}$ has the same information as he would have in the protocol $\mathcal{P}(\mathcal{N}_0)$ whenever they are required to forward messages. The protocol then proceeds to have each player send to the mediator the message they receive from Player 2 at time t, and the process repeats with the time $t+1$ pseudo history consisting of the reports of the messages received by players $i \in \{1,3\}$ (including h_i^t) and the messages that M *suggests* the players to forward at time $(t+1)^-$. Now, all that is left to prove is that there are no profitable deviations from the protocol $(\hat{T}, \hat{\rho}, \hat{\hat{\sigma}})$.

We will show that *if* there exists a profitable deviation in the protocol $\mathcal{P}(\mathcal{N}_6) = (\hat{T}, \hat{\rho}, \hat{\hat{\sigma}})$ then there exists an equivalent deviation that must also be profitable in the protocol $\mathcal{P}(\mathcal{N}_0)$. First, we know there are no possible deviations at all pre-time k communication periods k^- for all $t \leq k < T$ as M is the only one required to communicate at this stage. Further, note that if players 1 and 3 adhere to the protocol then the pseudo histories $\hat{h}_{i\to 2}^k \sim \pi_{i\to 2}^k(\cdot|h_i^k)$ for all $t < k < T$ and $i \in \{1,3\}$. This comes from the fact that whenever players 1 and

3 correctly forward all messages sent to them, then M draws the messages sent by each $i \in \{1,3\}$ according to the distribution $\rho^k_{i\to 2}$. In this case each $i \in \{1,3\}$ communicates according to the strategy ρ_i, except they have the mediator draw the messages from ρ_i from time t on and forward them the realization. One implication of this is that if Player 2 has a profitable unilateral deviation in the protocol $\mathcal{P}(\mathcal{N}_6)$, then clearly this deviation must be profitable in the protocol $\mathcal{P}(\mathcal{N}_0)$ as whenever she makes this deviation, Player 1, Player 3, and M all communicate *virtually* the same as in the protocol $\mathcal{P}(\mathcal{N}_0)$.

Now, by contradiction, suppose that Player $i \in \{1,3\}$ has a profitable deviation from the protocol $\mathcal{P}(\mathcal{N}_6)$ and that this deviation is not profitable in the protocol $\mathcal{P}(\mathcal{N}_0)$. Then, given that each Player $i \in \{1,3\}$ is only required to communicate at times k (by forwarding messages to Player 2) and k^+ (by forwarding messages to M), we can separate this deviation into two communication strategies $(d^k_{i\to 2}[h^k_i])_{t\leq k<T}$ and $(d^{k^+}_{i\to M}[h^k_i])_{t\leq k<T}$. We will first show that for every deviation $(d^{k^+}_{i\to M})_{t\leq k<T}$ of Player $i \in \{1,3\}$, whereby Player i forwards to the mediator the incorrect message that they received by Player 2 at time k, there exists an equivalent deviation where Player i follows the protocol $\mathcal{P}(\mathcal{N}_6)$ at each time k^+. Namely, suppose that upon receiving the history h^k_i Player i sends some message other than the message, $m^k_{2\to i}$, forwarded to them by Player 2 in the prior period. Then, the only effect this deviation has on the protocol $\mathcal{P}(\mathcal{N}_6)$ is that now Player i's time $k+1$ message, to be forwarded to Player 2, is distributed according to $\rho^{k+1}_{i\to 2}[h^{k+1}_{i\to Y(i)}, h^{k+1}_{X(i)\setminus\{2\}\to i}, m^d]$ where m^d is the realization of $d^{k^+}_{i\to M}[h^k_i]$. Thus, if Player i follows the protocol and sends the correct message $m^t_{2\to i}$ to M at time k^+, he can still induce the same outcome as he would have with the deviation $d^{k^+}_{i\to M}[h^k_i]$ by sending at time $k+1$ a message $m^{k+1}_{i\to 2} \sim \rho^{k+1}_{i\to 2}[h^{k+1}_{i\to Y(i)}, h^{k+1}_{X(i)-\{2\}\to i}, m^d]$ instead of forwarding the message sent to him at time $(k+1)^-$ by M. Therefore, it is with out loss to restrict attention to deviations of the form $(d^k_{i\to 2}[h^k_i])_{t\leq k<T}$.

Now, note that if there exists a profitable deviation $(d^k_{i\to 2}[h^k_i])_{t\leq k<T}$ from the protocol $\mathcal{P}(\mathcal{N}_6)$, then this deviation also exists in the protocol $\mathcal{P}(\mathcal{N}_0)$ as based on the construction of $\mathcal{P}(\mathcal{N}_6)$, deviations of the form $(d^k_{i\to 2}[h^k_i])_{t\leq k<T}$ only involve Player i sending messages to Player 2 according to some distribution other than $(\rho^k_{i\to 2}[h^k_i])_{t\leq k<T}$. This simply amounts to Player i incorrectly forwarding the time $k \geq t$ messages sent to them by M. Hence, given that Player $i \in \{1,3\}$ has the same information when forwarding messages to Player 2 at time $k \geq t$ in the protocol $\mathcal{P}(\mathcal{N}_6)$ as he does in the protocol $\mathcal{P}(\mathcal{N}_0)$, then any deviation of the form $(d^k_{i\to 2}[h^k_i])_{t\leq k<T}$ from the protocol $\mathcal{P}(\mathcal{N}_6)$ is also available in the protocol $\mathcal{P}(\mathcal{N}_0)$ contradicting the fact that $\mathcal{P}(\mathcal{N}_0)$ implements Q.

Now, by the above logic, if Player $i \in \{1,3\}$ has a profitable deviation where they forward to the mediator some time t history $\hat{h}^t_i \neq h^t_i$ at time $(t-1)^+$ in the protocol $\mathcal{P}(\mathcal{N}_6)$ then there exists a profitable deviation in the protocol $\mathcal{P}(\mathcal{N}_0)$ where upon receiving the history

h_i^t Player i communicates according to $\rho_i^t[\hat{h}_i^t]$ from that time on. Hence, any deviation in the protocol $(\hat{\rho}, \hat{\hat{\sigma}})$ is also available in the protocol $(\rho, \hat{\sigma})$ and therefore cannot be profitable. Finally, given that this construction exists for any $t < T$ we can simply take $t = 0$ and we arrive at our claim.

□

Lemma 2 tells us that, by contraposition, if there exists a CE Q that cannot be implemented on $\mathcal{N}_6$ where players 1 and 3 only forward messages between M and Player 2, then Q cannot be implemented on $\mathcal{N}_0$. Now, given that $\mathcal{N}_0$ is the network obtained by taking the complete network, where all players can communicate directly, and removing the edges $P2$ and $2P$, it can be seen that by adding either of these two missing edges to the network $\mathcal{N}_0$ that we will satisfy the conditions of Proposition 1. Therefore, proving that there exists a Q that cannot be implemented on $\mathcal{N}_6$ (and therefore on $\mathcal{N}_0$) is enough to prove that the conditions of Proposition 1 are necessary. Further, Lemma 2 tells us that throughout this proof we can restrict our attention without loss to protocols where players 1 and 3 only communicate by forwarding messages between M and Player 2.

Now, consider the game Γ_0 and CE Q_0 of figures 14 and 15 of Section 4 respectively. Here, Player 1 is the row player, Player 2 the column player, and Player 3 the matrix player. We will now show that there exists no protocol that implements Q_0 on $\mathcal{N}_6$. By contradiction, suppose there exists a protocol $\mathcal{P}(\mathcal{N}_6) = (T, \rho, \hat{\sigma})$ that implements Q_0 on $\mathcal{N}_6$. Then, there exists no profitable deviations by any players in the communication phase of $\mathcal{P}(\mathcal{N}_6)$. The next three steps will outline the conditions that the protocol must satisfy in order to prevent a simple deviation that we will now define.

Definition 1.21. A.1 Let D_1 denote the following deviation by Player 1: whenever Player 1 is required by the protocol $\mathcal{P}(\mathcal{N}_6)$ to forward some message m^t at time t, he instead forwards $\tilde{m}^t = m_0$, where m_0 is some fixed message for all $t = 0, 1, ..., T-1$.

In what follows, we will resort to the notation that $\hat{\sigma}_2(h_2^t) = \emptyset$ whenever $h_2^t \notin supp(\pi_2^t(\cdot))$ and we will utilize the terminology that *Player 2 does not learn her suggested action* whenever $\hat{\sigma}_2(h_2^T) = \emptyset$. Now, for any realization of the protocol where Player 1 makes the deviation D_1 note that there are three possible outcomes: (i) Player 2 learns the correct suggested action $(\hat{\sigma}(h_2^T) = \tilde{s}_2)$, (ii) Player 2 learns an incorrect suggested action $(\hat{\sigma}(h_2^T) = s_2^l \neq \tilde{s}_2)$, and (iii) Player 2 does not learn any suggested action $(\hat{\sigma}(h_2^T) = \emptyset)$. Denote by λ_1 the ex-ante probability that Player 2 learns the correct suggested action, λ_2 the ex-ante probability that Player 2 learns the incorrect suggested action, and $1 - \lambda_1 - \lambda_2$ the ex-ante probability that Player 2 does not learn any suggested action, given that Player 1 makes the deviation D_1. Note that the protocol $\mathcal{P}(\mathcal{N}_6)$ determines the probabilities λ_1 and λ_2, and what happens

whenever $\hat{\sigma}_2(h_2^T) = \emptyset$. Therefore, if $\mathcal{P}(\mathcal{N}_6)$ implements Q on $\mathcal{N}_6$, it must be the case that λ_1, λ_2, and the strategy profile (*i.e.,* punishment) that results whenever Player 2 does not learn any suggested action, are such that the unilateral deviation D_1 is not profitable for Player 1. In what follows we will show for any values of λ_1 and λ_2, and any strategy profile $\tilde{Q}$ resulting when $\hat{\sigma}_2(h_2^T) = \emptyset$, that there always exists a profitable deviation from $\mathcal{P}(\mathcal{N}_6)$. In order to proceed, we introduce the following notation and definitions.

Let us denote by $h_2^{t,1}$ a generic time t history of Player 2 given Player 1 has made the deviation D_1 up to time t and all other players have followed the protocol. Further, we denote by $[h_i^T]_0^t$ as the first t periods of the history h_i^T, which itself is a time t history of Player i. Finally, as an abuse of notation we write $\hat{\sigma}_i(\tilde{h}_i^t) = \emptyset$ for some history $\tilde{h}_i^t$ to mean that $\hat{\sigma}_i(h_i^T) = \emptyset$ for all h_i^T such that $[h_i^T]_0^t = \tilde{h}_i^t$. We will now proceed to introduce some new definitions necessary to continue the proof.

Definition 1.22. A.2 (a): **a.)** Player 2 *detects the deviation* D_1, if $\hat{\sigma}_2(h_2^{t,1}) = \emptyset$ and $h_{1\to 2}^{t,1} \notin supp(\pi_2^t(\cdot|h_{2,-1}^{t,1}, h_{2\to 1}^{t,1}))$ for some $t \leq T$.

We will now introduce one simplifying assumption which we will then show can be made without loss of generality in Step 4 below.

Assumption 1.23. 1: Player 2 can report D_1 and M can learn of this report with probability 1: For all $t < T-1$ there exists a communication strategy $(\rho_2^{r_1,l})_{t<l<T}$ of Player 2 that satisfies the following conditions: For all $l > t$, let $\pi^{r_1,l}$ be the distribution over time l histories induced by the communication strategy $(D_1, (\rho_2^{t'})_{t'\leq t}, (\rho_2^{r_j,t'})_{t'>t}, \rho_3)$. Further, for all $l > t$ let $\hat{\pi}^l$ be the distribution over time l histories induced by the communication strategy (D_1, ρ_2, ρ_3). Then, $supp(\pi_M^{r_1,T}) \cap supp(\hat{\pi}_M^T) = \emptyset$.

This assumption states that there exists a communication strategy of Player 2 that allows her to *report* to the mediator that she has detected the deviation D_1 and that the mediator learns of this report with probability 1 whenever it is made. Note here that such reports are independent of the universal message space, and could be any communication strategy satisfying the above conditions when coupled with the correct equilibrium beliefs of Player 2 and the mediator.

Before continuing, we will introduce some further notation regarding the marginal distribution of Q_0 with respect to the strategy s_i of Player $i \in \{1, 2, 3\}$. Namely, we will write Q_0^i to denote the marginal distribution of Q_0 with respect the strategy s_i. For example, one can verify that $Q_0^3(s_3^1) = \frac{3}{4}$ and $Q_0^3(s_3^1|s_2^1) = 1$. Namely, under the distribution Q_0, the ex-ante probability that Player 3 plays his first strategy s_3^1 (the left matrix) is $\frac{3}{4}$, and the probability that Player 3 plays s_3^1 conditional on Player 2 playing s_2^1 (the left most column) is 1. We will now proceed to introduce the last few definitions necessary to continue to Step 2.

Definition 1.24. A.2 (b-d):

b.) Player 2 *reports the deviation* D_1 *at time* t if she communicates according to the communication strategy $((\rho_2^{t'})_{t'<t}, (\rho_2^{r_1,t'})_{t'\geq T})$ defined in Assumption 1.

c.) The protocol $\mathcal{P}(\mathcal{N}_6)$ *adheres to the deviation* D_1 *reported by Player 2 with the strategy* $\tilde{Q}$ if whenever Player 2 reports the deviation D_1 at some time $t < T-1$, the protocol implements the distribution $\tilde{Q}$.

d.) The protocol *does not adhere to the deviations* D_1 *reported by Player 2* if $\hat{\sigma}_3 \sim Q_0^3$ whenever Player 2 reports the deviation D_1 at any time $t < T$.

These definitions capture the idea that the protocol may or may not allow reporting of deviations for Player 2, and if it does allow reporting the protocol must specify a continuation of play whenever such a report is made. Further, the protocol must specify whether Player 2, upon detecting a deviation, punishes the deviator independently (*i.e.,* if the protocol does not adhere to deviations reported by Player 2) or if there is a joint punishment of the deviator (*i.e.,* if the protocol adheres to the report with some strategy profile $\tilde{Q} \neq Q_0^{-2} \cdot \tilde{Q}^2$ for any $\tilde{Q}^2 \in \Delta(S_2)$). We will now proceed with the next three steps to show what conditions the protocol $\mathcal{P}(\mathcal{N}_6)$ must satisfy in order to deter the deviation D_1.

Step 2: $\mathcal{P}(\mathcal{N}_6)$ must adhere to the deviation D_1 reported by Player 2.

Suppose that $\mathcal{P}(\mathcal{N}_6)$ does not adhere to the deviation D_1 reported by Player 2. Then, the deviation D_1, when it is detected, must be deterred by an independent punishment strategy, which we will denote by $p_2(D_1) = (\sigma_1, \sigma_2, 1-\sigma_1-\sigma_2) \in \Delta(S_2)$, given that Player 3 is playing independently according to the marginal distribution Q_0^3 conditional on Player 2's suggestion. Then, we can see that the payoff of Player 1 from making the deviation D_1 in the communication phase and playing s_1^l, for some $l \in \{1,2\}$ in the play phase, denoted as $u_1^{D_1}(s_1^l)$ is

$$\begin{aligned} u_1^{D_1}(s_1^l) = \lambda_1[Q_0^2(s_2^1)u_1(s_1^l, s_2^1, Q_0^3(s_3|s_2^1)) + Q_0^2(s_2^2)u_1(s_1^l, s_2^2, Q_0^3(s_3|s_2^2))] \\ +\lambda_2[Q_0^2(s_2^1)u_1(s_1^l, s_2^2, Q_0^3(s_3|s_2^1)) + Q_0^2(s_2^2)u_1(s_1^l, s_2^1, Q_3(s_3|s_2^2))] \\ +(1-\lambda_1-\lambda_2)[Q_0^2(s_2^1)u_1(s_1^l, p_2(D_1), Q_0^3(s_3|s_2^1)) + Q_0^2(s_2^2)u_1(s_1^l, p_2(D_1), Q_0^3(s_3|s_2^2))] \end{aligned}$$

and when restricting attention to the game Γ and CE Q above, we obtain that whenever Player 1 makes the deviation D_1 and then plays s_1^1 their payoff is

$$u_1^{D_1}(s_1^1) = 60\lambda_1 + 64.75\lambda_2 - (1-\lambda_1-\lambda_2)\frac{1}{4}(1050 - 805\sigma_1 - 720\sigma_2) \geq 60 = u_1(Q_0)$$

Noting that $u_1(Q_0) = 60$, we can see that the only way that the protocol deters the deviation D_1 is if $\lambda_1 = 1$.[32] Namely, D_1 is not profitable if and only if Player 2 learns her correct suggested action with probability 1 whenever Player 1 makes the deviation D_1 and the protocol does not adhere to deviations reported by Player 2. The next two claims will prove that whenever the protocol $\mathcal{P}(\mathcal{N}_6)$ does not adhere to deviations reported by Player 2 and is such that $\lambda_1 = 1$ then either there exists a profitable deviation for Player 3 in the play phase or it is profitable for Player 1 to make the deviation D_1 in every realization of the protocol.

Claim 1: Whenever $\lambda_1 = 1$, then either Player 3 learns with probability 1 that Player 1 has made the deviation D_1, or with positive probability Player 3 receives a history that perfectly reveals $\tilde{s}_2$ when no player has deviated from $\mathcal{P}(\mathcal{N}_6)$.

Proof. By contraposition, suppose that $\mathcal{P}(\mathcal{N}_6)$ is such that Player 3 does not learn that D_1 has occurred with probability 1 whenever Player 1 makes the deviation D_1 and that the probability that Player 3 correctly learns $\tilde{s}_2$ when no player has deviated is zero. In what follows, we denote by $\pi^T_{j,k}(\cdot|D_1)$ the conditional distribution of history $h^T_{j,k}$ conditional on Player 1 making the deviation D_1.

Now, if $\lambda_1 = 1$, then it must be the case that $h^T_{2,3}$ carries all of the information regarding $\tilde{s}_2$ conditional on Player 1 making the deviation D_1. If we let $\hat{h}^T_{2\to1}$ denote the random variable distributed according to $\pi^T_{2\to1}(\cdot|D_1, h^T_{2,3})$, then one consequence of this is that for any realization $h^T_{2,3}$ of $\pi^T_{2,3}(\cdot|D_1)$, Player 3 knows that it must be the case that $\mathbb{P}(\hat{\sigma}_2((\hat{h}^T_{2\to1}, m^T_0), h^T_{2,3}) = \tilde{s}_2|D_1) = 1$. Therefore, if Player 3 knows that the deviation D_1 has occured, he can always learn $\tilde{s}_2$ by at time T drawing a *pseudo* history of Player 2, $h^T_{2\to1}$ from the distribution $\pi^T_{2\to1}(\cdot|D_1, h^T_{2,3})$, and computing $\hat{\sigma}_2((h^T_{2\to1}, m^T_0), h^T_{2,3})$. We call such a process by Player 3 a *decoding* and denote by $\zeta(h^T_{2,3}|D_1)$ the random variable $\hat{\sigma}_2((\hat{h}^T_{2\to1}, D_1), h^T_{2,3})$ where $\hat{h}^T_{2\to1} \sim \pi^T_{2\to1}(\cdot|D_1, h^T_{2,3})$.

Now, if Player 3 does not learn that the deviation D_1 has occurred with probability 1 whenever Player 1 makes the deviation D_1, then it must be the case that $\mathbb{P}_\rho(supp(\pi^T_3(\cdot|D_1))) > 0$; with positive probability Player 3 receives a history $h^T_{2,3}$ as if Player 1 made the deviation D_1 when all players obediently communicate according to ρ. Further, if Player 3 does not learn $\tilde{s}_2$ with positive probability when all players communicate according to ρ, then it must be the case that $\mathbb{P}_\rho(\zeta(h^T_{2,3}|D_1) \neq \tilde{s}_2|h^T_{2,3} \in supp(\pi^T_3(\cdot|D_1)) \cap supp(\pi^T_3)) > 0$; whenever Player 3 receives a history consistent with the deviation D_1 but no deviation has been made, then with positive probability he decodes the wrong action. But, this implies that if Player 1 makes the deviation D_1, then with positive probability Player 3 receives a history

[32] Note, that no punishment gives Player 1 a payoff less than 60 given that if $\sigma_1 = 1$ then the payoff of this punishment is $\frac{1050-805}{4} = 61.25$.

$h^T_{2,3} \in supp(\pi^T_3(\cdot|D_1)) \cap supp(\pi^T_3)$ and decodes the wrong suggested action even knowing that D_1 was made.[33] Finally, we note that if $\zeta(h^T_{2,3}|D_1) \neq \tilde{s}_2$, over any positive probability support of histories consistent with the deviation D_1, then by the construction of ζ, it must be the case that Player 2 receives the wrong suggested action with positive probability when Player 1 makes the deviation D_1, contradicting the fact that $\lambda_1 = 1$.

□

Claim 2: If Player 3 learns $\tilde{s}_2$ with positive probability whenever all players communicate according to ρ, then Player 3 has a profitable deviation from $\mathcal{P}(\mathcal{N}_6)$.

Proof. Suppose this is the case, and Player 3 learns $\tilde{s}_2 = s^2_2$ under some realization of the protocol and receives the suggested action $\tilde{s}_3 = s^1_3$. We need to show that under any consistent and sequentially rational belief system, Player 3 has a profitable deviation. Here the relevant off path beliefs are whether players 1 and 2 learn that Player 3 has learned the suggested action $\tilde{s}_2 = s^2_2$. We have thus the 4 following cases to consider:

Case 1: Neither Player 1 nor Player 2 learn that Player 3 has learned $\tilde{s}_2 = s^2_2$.

In this case any consistent belief system has players 1 and 2 play their correlated equilibrium strategies $\tilde{s}_2 = s^2_2$ and $\tilde{s}_1 = s^1_1$.[34] In this case, by playing s^2_3 Player 3 gets a payoff of 65 as opposed to 35 and therefore has a profitable deviation.

Case 2: Player 1 learns that Player 3 has learned $\tilde{s}_2 = s^2_2$ but Player 2 does not.

In this case, any sequentially rational belief of Player 1 must have him expect that, conditional on learning that Player 3 has learned $\tilde{s}_2$ and receiving the suggested action $\tilde{s}_1 = s^1_1$, Player 3 will play s^2_3 if Player 3 believes that Player 1 will play s^1_1. Now, conditional on Player 2 playing s^2_2, it is always strictly optimal for Player 1 to play s^1_1 no matter the strategy of Player 3; he gets a payoff of 80 v. 78 if Player 3 plays s^1_3 and a payoff of 90 v. 16 if Player 3 plays s^2_3. Therefore, any sequentially rational beliefs of Player 3 must ascribe probability 1 to Player 1 playing s^1_1 if he learns that Player 3 has learned $\tilde{s}_2$. Hence the deviation to play s^2_3 is still profitable for Player 3.

Case 3: Player 2 learns that Player 3 has learned $\tilde{s}_2 = s^2_2$ but Player 1 does not.

If Player 2 believes that Player 3 will play s^1_3 conditional on learning $\tilde{s}_2$, then given Q_0 is a correlated equilibrium distribution, the only sequentially rational beliefs of Player 3 are that Player 2 play s^2_2 in which case Player 2's beliefs are not sequentially rational (Player 3 strictly profits by playing s^2_3 in this case). Hence, any sequentially rational beliefs of Player

[33] If Player 3 receives $h^T_{2,3} \in supp(\pi^T_3(\cdot|D_1)) \cap supp(\pi^T_3)$ with positive probability only when Player 1 communicates obediently, then this implies that with probability 1 he receives a history $h^T_{2,3} \in supp(\pi^T_3(\cdot|D_1)) \backslash supp(\pi^T_3)$ whenever D_1 was made and therefore learns with probability 1 whenever Player 1 makes D_1, a contradiction.

[34] Under Q_0 Player 1 is always suggested to play s^1_1 whenever Player 2 is suggested to play s^2_2.

2 are that Player 3 play s_3^2 conditional on learning $\tilde{s}_2$. Further, in this case Player 3 must believe that Player 2 will play s_2^3. Finally, if Player 2 believes Player 3 will play s_3^2 and Player 3 believes that Player 2 will play s_2^3 then Player 3 still receives a payoff of 40 which is higher than his payoff of playing s_3^1 (which is 35). Thus, Player 3 still has a profitable deviation to play s_3^2 conditional on learning $\tilde{s}_2 = s_2^2$.

Case 4: Both players 1 and 2 learn that Player 3 has learned $\tilde{s}_2 = s_2^2$.

In this case the only way to deter the aforementioned deviation of Player 3 is for Player 3 to believe that players 1 and 2 will play something other than $(\tilde{s}_1, \tilde{s}_2) = (s_1^1, s_2^2)$. Given that Player 3 learns $\tilde{s}_2$ with positive probability when all players follow the protocol implies, in order for these beliefs to be consistent, that players 1 and 2 do not play their suggested actions $(\tilde{s}_1, \tilde{s}_2) = (s_1^1, s_2^2)$ with positive probability under the protocol $\mathcal{P}(\mathcal{N}_6)$. Therefore, $\mathcal{P}(\mathcal{N}_6)$ does not implement Q_0.

Finally, we should account for the case where Player 3 is uncertain as to whether he is in one of the above 4 cases. First, if Player 3 is uncertain whether he is in cases $(1) - (3)$, then playing s_3^2 is still profitable. This simply comes from the fact that Player 3's optimal continuation in cases $(1) - (3)$ is to play s_3^2. Finally, if Player 3 is unsure whether he is in case (4), we simply note that if he obediently plays s_3^1, then he must have the belief that players 1 and 2 will not play (s_1^1, s_2^2). Further, given that this is only relevant if case (4) happens with positive probability (whether or not Player 3 knows of it), then consistency implies that on the equilibrium path, with positive probability players 1 and 2 are suggested to play (s_1^1, s_2^2) but do not. Ruling out this case where $\mathcal{P}(\mathcal{N}_6)$ does not implement Q_0, we see that playing s_3^2 is always an optimal continuation for Player 3 whenever he learns $(\tilde{s}_2, \tilde{s}_3) = (s_2^2, s_3^1)$.

Claim 3: If Player 3 learns that some D_1 has occurred with probability 1, then D_1 is a profitable deviation for Player 1.

First, we note that if Player 3 learns that D_1 has occurred with probability 1, then whenever Player 1 makes the deviation D_1, as shown above, Player 3 learns $\tilde{s}_2$. In what follows we assume that we are in the case where Player 1 makes the deviation D_1, Player 3 learns $\tilde{s}_2 = s_2^2$, and is suggested the action $\tilde{s}_3 = s_3^1$. Given that both Player 1 and 3 are aware whenever D_1 occurs, then we only need to consider the following two cases.

Case 1: Player 3 learns Player 1 has made the deviation D_1 but Player 2 is unaware of this.

Just as in Case 2 of Claim 2 above, any sequentially rational belief system has Player 3 play s_3^2 in this case. Given that such a deviation by Player 3 gives Player 1 a higher payoff than following his suggested action implies that D_1 is profitable.

Case 2: It is common knowledge that Player 1 has made the deviation D_1.

Here we note that if it is common knowledge that D_1 has been made, then it is common knowledge that Player 3 learns $\tilde{s}_2$. Therefore, if D_1 is not profitable, then Player 1 must

believe that players 2 and 3 will play some (sequentially rational) strategy p_{-1} that gives Player 1 a payoff lower than 80 no matter whether he plays s_1^1 or s_1^2. What we will now prove is that no such punishment strategy exists if after making the deviation D_1, Player 1 plays s_1^1.

First note that if Player 1 plays s_1^1, then the only sequentially rational pure strategy profile p_{-1} is such that players 2 and 3 play (s_2^3, s_3^2) which gives Player 1 a payoff of $810 > 80$. Hence, the punishment must be mixed. Let us denote $p_{-1} = ((\sigma_1, \sigma_2, 1-\sigma_1-\sigma_2), (\alpha, 1-\alpha))$ the punishment strategies for Player 2 and Player 3. First, we can see that if Player 1 plays s_1^1, then for Player 2, s_2^2 is strictly dominated by s_2^1. Therefore, if p_{-1} is sequentially rational it must be the case that $\sigma_2 = 0$. Further, sequential rationality requires that

$$u_2(s_1^1, s_2^1, (\alpha, 1-\alpha)) := 80\alpha + 40(1-\alpha) = 20\alpha + 140(1-\alpha) =: u_2(s_1^1, s_2^3, (\alpha, 1-\alpha))$$

or $\alpha = \frac{5}{8}$. Similarly, sequential rationality for Player 3 requires that

$$u_3(s_1^1, (\sigma_1, 0, 1-\sigma_1), s_3^1) := 70\sigma_1 = 65\sigma_1 + 40(1-\sigma_1) =: u_3(s_1^1, (\sigma_1, 0, 1-\sigma_1), s_3^2)$$

or $\sigma_1 = \frac{8}{9}$. Finally, we note that

$$u_1(s_1^1, (\frac{8}{9}, 0, \frac{1}{9}), (\frac{5}{8}, \frac{3}{8})) = \frac{5}{8}[35 * \frac{8}{9} + 80 * \frac{1}{9}] + \frac{3}{8}[64 * \frac{8}{9} + 810 * \frac{1}{9}] = \frac{961}{12} > 80$$

and therefore there does not exist a sequentially rational punishment for Player 1 and D_1 is profitable in this case. □

Step 3: There exists no distribution $\tilde{Q}$ that enforces any strategy $(\rho, \hat{\sigma})$ that induces the same distributio as Q_0 as a perfect Bayesian equilibrium of $(\Gamma, \mathcal{N}_6)$ when the protocol $\mathcal{P}(\mathcal{N}_6)$ adheres to the deviation D_1 reported by Player 2.

We now know from Step 2 that if $\mathcal{P}(\mathcal{N}_6)$ implements Q_0 on $\mathcal{N}_6$ then $\lambda_1 < 1$ and $\mathcal{P}(\mathcal{N}_6)$ adheres to the deviation D_1 reported by Player 2 with some strategy $\tilde{Q} \neq Q_0$. Namely, if $\mathcal{P}(\mathcal{N}_6)$ implements Q_0 on $\mathcal{N}_6$, then whenever Player 2 detects the deviation D_1 made by Player 1 (which happens with positive probability[35]), she reports this deviation by using a different communication strategy (assumed to exist in Assumption 1) and then the protocol implements some distribution $\tilde{Q}$ that punishes Player 1. It is worth noting that it may be the case that Player 2 detects a deviation but is not necessarily aware that it was the deviation D_1. Either way, in this case the protocol must adhere to the deviation reported by Player 2 with $\tilde{Q}$ that punishes D_1 and potentially other deviations. In what follows we will see that

[35] Player 2 detects D_1 with positive probability due to the fact that $\lambda_1 < 1$ and $\lambda_2 = 0$.

even if Player 2 can detect D_1, there is still a profitable deviation from $\mathcal{P}(\mathcal{N}_6)$. Therefore in what follows we assume, without loss, that when Player 2 detects a deviation she knows it was D_1.

Now, in order for the strategy $\tilde{Q}$ to prevent the deviation D_1, it must be credible in the sense that $\tilde{Q}$ is incentive compatible for Player 2. Second, if $\mathcal{P}(\mathcal{N}_6)$ implements Q_0 on $\mathcal{N}_6$ it must be the case that $u_2(\tilde{Q}) \leq u_2(Q_0)$ as otherwise Player 2 can *fake* the deviation D_1 (see the next paragraph) and then report that D_1 has occurred and this would be a profitable deviation. Now, based on the construction of Γ_0, it can be seen by running a simple linear program that the unique $\tilde{Q}$ satisfying $u_1(\tilde{Q}) \leq u_1(Q_0)$, $u_2(\tilde{Q}) \leq u_2(Q_0)$, and $\tilde{Q}$ is incentive compatible for Player 2, is the CE distribution Q_0 itself. Therefore, given that we are considering only finite protocols implies that either the condition $u_2(\tilde{Q}) \leq u_2(Q_0)$ is relaxed, or D_1 is profitable for Player 1.

Finally, we argue that whenever the condition $u_2(\tilde{Q}) \leq u_2(Q_0)$ is relaxed, then Player 2 has a profitable deviation. Namely, suppose that for all $t = 0, 1, ..., T$ Player 2 draws $m^t_{2\to1} \sim \rho^t_{2\to1}[h^t_{2,3}, h^t_{1\to2} = m^t_0]$ and denote the history $\hat{h}^t_{2\to1}$ the result of such draws. If for all t Player 2 sends $m^t_{1\to2} = m_0$ and $m^t_{2\to3} \sim \rho^t_{2\to3}[h^t_{2,3}, \hat{h}^t_{2\to1}, h^t_{1\to2} = m^t_0]$ then whenever Player 1 follows the protocol, the mediator cannot detect whether Player 1 is making the deviation D_1 or Player 2 is making the aforementioned deviation. Hence, if the protocol adheres to deviations D_1 reported by Player 2, then whenever she reports D_1 after communicating in this fashion we have just shown that either D_1 is profitable, or $\mathcal{P}(\mathcal{N}_6)$ must adhere to that report with some strategy $\tilde{Q}$ such that $u_2(\tilde{Q}) > u_2(Q_0)$. Thus, we arrive at our final contradiction that for all protocols $\mathcal{P}(\mathcal{N}_6)$, either Player 1 has an optimal deviation D_1 or Player 2 has an optimal deviation to falsely report that D_1 has occurred. This takes us to our last step of the proof of Proposition 1.

Step 4: Assumption 1 can be made without loss of generality.

Assumption 1 states that there exists a reporting strategy of Player 2 that allows her to report that D_1 has occurred *and* allows the mediator to receive this report with probability 1. First, note that this assumption is only relevant if the protocol adheres to deviations reported by Player 2. Then, we have just shown that even if Player 2 can detect and report the deviation D_1, there is still a profitable deviation from the protocol. Furthermore, if the mediator does not receive the report by Player 2 with probability 1, then with positive probability Player 2 receives no suggestion at the end of the protocol and must therefore play independently. In this case Player 1 receives a payoff identical to the case where $\mathcal{P}(\mathcal{N}_6)$ does not adhere to deviations reported by Player 2 and with $\lambda_1 < 1$ and $1 - \lambda_1 - \lambda_2 > 0$ in which case the deviation D_1 is still profitable. □

1.8.4 Proof of Corollary 1

Proof. Let Γ_0 and Q_0 be the game and CE from the proof of Proposition 1 (see figures 14 and 15) with $n-3$ extra players who each have a single trivial strategy and trivial payoffs independent of the strategy profile played by players 1, 2, and 3. Further, let $\mathcal{N}$ be a network such that $\mathcal{N}$ is only strongly 1-connected from M to 2 and strongly k-connected from 2 to M for some $k \leq n$. Finally, let Player 1 be on every path connecting M to 2.

Now, note that if Player 1 makes the deviation D_1 from the proof of Proposition 1, then given the strong 1-connectedness of $\mathcal{N}$, Player 2 cannot ever learn her strategy with probability 1. In the context of the proof of Theorem 1, whenever Player 1 makes the deviation D_1 it must be the case that $\lambda_1 < 1$. The only other way that Player 2 would receive her suggested action is if she chose it herself. But then Player 2 must be indifferent between choosing one suggested action over another so that her ex interim payoff when receiving the strategy suggestion $\tilde{s}_2 = s_2^1$ is the same as when she receives the strategy suggestion $\tilde{s}_2 = s_2^2$. Thus, it must be the case that $u_2(s_2^1, Q_0^{-2}(s_{-2}|s_2^1)) = 0 = 80 = u_2(s_2^2, Q_0^{-2}(s_{-2}|s_2^2))$, a contradiction.

Finally, if the protocol adheres to deviations reported by Player 2, we note that just as in the proof of Theorem 1, the unique $\tilde{Q}$ such that both $u_1(\tilde{Q}) \leq u_1(Q_0)$ and $u_2(\tilde{Q}) \leq u_2(Q_0)$ is Q_0 and therefore, either D_1 is profitable or Player 2 has an optimal strategy to falsely report D_1 in every realization of the protocol. □

1.8.5 Proof of Corollary 2

Proof. We will now construct a network $\mathcal{N}$ that is $n-1$ weakly connected from M to some Player $i \notin Y(M)$ such that there exists a CE of the n-player game that is not implementable on $\mathcal{N}$. Again we look at the 3-player game in the proof of Theorem 1 and add $n-3$ players, each with a trivial strategy so that the game played between players 1,2, and 3 remains unchanged. Then, we add each of these $n-3$ players to the network $\mathcal{N}_0$ in the following way. For each Player $j \in I \backslash \{1,2,3\}$ add the arcs Pj and $2j$ to the network $\mathcal{N}_0$.

The resulting network, with $I \backslash Y(M) = \{2\}$, is then weakly $(n-1)-$connected between M and 2 via the undirected paths $(Pj, 2j)$ for all $j \in I \backslash \{2\}$. Now, given the structure of this network, we can see that players $4, 5, ..., n$ cannot communicate (*i.e.,* they are *sinks* in the network $\mathcal{N}_0 \cup \{\{Pj, 2j\} : j \in I \backslash \{1,2,3\}\}$), thus if the game and CE from the proof of Proposition 1 cannot be implemented on $\mathcal{N}_0$, then it also cannot be implemented on this resulting network. □

1.8.6 Proof of Theorem 2

Proof. ($\Leftarrow$) First, we will illustrate the necessity of the conditions of Theorem 1. To do this, note that the network $\mathcal{N}_0$ satisfies 1-connectivity from 2 to M but does not satisfy the conditions of Theorem 1. Then, consider the trivial Bayesian game G_0 such that $\Theta = \{\theta\}$ and payoffs conditional on the single type θ are equal to those of the game Γ_0. Then the COE q_0 of G_0 with $q(\cdot|\theta) = Q_0(\cdot)$ is a communication equilibrium that cannot be implemented on $\mathcal{N}_0$ as was proven in the proof of Proposition 1. In order to prove necessity of 2-connectivity from $i \in I\backslash X(M)$ to M and the conditions of Theorem 1, consider the game G_0 and the network $\mathcal{N}_7$ of Figure 18.

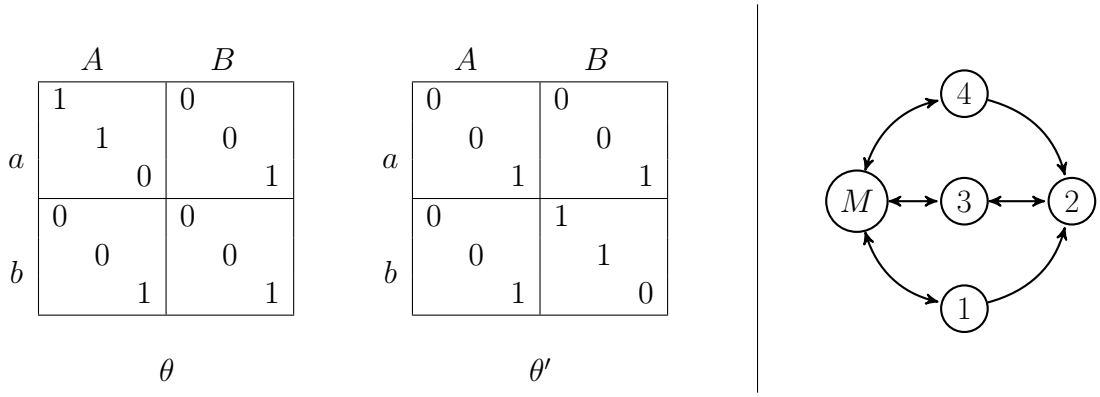


Figure 18: The game G_0 and the network $\mathcal{N}_7$.

The game G_0 is such that players 3 and 4 have no actions nor information and only take part in the communication phase. Meanwhile, Player 2 learns whether the state is $\theta_2 \in \{\theta, \theta'\}$ and has actions $A_2 = \{A, B\}$ and Player 1 has no private information and actions $A_1 = \{a, b\}$. We assume Player 4 has a constant payoff and therefore we can ignore any strategic concerns regarding his strategy. Thus, the matrices represent the payoffs to players 1, 2, and 3 in each state θ and θ'. Finally, we assume that the common prior p puts probability $\alpha \in (0, 1)$ on state θ and probability $(1 - \alpha)$ on state θ'. This game has a simple communication equilibrium q_0 whereby players 1 and 2 play (a, A) in state θ and (b, B) in state θ'.

Now, note that $\mathcal{N}_7$ satisfies both conditions of Theorem 1. What we will show is that the communication equilibrium q_0 cannot be implemented on the network $\mathcal{N}_7$ which is only 1-connected from Player $2 \in I\backslash X(M)$ to M. To see this, we simply note that if Player 3 randomly sends a message $\tilde{m} \sim U[0, 1)$, whenever he is required to forward any message in the first communication phase where Player 2 must report her type to M, then M can never learn θ_2 with probability 1. This simply comes from the fact that the only edge from 2 to M is $23P$, thus any information about θ_2 is contained in $h^T_{2,3}$. Finally, given that q_0 gives Player

3 a strictly worse payoff over any other communication equilibrium of this game, means that the aforementioned deviation is always profitable for Player 3; upon making this deviation there is positive probability players 1 and 2 do not perfectly coordinate on (a, A) in state θ or (b, B) in state θ'.

Now, similar to Corollary 2, we could always augment the game G_0 and network $\mathcal{N}_7$ such that q_0 cannot be implemented on a network $\mathcal{N}_7'$ that satisfies both of the conditions of Theorem 1 and is weakly k-connected between M and Player 2. Further, in the same vein we can augment $\mathcal{N}_7$ so that q_0 cannot be implemented on a network $\mathcal{N}_7''$ that satisfies both of the conditions of Theorem 1 and is strongly k-connected from M to 2. Therefore, it must be the case that there is an additional disjoint directed path from 2 to M. □

Proof. ($\Rightarrow$) To prove sufficiency we first note that given the network satisfies the conditions of Theorem 1, implies that M can send the suggested strategy to each player in a perfectly secure fashion. But, this also implies that M can use the protocols constructed in the proof of sufficiency of Theorem 1 to send any message to any Player $i \in I \backslash X(M)$ in a perfectly secure fashion. In what follows we assume (without loss) that Player $i \in I \backslash X(M)$ has a type space $\Theta_i = \{\theta_i^{(1)}, ..., \theta_i^{(l)}\}$. We will now construct a protocol that has Player i send his type $\theta_i^{(k)}$ to M in a perfectly secure fashion.

Communication Strategy α: **Step 1:** M draws a vector $\mathbf{a} = (a_1, ..., a_l) \sim U[0,1)^l$ and a key $x_i \sim U[0,1)$ and sends $(\mathbf{a}, x_i)$ to Player i using the appropriate secure protocol constructed in Theorem 1. **Step 2:** Player i of type $\theta_i^{(k)}$ sends $x \oplus a_k$ on both disjoint paths $p_1(i, M)$ and $p_2(i, M)$ guaranteed to exist by 2-connectivity.

Type Decoder θ^M: **Step 1:** Conditional on receiving message m_1 and m_2 from the paths $p_1(i, M)$ and $p_2(i, M)$ respectively at the end of the first communication phase, M decodes $m_1 \ominus x$ and believes that Player i is of type $\theta_i^{(z)}$ for $z \in \{1, ..., l\}$ if $m_1 \ominus x = a_z$. **Step 2:** Otherwise, M decodes $m_2 \ominus x$ and believes that Player i is of type $\theta_i^{(z)}$ for $z \in \{1, ..., l\}$ if $m_2 \ominus x = a_z$. Finally, if $m_1 \ominus x \neq a_z$ and $m_2 \ominus x \neq a_z$ for all $z \in \{1, ..., l\}$ then M believes that Player i is of type $\theta_i^{(1)}$.

Finally, when suggesting strategy a_i to each $i \in I$ we use the grand protocol constructed in the proof of sufficiency of Theorem 1. Now, in order to show that this protocol constitutes a PBE of the game $(G_0, \mathcal{N})$ we only need to prove that types are sent in a perfectly secure fashion to M under the communication strategy (α, θ^M). To see that (α, θ^M) guarantees secrecy, we simply note that given no player has any information regarding $(\mathbf{a}, x)$ implies that $a_k \oplus x \sim U[0,1)$ with respect to any players information besides Player i. To see that (α, θ^M) guarantees resiliency, we simply note that if some player deviates on the path $p_1(i, M)$, then $\tilde{m}_1 \neq m_{2\to 1}$ and $\mathbb{P}(\tilde{m}_1 \ominus x \in \{a_1, ..., a_l\}) = 0$. Therefore, M decodes correctly $m_2 \ominus x = a_k$ in Step 2. If some player deviates on the path $p_2(i, M)$ then M decodes correctly

$m_1 \ominus x = a_k$ in Step 1. Therefore, (α, θ^M) guarantees perfectly secure communication from Player i to M. Finally, we know that the grand protocol of Theorem 1 guarantees perfectly secure communication from M to i and therefore we have constructed a protocol that allows for perfectly secure communication between each Player $i \in I$ and M. □

1.8.7 Proof of Corollary 3

Proof. ($\Leftarrow$) We first note that as in the proof of necessity of Theorem 2, the conditions of Theorem 1 are also necessary when considering Bayesian games. Finally, to illustrate the necessity of strong 1-connectivity from each player to M, we simply note that no COE q such that $q(\cdot|\theta_i, \theta_{-i}) \neq q(\cdot|\theta_i', \theta_{-i})$ for some $\theta_i, \theta_i' \in \Theta_i$ and $\theta_{-i} \in \Theta_{-i}$ can be implemented on $\mathcal{N}$ unless $\mathcal{N}$ is strongly 1-connected from i to M. □

Proof. ($\Rightarrow$) Here we will sketch the protocol used in Renou and Tomala (2012); an incentive compatible mechanism for sending θ_i to the mediator guaranteeing secrecy of the information regarding types sent to M. Suppose without loss that $\theta_i \in \Theta_i := \{1, 2, ..., t_i, ..., T_i\}$ for all $i \in I$ and denote by p^i the marginal distribution of the belief p on Θ_i, and by $\bar{p}^i(t_i) = \sum_{\theta_i \leq t} p^i(\theta_i)$ the cumulative distribution function of p^i. Then, define a partition $\Pi_i = \{\Pi_i(1), ..., \Pi_i(T_i)\}$ of the message space $\mathcal{M} = [0, 1)$ into T_i subsets such that $\Pi_i(t_i) = [\bar{p}^i(t_i - 1), \bar{p}^i(t_i))$, defining $\bar{p}^i(0) = 0$.

Renou and Tomala (2012) then develop a protocol that allows players to secretly send their type to the mediator by drawing a message m_i uniformly on the subset $\Pi_i(t_i)$ whenever they are of type t_i. Then if all players follow the protocol, the mediator receives m_i with probability 1 and deduces that Player i is of type t_i whenever $m_i \in \Pi(t_i)$. Then, they use the fact that the network is weakly 2-connected to have the remaining players generate keys and send them to their neighbors in such a way that Player i can encrypt her message m_i with one of the keys and send it along the path to the mediator (1) without any player on that path learning the key (and therefore the message), and (2) such that the mediator learns the key at the end of the protocol (and therefore the message). The main feature that prevents deviations is that their protocol is constructed such that no player has an incentive to deviate during the communication phase due to a specific property of such a deviation: whenever Player $j \in I$ deviates in the phase of the protocol where Player $i \in I$ is forwarding m_i to the mediator, this results in the mediator receiving some message $\tilde{m}_i$ such that the probability that $\tilde{m}_i \in \Pi_i(t_i)$ is equal to $\Pi_i(t_i)$. Hence, the distribution of θ_{-j}^M, given that all players $i \in I \backslash \{j\}$ follow the protocol and Player j makes any deviation during the communication phase, is the same as if Player j had followed the protocol. The protocol we will use to prove sufficiency will simply be the two stage protocol that uses the aforementioned protocol

of Renou and Tomala (2012) to have each player report their types to M, combined with the grand protocol constructed in the proof of Theorem 1 to have M send suggested actions distributed according to $q(\cdot|\theta)$ to each player. Now, based on the respective constructions we know that if all players follow the protocol $((\alpha, \theta^M), (\rho, \hat{\sigma}))$ then $\mathbb{P}_\alpha(\theta^M = \theta|\theta) = 1$ and $\mathbb{P}_\rho(\hat{\sigma} = s|\theta) = q(s|\theta)$. What is left to prove is that $((\alpha, \theta^M), (\rho, \hat{\sigma}))$ is a PBE of the game $(G, \mathcal{N})$. Namely, we will now prove that there are no compound deviations where some Player $i \in I$ deviates in both phases of the protocol and is made better off by doing so.

Let us introduce the following notation. For a given *first phase* communication strategy α' and any $i \in I$, denote by $p_{\alpha'}(\theta_{-i}) := \mathbb{P}_{\alpha'}(\theta^M_{-i} = \theta_{-i}|\theta_{-i})$ the probability that the mediator learns the true type θ_{-i} under the communication strategy α'. Similarly, for a given *second phase* communication strategy ρ' denote by $q_{\rho'}(s|\theta^M) = \mathbb{P}_{\rho'}(\hat{\sigma} = s|\theta^M)$ the probability that $\hat{\sigma} = s$ given some fixed $\theta^M \in \Theta$ and the communication strategy ρ'. Now, note that the protocol (α, θ^M) of Renou and Tomala (2012) satisfies the property that $p_{\alpha_{-i}, \alpha'_i}(\theta_{-i}) = p_\alpha(\theta_{-i}) = p(\theta_{-i})$ for any $i \in I$, $\theta_{-i} \in \Theta_{-i}$ and any communication strategy α'_i due to the aforementioned property that prevents players from deviating in their protocol. Now, note that resiliency of the protocol $(\rho, \hat{\sigma})$ constructed in the proof of Theorem 1 implies that $q_{\rho_{-i}, \rho'_i}(s|\theta) = q_\rho(s|\theta) = q(s|\theta)$ for all $\theta \in \Theta$ and communication strategies ρ'_i. What we will now show is that these two properties guarantee that no player has a profitable deviation from the protocol $((\alpha, \theta^M), (\rho, \hat{\sigma}))$ whenever q is a COE.

Suppose there exists a profitable deviation by Player i, and denote this by (α'_i, ρ'_i), and $(\theta'_i, \delta(\hat{\sigma}_i))$ where α'_i and ρ'_i are the communication strategies that Player i uses in the first and second communication phases respectively, θ'_i is the type that Player i reports, and $\delta : S_i \to S_i$ maps the action suggested to Player i after the deviation $(\theta'_i, \alpha'_i, \rho'_i)$ to some other action of Player i. Now, suppose that this is a profitable deviation. Then it must be the case that

$$\sum_{\theta_{-i} \in \Theta_{-i}} \sum_{s \in S} p(\theta_{-i}) q(s|\theta_i, \theta_{-i}) u_i(s|\theta_i) < \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{s \in S} p_{\alpha_{-i}, \alpha'_i}(\theta_{-i}) q_{\rho_{-i}, \rho'_i}(s|\theta'_i, \theta_{-i}) u_i(\delta_i(s_i), s_{-i}|\theta_i)$$

but, given the aforementioned properties of the two protocols with respect to the distributions $p_\alpha(\cdot)$ and $q_\rho(\cdot)$, this implies that

$$\begin{aligned}\sum_{\theta_{-i} \in \Theta_{-i}} \sum_{s \in S} p(\theta_{-i}) q(s|\theta_i, \theta_{-i}) u_i(s|\theta_i) &< \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{s \in S} p_\alpha(\theta_{-i}) q_\rho(s|\theta'_i, \theta_{-i}) u_i(\delta_i(s_i), s_{-i}|\theta_i) \\ &= \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{s \in S} p(\theta_{-i}) q(s|\theta'_i, \theta_{-i}) u_i(\delta_i(s_i), s_{-i}|\theta_i)\end{aligned}$$

which contradicts the fact that q is a COE. □

1.8.8 Proof of Corollary 4

Proof. ($\Rightarrow$) Given that $\mathcal{N}$ is such that M and $i \in I \backslash Y(M)$ are 3-connected implies that $\mathcal{N}$ is strongly 3-connected from M to $i \in I \backslash Y(M)$ and strongly 3-connected from $i \in I \backslash Y(M)$ to M. Therefore, we can use the protocol from the proof of Theorem 2. □

Proof. ($\Leftarrow$) The necessary conditions of this Theorem guarantee that M and i are 3-connected whenever we are interested in undirected (*i.e.,* 2-way communication) networks. Further, if we try to relax the assumptions of this theorem, we would end up with a network such that M and some Player $i \in I \backslash Y(M)$ are at most 2-connected. But, we have already constructed in the proof of Proposition 1 a game and CE that cannot be implemented on the network $\mathcal{N}_6$ where M and each $i \in I \backslash Y(M)$ are 2-connected[36]. Namely, the network $\mathcal{N}_6$ is such that there are two disjoint directed paths from M to $i \in I \backslash Y(M)$ and two disjoint directed paths from i to M. Therefore, as far as this problem is concerned, the network $\mathcal{N}_6$ and its underlying undirected network are equivalent. Thus, we cannot implement on the undirected network analogous to $\mathcal{N}_6$ and hence must add an additional undirected edge between M and i which implies that the network be 3-connected between M and i. □

1.8.9 Proof of Lemma 1

Proof. Protocol for when $\mathcal{N}$ satisfies Condition (1) of Theorem 1: Assume that the network satisfies Condition (1) of Theorem 1, and has an additional disjoint path from M to each $i \in I \backslash Y(M)$. Then the network is strongly 4-connected from M to i and the mediator can send any message $m \in \mathcal{M}$ to Player i in a perfectly secure fashion by utilizing the following protocol.

- Communication strategy ρ: The mediator draws 4 keys x_1, x_2, x_3, x_4 uniformly from $\mathcal{M}$ and, as illustrated in Figure 19, sends $(m \oplus_K x_1, m \oplus_K x_2, m \oplus_K x_3, x_4)$ on Path 1, $(m \oplus_K x_2, m \oplus_K x_3, m \oplus_K x_4, x_1)$ on Path 2, $(m \oplus_K x_1, m \oplus_K x_3, m \oplus_K x_4, x_2)$ on Path 3, and $(m \oplus_K x_1, m \oplus_K x_2, m \oplus_K x_4, x_3)$ on Path 4, where $\oplus_K$ and $\ominus_K$ are the finite version of $\oplus$ and $\ominus$ with a message space of size K given by

$$x \oplus_K y := \begin{cases} x+y & \text{if } x+y < K \\ x+y-K & \text{if } x+y \geq K \end{cases} \qquad x \ominus_K y := \begin{cases} x-y & \text{if } x-y \geq 0 \\ x-y+K & \text{if } x-y < 0 \end{cases}$$

All players on forward any message sent to them from their predecessors to their successors.

- Decoding Step: Once Player i receives all four vectors, she decodes the first three elements of each vector using the appropriate key (*i.e.,* to decode $(m \oplus_K x_1, m \oplus_K x_2, m \oplus_K x_3)$

[36] Note here that any undirected network $\mathcal{N}$ can be written as a directed network $\vec{\mathcal{N}}$ such that for all $ij \in A(\mathcal{N})$ we have $\{ij, ji\} \in A(\vec{\mathcal{N}})$.

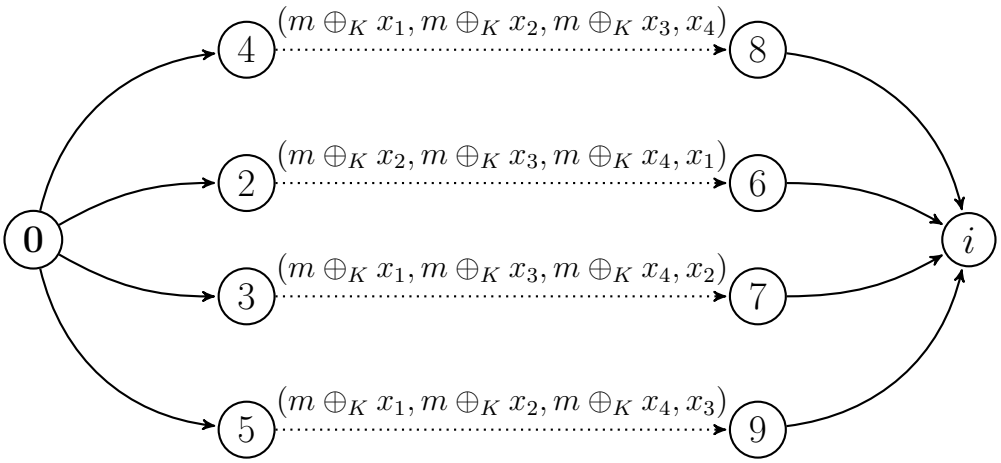


Figure 19: Protocol for implementation with finite message space.

sent on Path 1 (by subtracting using $\ominus_K$), she uses the last element of the vector sent on Path 2 to decode the first element $m \oplus_K x_1$, the last element of the vector sent on Path 3 to decode the second element $m \oplus_K x_2$, etc.). Finally, after decoding all four sets of triples, she interprets the message to be m' only if there are three triples where m' shows up *at least* twice.

To show that this protocol sends message m in a perfectly secure fashion, first note that secrecy is satisfied as all messages are uniformly distributed and therefore no information regarding m is revealed (this holds via the finite extension of Lemma 0).[37] To see that it satisfies resiliency, suppose that a player on Path 1 deviates and sends $(\tilde{m}_1, \tilde{m}_2, \tilde{m}_3, \tilde{x}_4)$ instead of the intended vector. The result of the decoding phase is given in the following table.

Decoding:	First Element	Second Element	Third Element
Vector From Path 1	$\tilde{m}_1 \ominus_K x_1$	$\tilde{m}_2 \ominus_K x_2$	$\tilde{m}_3 \ominus_K x_3$
Vector From Path 2	m	m	$m \oplus_K x_4 \ominus \tilde{x}_4$
Vector From Path 3	m	m	$m \oplus_K x_4 \ominus \tilde{x}_4$
Vector From Path 4	m	m	$m \oplus_K x_4 \ominus \tilde{x}_4$

Namely, whenever a player on Path 1 deviates and sends any vector $(\tilde{m}_1, \tilde{m}_2, \tilde{m}_3, \tilde{x}_4)$, there is no way for them to produce a message $m' \neq m$ twice in 3 or more of the decoded vectors (*i.e.,* the player can only produce $m' \neq m$ more than twice in the decoded vector from his own path, and otherwise can only produce a message $m' \neq m$ once in the remaining 3 vectors.). Hence, Player i will always interpret the result as the intended message m under any deviation from Path 1. Thus, by symmetry we can see that the protocol satisfies

[37] Further, no deviation is detectable.

resiliency for deviations on any of the remaining paths.

Protocol for when $\mathcal{N}$ satisfies Condition (2) of Theorem 1: Assume that the network satisfies Condition (2) of Theorem 1 for Player $i \in I \backslash Y(M)$. Then there are three paths from M to i (paths 1, 2, and 3) and one path from i to M (Path 4) such that all four paths are disjoint. We will construct a protocol with a finite message space that allows for the mediator to send any message m in a perfectly secure fashion to Player $i \in I \backslash Y(M)$ whenever it satisfies these conditions.

• Communication strategy ρ: **Step 1:** Player i draws a key x_0 uniformly from $\mathcal{M}$ and sends it to the mediator on Path 4. **Step 2:** The mediator draws a vector of keys x_1, x_2, x_3 uniformly from $\mathcal{M}$ and sends $m_1 = (x_1 \oplus m, x_2, x_0 \oplus m)$ on Path 1, $m_2 = (x_2 \oplus m, x_3, x_0 \oplus m)$ on Path 2, and $m_3 = (x_3 \oplus m, x_1, x_0 \oplus m)$ on Path 3. All players forward any messages received by their predecessors to their successors.

• Decoding Strategy: **Step 1:** Upon receiving the three vectors m_1, m_2, and m_3, Player i decodes the first element of each vector (*i.e.,* subtracts the first element of m_1 by the second element of m_3, subtracts the first element of m_2 by the second element of m_1, and subtracts the first element of m_3 by the second element of m_2) and interprets message m' if all three of the decoded first elements are equal to m'. If any of the decoded first elements differ she proceeds to Step 2. **Step 2:** She decodes the third element of each vector m_1, m_2, and m_3 by subtracting it with the key she sent to the mediator x_0 and interprets the message to be m' if at least two of the decoded messages are equal to m'.

To note that this protocol satisfies secrecy, we simply note that all messages sent are uniformly distributed over $\mathcal{M}$ and therefore reveal no information regarding m. To note that it satisfies resiliency, suppose that some Player j on Path 1 deviates and sends a message $\tilde{m}_1$ instead of the true message. In Step 1 of the decoding strategy, the only way Player 2 will receive the same message m' in all three of the decoded first elements is if $m' = m$. Therefore, unless $\tilde{m}_1$ is a trivial deviation she will proceed to Step 2. Then, given that a player on Path 1 has deviated, this implies that all messages on paths 2, 3, and 4 were sent truthfully, and therefore the decoded last element of m_2 is the same as the decoded last element of m_3 and equal to m, in which case Player i receives m. This shows that the protocol is resilient to deviations on paths 1, 2, and 3. To show that it is resilient to deviations on Path 4, we simply note that if a player on Path 4 deviates and sends $\tilde{x}_0$ instead of x_0, then all messages have been sent truthfully on paths 1, 2, and 3 and therefore Player i decodes m in Step 1 of the decoding strategy.

Now, we have shown that the mediator can send any message m in a perfectly secure fashion with a finite message space to any $i \in I$ such that the network satisfies the conditions of Theorem 1 and has an additional disjoint path from M to every $i \in I \backslash Y(M)$.

To extend this result to implement any correlated equilibria, the mediator can first send a vector of messages (using the above protocols), where each message in the vector represents a strategy of Player i. Then after sending drawing the strategy profile $\tilde{s}$ from the correlated equilibrium distribution the mediator simply sends the message corresponding to strategy $\tilde{s}_i$ using the appropriate protocol above. Finally, given that the resulting protocol satisfies security and resiliency (all messages are being send in a perfectly secure fashion using the above protocols), when we couple it with the beliefs from the proof of sufficiency of Theorem 1 (that upon detecting any deviation the resulting distribution of actions will still be the correlated equilibrium distribution), it is easy to see that this pair constitutes a perfect Bayesian Equilibrium of the extended game. □

References

[1] Alonso, R., Dessien, W., and Matouschek, N. (2008): "When Does Coordination Require Centralization?" *American Economic Review*, 98 (1): 145-179.

[2] Aumann, R. J. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67-96.

[3] Bang-Jensen, J. and Gutin, G., *Digraphs: Theory, Algorithms, and Applications*, 2002, Springer-Verlag.

[4] Barany, I. (1992): "Fair Distribution Protocols or How the Players Replace Fortune," *Journal of Mathematical Economics*, 17, 327-340.

[5] Ben-Or, M., Goldwasser, S., and Wigderson, A. (1988): "Completeness Theorems for Non-Cryptographic Fault-Tolerant Distributed Computation." *The Proceedings of the 20th Annual ACM Symposium on Theory of Computing.*

[6] Ben-Porath, E. (2003): "Cheap Talk in Games with Incomplete Information," *Journal of Economic Theory*, 108, 45-71.

[7] Chaum, D., Crépeau, C., and Damgård, I. (1988): "Multiparty Unconditionally Secure Protocols." *Proceedings of the twentieth annual ACM symposium on Theory of computing.* ACM.

[8] Crawford, V. M. and Sobel, J. (1982): "Strategic Information Transmission," *Econometrica*, 50 (6), 1431-1451.

[9] Dessein, W. (2002): "Authority and Communication in Organizations," *Review of Economic Studies*, 69, 1431-1452.

[10] Dessein, W., and Santos, T. (2006): "Adaptive Organizations." *Journal of Political Economy*, 114 (50), 956-995.

[11] Dewatripont, M. and Tirole, J. (2005): "Modes of Communication," *Journal of Political Economy*, 113 (6), 1217-1238.

[12] Dewatripont, M. (2006): "Costly Communication and Incentives," *Journal of the European Economic Association*, 4 (2-3), 253-268.

[13] Dolev, D. (1982): "The Byzantine Generals Strike Again." *The Journal of Algorithms*, 3 (1), 14-30.

[14] Dolev, D., Dwork, C., Waarts, O., and Yung, M. (1993): "Perfectly Secure Message Transmission," *Journal of the ACM*, 40 (1), 17-47.

[15] Forges, F. (1986): "An Approach to Communication Equilibria," *Econometrica*, 54 (6), 1375-1385.

[16] Forges, F. (1990): "Universal Mechanisms," *Econometrica*, 58, 1341-1364.

[17] Forges, F. (2009): "Correlated Equilibria and Communication in Games," in *Encyclopedia of Complexity and Systems Science (Robert A. Meyers ed.)*, 1587-1596, Springer, New York.

[18] Galeotti, A., Ghiglino, C., and Squintani, F. (2013). "Strategic Information Transmission in Networks." *Journal of Economic Theory*, 148 (5), 1751-1769.

[19] Gerardi, D. (2004): "Unmediated Communication in Games with Complete and Incomplete Information," *Journal of Economic Theory*, 114, 104-131.

[20] Gossner, O. (1998): "Secure Protocols or How Communication Generates Correlation," *Journal of Economic Theory*, 83, 69-89.

[21] Hagenbach, J., and Koessler, F. (2010): "Strategic Communication Networks," *Review of Economic Studies*, 77 (3), 1072-1099.

[22] Halpern, J. (2007): "Computer Science and Game Theory: A Brief Survey." in *Palgrave Dictionary of Economics (S. N. Durlauf and L. E. Blume, eds.)*, Palgrave MacMillan, 2008.

[23] Lamport, L., Shostak, R. E., and Pease, M. C. (1982): "The Byzantine Generals Problem" *The ACM Transactions on Programming Languages and Systems*, 4 (3), 382-401.

[24] Marshak, J., and Radner, R. *Economic Theory of Teams*, 1972, Yale University Press.

[25] Mookherjee, D. (2006): "Decentralization, hierarchies, and incentives: A mechanism design perspective," *Journal of Economic Literature*, 44 (2), 367-390.

[26] Myerson, R. B. (1982): "Optimal Coordination Mechanisms in Generalized mediator-Agent Problems," *Journal of Mathematical Economics*, 10 (1), 67-81.

[27] Myerson, R. B. (1986): "Multistage Games with Communication," *Econometrica*, 54, 323-358.

[28] Myerson, R. B. and Reny, M. (2015): "Sequential Equilibria of Games with Infinite Sets of Types and Actions," *Working Paper*.

[29] Rantakari, H. (2008): "Governing Adaptation," *Review of Economic Studies*, 75, 1257-1285.

[30] Renault, J., Renou, L., and Tomala, T., (2014): "Secure Message Transmission on Directed Networks," *Games and Economics Behavior*, 85 (1), 1-18.

[31] Renou, L. and Tomala, T. (2012): "Mechanism Design and Communication Networks." *Theoretical Economics*, 7 (3), 489-533.

2 Strategic Inventory Management in Capacity Constrained Supply Chains

Thomas J. Rivera, Marco Scarsini, Tristan Tomala

Abstract

We consider a supply chain with a single wholesaler facing random production disruptions and multiple retailers who decide how early to order their seasonal inventory. When the wholesaler is capacity constrained, there is a production bottleneck which can result in later orders not being fulfilled, imposing a penalty cost on those retailers. When holding inventory is costly, this makes order timing a strategic decision among retailers. We show that when the penalty cost is large then, in any Nash equilibrium, retailers stock their inventory inefficiently early as compared to the centralized optimum, imposing high inventory costs. We then show how pricing can help reduce this inefficiency but that above a certain penalty cost threshold, it is instead optimal to utilize a correlated equilibrium implementation scheme, generating a system of order time recommendations drawn from a joint distribution that are incentive compatible for the retailers to obey.

2.1 Introduction

In this paper we study a model of supply chain congestion whereby multiple retailers source their inventory from the same producer (which we refer to as the wholesaler). Retailers face deterministic seasonal demand which they must stock before the season begins to avoid paying a *penalty cost* (e.g. lost revenue from excess demand). Assuming holding inventory is costly, if the wholesaler is unconstrained in its production capacity then retailers will optimally order their inventory with just enough time for it to be produced and delivered before the deadline. On the other hand, if there is positive probability that the wholesaler is constrained, for example due to production disruptions coming from machine breakdowns and labor shortages, then the retailer's decision of when to order becomes strategic: multiple orders exceeding capacity creates a backlog which leads to longer lead times and a higher probability of paying the penalty cost. While inventory management has been studied in the context of supply disruptions and stochastic lead times, the innovation of this paper is to study a situation whereby the lead time is endogenous to the ordering decisions of the retailers. If all retailers order at the same time, this creates a production bottleneck whereas if orders are spread out, they can all be produced and delivered without a backlog and minimal lead times. Hence, this paper draws attention to the question of *when* to order seasonal inventory in the fact of supply disruptions.

The main result of the paper is to show that when the penalty for not stocking prior to the deadline is sufficiently large, then this leads to inefficiently early ordering of inventory when compared to the case whereby the retailers centralize their activities to minimize their total expected costs (inventory plus potential penalties). Hence, when multiple retailers independently order from a single wholesaler, this produces a situation with inefficiently high inventory costs, even if demand is deterministic. More concretely, we show that in any Nash equilibrium, retailers bunch their orders around the latest *individually rational* ordering time: the time at which the constrained wholesaler is guaranteed to meet any retailer's order prior to the deadline, independent of the ordering behavior of the other retailers. This implies that the risk of the wholesaler being constrained, and therefore the threat of missing the deadline, leads the retailers to pay the highest possible inventory cost. In contrast, a centralized system that minimizes the sum of total cost of all retailers would instead spread out their order times to ensure that all retailers get their orders before the deadline, without the duplication of the inventory costs inherent in the decentralized Nash Equilibria. The reason why the centralized optimum is not a Nash equilibrium is that the retailer with the earliest order will always have an incentive under the centralized optimum to delay their order. This comes from the fact that when doing so the retailer with the earliest order does not increase its own probability of being late but instead increases the probability of being late of the retailers who make the later orders. Hence, the retailer with the earliest order will always find it optimal to deviate from the centralized optimum in this fashion to save on inventory costs.

In order to remedy this issue, we show how prices can help resolve the inefficiency and their optimal design. Although prices help, they impose higher costs on the retailers and therefore are limited in their ability to achieve the same low cost of the centralized order flow. The optimal pricing scheme has an interesting structure: the wholesaler charges a premium starting from the *earliest* order time and increasing up to some later threshold time k and then charges zero premium for all orders made between the threshold time and the deadline. This pricing scheme generates a Nash equilibrium whereby a subset of retailers spread out their orders from the earliest ordering time to the threshold time and the remaining retailers play the symmetric Nash equilibrium of the remaining game with fewer players. While this Nash equilibrium is such that all retailers who spread out their orders pay the same *bad* equilibrium cost of the Nash equilibrium without prices, it allows for the remaining retailers to pay a lower cost in the Nash equilibrium of the game with fewer retailers. This also yields intuition for why the threshold time is interior (i.e. strictly between the earliest ordering time and the deadline): if it were equal to the earliest time then it generates the same game as without prices and if it were equal to the deadline time, then it would generate the

centralized optimum outcome (all orders are spread out), but at a cost equal to the worst Nash equilibrium cost without prices.

We believe that the inefficiency created by supply disruptions is relevant for the design of automated replenishment in inventory management. In an empirical study, van Donselaar et. al. (2010) show how retail managers in a supermarket chain regularly alter automated order suggestions (utilized to meet seasonal weekend demand) by pushing the order times to earlier days in the week. While those authors go on to explain this phenomenon as one of in-store handling costs, this paper proposes another theoretical justification: if the cost of holding inventory is small with respect to the lost revenue from not having inventory on hand, then retailers will optimally order their inventory much earlier whenever they believe there is a chance that the wholesaler may be capacity constrained. We illustrate in this paper how any system that suggests order times must be *incentive compatible* in the sense that it is a Nash equilibrium of the underlying inventory order game. To this end, we show how the wholesaler can utilize the correlated equilibrium solution concept to design a system that draws order times from a joint lottery and suggests to each retailer their order time, with the property that it is always optimal (i.e. incentive compatible) for the retailers to follow the suggestion. Importantly, there exists a threshold penalty cost such that this system of recommendations (which does not involve prices) achieves a higher level of efficiency than the pricing scheme. In addition, as the penalty cost increases, the sum of total expected costs to the retailers under the optimal correlated equilibrium approaches the cost of the centralized optimum. Using this insight, the retailer could implement an approximation of the centralized optimum by artificially increasing the penalty cost, for example, by refusing delivery or charging a high fee for deliveries made after the deadline. We further show how solving for the optimal correlated equilibrium consists in solving a linear program with the number of constraints weakly less than the number of players (depending on the underlying parameters).

2.1.1 Related Literature

In this paper we study a model of inventory management assuming that retailers compete for inventory when the wholesaler is capacity constrained. We believe that this is the first study of this type but relate the issue of order times to that of traffic congestion in a bottleneck first introduced by Vickrey (1969) and later generalized by Arnott et. al. (1990) and Hendrickson and Kocur (1981) in a continuous time framework. We show how when inventory orders create externalities for other retailers (by increasing their probability of not receiving their full shipment) then this creates a strategic game with very inefficient Nash equilibria. No such results exist in the congestion literature as this is the first paper to study

the case with a large penalty cost for arriving late.[38]

Another component of this paper is the fact that failing to stock inventory before a certain deadline leads to lost excess demand and therefore a large penalty. This was illustrated by the case study of a supply disruption incurred by Ericsson Corp. whereby a fire in the semiconductor plant of one of their component producers led to roughly $400 million in lost sales (see e.g. Mukherjee (2008)).[39] This further highlights another modeling assumption of our model which is that it is possible for such disruptions to go undetected (as had been the case for Ericsson) until it is essentially too late to find an alternative producer. A review of the literature on this issue of *lost sales inventory theory* is given by Bijvank & Vis (2011). While many papers study the issue of supply disruptions in the face of lost inventory, they typically assume a single retailer who must decide when to order from a single or multiple wholesalers (see e.g. Li (2017)).

Lost sales inventory models have also been studied with stochastic lead times: Ravichandran (1984), Buchanan and Love (1985), Beckmann and Srinivasan (1987), Johansen and Thorstenson (1993). These papers assume the production process can be disrupted leading to longer lead times in a situation with a single supplier and single retailer whereby disruptions are modeled as different exogenous random processes over lead times. Our contribution is to show how this increase in lead times can be endogenously determined by congestion in the ordering system when there are multiple retailers after a disruption has occurred.

The issue of inventory replenishment with continuous review and lost sales typically has focused on fixed order size policies (s, Q) whereby a fixed order of size Q is made once inventory drops below the reorder level s. The drawback of this literature is that it typically makes a restriction that at most one or two orders be outstanding at any given time (see. e.g. Hill (1992,1994) who study the two outstanding orders case with deterministic and stochastic lead times). We hope to illustrate with our model the necessity for understanding how multiple retailers utilizing a single supplier can naturally lead to situations whereby these assumptions break down. Another line of literature looks at periodic review policies with lost sales, holding costs, and stochastic lead times (see Karlin and Scarf (1958)). Again though, these papers focus on the issue of a single retailer and supplier and exogenously driven stochastic lead times.

The exercise closest in spirit to ours is done in Chen, Federguen, and Zheng (2001) who study a supply chain model with a single supplier and multiple retailers. They show how when demand fluctuates as a function of the retail price in the market then a centralized

[38]The closest paper resembling our model is the experimental study of Ziegelmeyer et. al. (2008).

[39]This case study typically focuses instead on Nokia who was also utilizing the same plant for components but who caught wind to the potential supply disruption due to the fire and avoided the losses faced by Ericsson by finding alternate sources for their components.

system maximizing total profits does substantially better than the decentralized problem and how a pricing mechanism can be devised that attains this optimal level of systemwide profits can be designed. In a similar vein, Perakis and Roels (2007) study the *Price of anarchy* in supply chain— the ratio of profits of the fully coordinated supply chain and the worst case decentralized supply chain — and study the efficiency of differing supply chain configurations in aiding in coordinating incentives and mitigating decentralized supply chain inefficiencies.

2.2 Model

We consider a supply chain whereby a single wholesaler supplies $I > 1$ independent retailers. Each retailer must build their inventory in order to meet some seasonal demand which we assume is deterministic and normalized to 1 for each retailer. Therefore, the only decision of the retailer is *when* to order its needed unit of inventory from the wholesaler in order to meet its demand. We assume that time is discrete (e.g. days) and that the production time plus delivery time (not necessarily equal to lead time) is equal to some integer β. Retailers face a deadline $t^\star$ so that all retailers would like have their inventory delivered before time $t^\star$ and face a penalty cost C if their order is delivered after $t^\star$. The common interpretation of this is that if the shelves are not stocked before the season begins then the excess demand will be lost, generating an opportunity cost of lost revenue equal to C. Further, retailers face an inventory storage cost g for each period that they store their inventory before $t^\star$: storing inventory for k periods costs $g \cdot k$. In such a setting, if the wholesaler had a large enough capacity, then all retailers would order their supply at time $t^\star - \beta$ to minimize on the storage cost while at the same time ensuring they will have their orders delivered before the deadline.

The main innovation is to assume that there is uncertainty about the capacity of the wholesaler. Namely, denoting by γ the per period capacity of the wholesaler — i.e. how many orders the wholesaler can deliver per one unit of time — we will assume that with probability $(1-p)$ the wholesaler has no capacity constraint $\gamma = +\infty$ so that all orders could be serviced in a single time period. With probability p though, the wholesaler is capacity constrained so that $\gamma < n$. A natural interpretation of this problem is that the wholesaler faces uncertainty in the production process so that in *normal times* the production capacity is large enough to meet all possible demand at any given time but with probability p production is disrupted (e.g. due to a break down of a machine or labor shortage) and therefore the wholesaler can only meet $\gamma < n$ demand per period. Importantly, we assume that the capacity is private information to the wholesaler and for simplicity that the disruption happens sufficiently early before retailers would optimally order their inventory (this time is specified below).

The role of the order timing will only matter when the wholesaler is capacity constrained in which case there will be a bottleneck in production which can lead to a backlog of orders. In this case, we assume that the wholesaler uses a first come first served priority to meet the orders with a uniform tie breaking rule:

Uniform Random Priority (URP): if Retailer i orders before Retailer j, then Retailer i's order is served before Retailer j's order (i has priority over j) regardless of whether there is a backlog or not. If a subset K of retailers order their inventory at the same time t, then retailer $i \in K$ is served with probability $\min\{1, \gamma/|K|\}$.[40] If Retailer i is not served at the time t at which they made their order, then a backlog is formed and they wait one period (with all other unserved retailers) and receive priority in the following periods over any new orders until they are served.

The following example demonstrates the priority rule.

Example 2.1. Suppose that the capacity of the wholesaler is $\gamma = 2$ and the production/delivery time is $\beta = 1$ day:

1.) If there is no backlog and 2 retailers submit their orders at the same time then both orders are served in $\beta = 1$ day.

2.) If there is no backlog and 3 retailers submit their orders on the same day then two of the retailers' orders are served in 1 day but a third user must queue in the backlog for an additional day and therefore have the order served in 2 days. The priority rule in this case assumes that the retailers whose orders are served first are chosen uniformly at random so that each of the three users has an ex-ante probability $\frac{1}{3}$ of having to queue and thus have their order served in 2 days and the complementary probability $\frac{2}{3}$ of not having to queue and have their project completed in 1 day.[41]

3.) If there is a backlog of 2 orders from retailers i_0 and j_0 at the time that two other retailers i_1 and j_1 submit their orders, then the tasks $\{i_0, j_0\}$ are completed first (taking 1 day plus the amount of time their orders were backlogged) and the tasks of users $\{i_1, j_1\}$ are backlogged for one day and then completed the following day, taking a total of 2 days to complete (regardless of any tasks submitted thereafter).

Note that assuming that the inventory cost is the opportunity cost of the space reserved to store the existing inventory (and therefore is paid whether the order is delivered or not) then this rule is equivalent to the rule whereby early orders are given priority but when $k > \gamma$ orders are made at the same time then the wholesaler rations the existing capacity, serving

[40] Equivalently, a random permutation π of K is drawn uniformly from the set of all permutations over K and Retailer i has priority over task j if and only if $\pi(i) < \pi(j)$.

[41] Of course, the URP correlates the tasks so that one retailer queues and 2 retailers do not queue with probability 1.

each retailer $\frac{\gamma}{k} < 1$ of their order. It may seem restrictive that retailers pay the inventory cost each period once the order is made, whether or not the order is physically delivered, but the results would still go through (albeit for higher cost of missing the deadline C) as long as the inability to meet the excess demand $1 - \frac{\gamma}{k}$ leads to an opportunity cost $(1 - \frac{\gamma}{k}) \cdot C$.

We will now summarize the *Inventory Ordering Game* played between the retailers. There are I retailers, each of which has one order to submit. Each retailer $i \in \mathcal{I} = \{1, ..., I\}$ chooses when to submit their order from the set of times $S_i = \mathbb{I}$, the set of integers, with the interpretation that if Retailer i chooses strategy s_i then she orders s_i periods before the deadline $t^\star$ at time $t^\star - s_i$. A strategy profile $\boldsymbol{s} \in S = S_1 \times \cdots \times S_I$ determines how early orders are submitted to the facility before the deadline and the uniform random priority establishes the waiting times.

Each retailer's objective is to minimize their total cost. A strategy profile $\mathbf{s} = (s_1, ..., s_I)$ determines when each retailer submits their order, and therefore, when each order is delivered. Together with the priority, this establishes the backlog at any given time, the lag times for each retailer (due to the backlog), and therefore the arrival time a_i for each Retailer i. Given a strategy profile $\mathbf{s}$, denoting by $\mathbf{s}_{-i}$ the profile of strategies excluding Retailer i's strategy, then in general the expected cost of Retailer i is given by

$$c_i(s_i, \mathbf{s_{-i}}) = g \cdot s_i + p \cdot f(a_i) \cdot \mathbb{P}(a_i > t^\star) \tag{2.1}$$

Namely, if Retailer i is late, that is if $a_i > t^*$, then she pays a penalty $f(a_i)$ where $f : \mathbb{I} \to \mathbb{R}$ is weakly increasing in a_i. In what follows, we will denote by $C := p \cdot f(t^\star + 1)$ the expected cost of being one period late so that $c_i(s_i, \mathbf{s_{-i}}) = g \cdot s_i + C \cdot \mathbb{P}(a_i > t^\star)$ and study the case where C is *large* with respect to g. As we note below, when C is large, then each retailer will receive their order at the latest one period after the deadline in all equilibria of the inventory game and therefore we will simply refer to the cost of missing the deadline by C. Note that in this expression, $\mathbb{P}(a_i > t^\star)$ is the probability that Retailer i's delivery is late conditional on the wholesaler being constrained so that $\mathbb{P}$ is determined only by the strategy profile s and the uniform random priority rule.

We will denote by $\sigma_i \in \Delta(S_i)$ a mixed strategy of Retailer i and $\boldsymbol{\sigma} \in \Sigma := \prod_{i \in \mathcal{I}} \Delta(S_i)$ a mixed strategy profile. The cost functions are extended to mixed strategies by taking expectations in the usual way, so that:

$$c_i(\boldsymbol{\sigma}) = \sum_{\boldsymbol{s} \in S} c_i(\mathbf{s}) \cdot \sigma_1(s_1) \dots \sigma_I(s_I)$$

is the expected cost of Retailer i under $\boldsymbol{\sigma}$. Given that the optimal ordering time of Retailer

i depends on the order time of the other retailers, then the parameters of this model $\Gamma := (\mathcal{I}, S, \mathbf{c}, g, C, p, \gamma, \beta)$ represents a strategic game which we dub the *Inventory Ordering Game*. In what follows we will be interested in the following two equilibrium solution concepts:

Definition 2.2. A *Nash equilibrium* of Γ is a profile of mixed strategies $\boldsymbol{\sigma}$ such that

$$c_i(\boldsymbol{\sigma}) \leq c_i(s_i, \boldsymbol{\sigma}_{-i}), \quad \text{for all } i \in \mathcal{I}, \text{ and all } s_i \in S_i.$$

A *correlated equilibrium* of Γ is a distribution $Q \in \Delta(S)$ over profiles of pure strategies such that

$$\sum_{\mathbf{s}_{-i}} c_i(s_i, \mathbf{s}_{-i}) Q(s_i, \mathbf{s}_{-i}) \leq \sum_{\mathbf{s}_{-i}} c_i(s_i', \mathbf{s}_{-i}) Q(s_i, \mathbf{s}_{-i}) \quad \text{for all } i \in \mathcal{I}, \text{ and all } s_i \in S_i.$$

The usual interpretation of a correlated equilibrium is that a *mediator* (in this case the wholesaler) draws a profile of strategies $\mathbf{s}$ with probability $Q(\mathbf{s})$ and recommends each Retailer i to order at time s_i without any further information. The distribution Q is then a correlated equilibrium if each retailer has an incentive to obey the recommendation knowing that the other order times $\mathbf{s}_{-i}$ are distributed according to the distribution $Q(\mathbf{s}_{-i}|s_i)$.

We will also consider the case whereby the retailer imposes a schedule of premium $r : \mathbb{I} \to \mathbb{R}$ such that $r(t) \geq 0$ is the premium (over price of inventory) each retailer pays when their order is made at time $t^\star - t$. In this case, the cost of retailer i under the strategy profile $\mathbf{s}$ is given by

$$c_i(s_i, \mathbf{s}_{-i}|r) := g \cdot s_i + r(s_i) + C \cdot \mathbb{P}(a_i > t^\star)$$

We denote by $\Gamma(r)$ the game Γ augmented by a premium schedule r with the Nash equilibrium solution concept extending in the natural way to this game.

The main focus of our paper is to study the efficiency of Nash and correlated equilibria of the inventory games Γ and $\Gamma(r)$. For each $\mathbf{s} \in S$ the *social cost* is defined as

$$\mathsf{SC}(\mathbf{s}) := \sum_{i=1}^{I} c_i(\mathbf{s}).$$

By taking expectations, we naturally extend this definition to any profile of mixed strategies $\boldsymbol{\sigma}$ and to any correlated distribution $Q \in \Delta(S)$.

Let $\mathcal{E}$ and $\mathcal{E}(r)$ represent the set of all Nash equilibria of Γ and $\Gamma(r)$ repspectively, and $\mathcal{C}$ the set of correlated equilibria of the game Γ. Then we will be interested in the following

objects:

$$\mathsf{Opt} = \min_{\mathbf{s}\in S} \mathsf{SC}(\mathbf{s}) \qquad \mathsf{WorstEq} = \max_{\boldsymbol{\sigma}\in\mathcal{E}} \mathsf{SC}(\boldsymbol{\sigma}) \qquad \mathsf{BestEq} = \min_{\boldsymbol{\sigma}\in\mathcal{E}} \mathsf{SC}(\boldsymbol{\sigma})$$

$$\mathsf{Premium} = \min_{r} \max_{\boldsymbol{\sigma}\in\mathcal{E}(r)} \mathsf{SC}(\boldsymbol{\sigma}) \qquad \mathsf{BestCE} = \min_{Q\in\mathcal{C}} \mathsf{SC}(Q)$$

whereby Opt is the optimal social cost if orders were centralized in order to minimize the sum of expected costs of the retailers, $\mathsf{WorstEq}$ the highest Nash equilibrium cost, BestEq the lowest Nash equilibrium cost, $\mathsf{Premium}$ is the lowest cost of the worst Nash equilibrium over all premium schedules r, and BestCE the lowest cost correlated equilibrium cost of the game Γ.

2.3 Simple Case

In what follows we will present the main results for the notationally simple case where $\gamma = \beta = g = 1$. The results then will be restated and proven for general values of γ, β, and g in the next section. Before stating these results we should note that in this paper we are interested in the case where C is large. Further, given that backlogs and lag times are relative to the number of retailers (which also dictates the relevant window of order times), our bounds for C will always be a function of the number of retailers I. To clarify, we note that in this simple case of $\gamma = \beta = g = 1$, no retailer will ever order before time $t^\star - I$. This is due to the fact that $t^\star - I$ is the latest time with the property that when ordering at time $t^\star - I$ Retailer i can be sure that his inventory will arrive before time $t^\star$ with probability 1. Therefore, ordering before time $t^\star - I$ is strictly suboptimal compared to ordering at time $t^\star - I$, independent of the remaining retailer's order times. Furthermore, this implies that the maximal cost that any retailer will pay is I (i.e. I times the inventory cost). The social optimum in this context is the outcome whereby one retailer orders in each period from time $t^\star - I$ to time $t^\star - 1$ yielding a social cost of $\sum_{j=1}^{I} j = \frac{I(I+1)}{2}$. The following theorem states how the best and worst Nash equilibrium compare to the social optimum as well as the best premium schedule and the best correlated equilibrium.

Theorem 2.3. *Suppose that $C > I^2$, then*
(1) $\mathsf{Opt} = \frac{I(I+1)}{2}$
(2) $\mathsf{WorstEq} = I^2 = \mathsf{Opt} + \frac{I(I-1)}{2}$
(3) $\mathsf{BestEq} = I + (I-1)^2 = \mathsf{Opt} + \frac{(I-1)(I-2)}{2}$
(4) $\mathsf{Premium} = \frac{3}{4}I^2 = \mathsf{Opt} + \frac{I(I-2)}{4}$

Further, when $C > 2 \cdot (I+1) \cdot I$ then

(5) $\mathsf{BestCE} < (1-\alpha(C)) \cdot \mathsf{Opt} + \alpha(C) \cdot \mathsf{BestEq} = \mathsf{Opt} + \alpha(C) \cdot \frac{(I-1)(I-2)}{2}$

where $\alpha(C) = \frac{(I+1)\cdot I}{C} < \frac{1}{2}$

Remark 2.4. The theoretical bounds obtained for our results are weak as can be seen from the inequality in (5). The reason why we cannot obtain tight bounds is that the equations which characterize the optimal probabilities that the best correlated equilibrium puts on outcomes in the support are recursive and therefore it is not possible to obtain closed form solutions, making the analysis for tight bounds intractable. Important to note though is that we solve for the optimal support of the cost minimizing correlated equilibrium and only lack a closed form solution for the probabilities that that equilibrium puts on the optimal support. Further, solving numerically for those probabilities is extremely simple as solving for the best correlated equilibrium in general amounts to a simple linear program (especially given that the support is known). At some point we utilize a simple correlated equilibrium with closed form probabilities over the support to help us obtain our results which could also be utilized as a simple heuristic.

The results of Theorem 2.3 are illustrated in Figure ?? when $I = 10$ and increasing values of $C = I^2, ..., I^3$. The point to make here is that the best correlated equilibrium

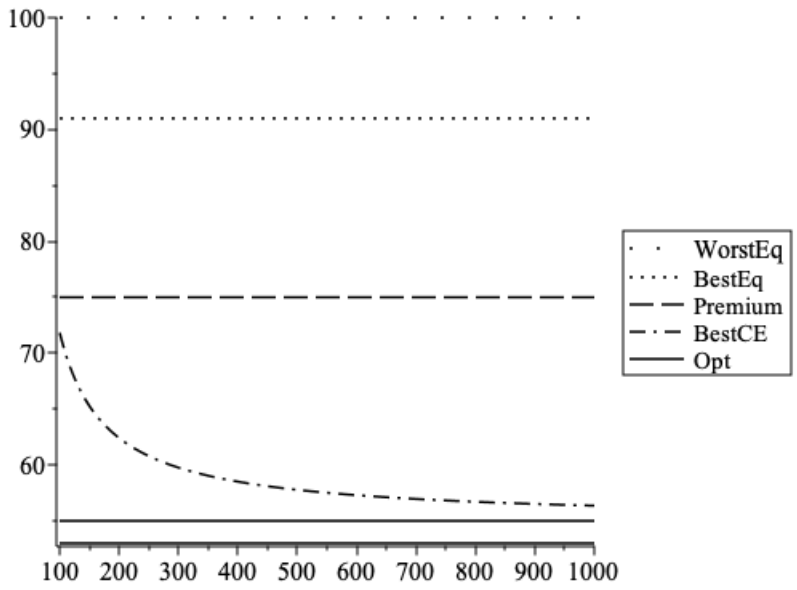


Figure 20: The social cost of the cost minimizing correlated equilibrium $Q^\star$ as a function of C in the simple case $\gamma = \beta = 1$ and $g = 1$.

performs substantially better than the best and worst Nash equilibria, and substantially

better than the Nash equilibrium induced by the optimal premium schedule whenever C is large. Further, as can be observed from Figure ??, as the penalty cost increases (in relation to the inventory cost), the best correlated equilibrium cost approaches the centralized optimum. This can also be observed from the statement of Theorem 2.3 whereby $\alpha(C) \to 0$ as $C \to +\infty$. Finally, we would like to point out the fact that if the wholesaler had the ability to increase C, for example by refusing delivery or charging a sufficiently large premium for orders delivered after the deadline, then they could obtain a total cost arbitrarily close to the social optimum. Importantly, the mere act of increasing C would be simply to enforce the correlated equilibrium, but in fact the correlated equilibrium property is such that no retailer is never late so that such a cost, if credible, would never have to be levied in equilibrium.

We will now illustrate these results with an example.

Example 2.5. Suppose $I = 4$. Then, as explained above, no retailer will ever order earlier than time $t^\star - 4$. First note that the socially optimal *strategy* $\boldsymbol{\sigma}^{opt}$ has one retailer order at each time $t = t^\star - 4, ..., t^\star - 1$ and in this case has a social cost of $\mathsf{Opt} = \mathsf{SC}(\boldsymbol{\sigma}^{opt}) = \frac{I(I+1)}{2} = 4 + 3 + 2 + 1 = 10$.

Whenever $C > 4^2$ the worst Nash equilibrium is given by the symmetric strategy $\boldsymbol{\sigma}^{worst}$ whereby all retailers order at time $t^\star - 3$ with probability $\hat{\sigma} \in (0, 1)$ and order at time $t^\star - 4$ with probability $1 - \hat{\sigma}$. We also know that $t^\star - 4$ is a *safe order time* so that $c_i(t^\star - 4, \sigma_{-i}) = 4$ for all σ_{-i}. Now, if Retailer i orders at time $t^\star - 3$ while the remaining retailers order according to $\boldsymbol{\sigma}_{-i}^{worst}$, then the expected cost of Retailer i is

$$c_i(t^\star - 3, \boldsymbol{\sigma}_{-i}^{worst}) = 3 + \frac{\hat{\sigma}^3}{4} \cdot C$$

To understand why, we note first that Retailer i orders 3 periods early and therefore pays an inventory cost of 3 plus C times the probability of arriving late (note that no retailer is more than one period late given the support of $\boldsymbol{\sigma}^{worst}$). In this case, given that the capacity is 1, a retailer is late when ordering at time $t^\star - 3$ only if there is a production disruption (which happens with probability p) and *all other* retailers order at time $t^\star - 3$. Therefore, from Retailer i's perspective, he can only be late if every other retailer also orders at time $t^\star - 3$ which happens with probability $\hat{\sigma}^3$. Finally, given that only one retailer is late in this scenario — the wholesaler can service 3 of 4 retailers before the deadline— it is the uniform random priority determines which retailer is late. In this case, conditional on all retailers ordering at time $t^\star - 3$, each is late with probability $\frac{1}{4}$ as dictated by the URP so that Retailer i is late with probability $\frac{\hat{\sigma}^3}{4}$ when ordering at time $t^\star - 3$.

Now, if $\boldsymbol{\sigma}^{worst}$ is a Nash equilibrium then it must be the case that retailers are indifferent

between ordering at time $t^\star - 3$ and $t^\star - 4$ which is the case whenever

$$4 = 3 + \frac{\hat{\sigma}^3}{4} \cdot C$$

which implies

$$\hat{\sigma} = (\frac{4}{C})^{\frac{1}{3}}$$

Most importantly, given $supp(\boldsymbol{\sigma}_i^{worst}) = \{t^\star - I, t^\star - (I-1)\}$ implies $c_i(\boldsymbol{\sigma}^{worst}) = 4$ for all $i \in \{1,2,3,4\}$ which implies that $\mathsf{WorstEq} = \mathsf{SC}(\boldsymbol{\sigma}^{worst}) = I^2 = 16$.

We show below that the best Nash equilibrium $\boldsymbol{\sigma}^{best}$ is such that one retailer orders at time $t^\star - 4$ with probability 1, and the remaining three retailers randomize between ordering at time $t^\star - 3$ and time $t^\star - 2$. In this case the social cost is $\mathsf{WorstEq} = \mathsf{SC}(\boldsymbol{\sigma}^{best}) = I + (I-1)^2 = 4 + 3 \cdot 3 = 13$.

Iext, note that the optimal premium schedule r is such that $r(1) = r(2) = r(4) = 0$ while $r(3) = 1$. In this case, any Nash equilibrium has two retailers order at times $t^\star - 4$ and $t^\star - 3$ with probability 1 (both paying a cost of 4), and the remaining retailers play the worst Nash equilibrium of the 2-retailer game costing in expectation 2 each. This is the general design of the best premium schedule which sets $r(t) = I - t$ for $t = k+1, ..., I$ and $r(t) = 0$ for all $t = 1, ..., k$ with the optimal $k^\star = \frac{I}{2}$. This leads the $\frac{I}{2}$ retailers ordering $t = \frac{I}{2} + 1, ..., I$ periods early to all pay a cost of I and the remaining $I - k^\star = \frac{I}{2}$ retailers to pay the worst Nash equilibrium cost of the $\frac{I}{2}$ retailer game which is equal to $\frac{I}{2}$. In this case we obtain $\mathsf{Premium} = \frac{I}{2} \cdot I + (\frac{I}{2}) \cdot (\frac{I}{2}) = \frac{3}{4} I^2$

Below we characterize in the general case the best correlated equilibrium of this game and show that it only randomizes over $I - 1$ *outcomes*. In our simple case, these three outcomes are $\boldsymbol{\xi}^2 = (1,1,1,1)$, $\boldsymbol{\xi}^3 = (0,2,1,1)$, and $\boldsymbol{\xi}^4 = (0,0,3,1)$. For example, $\boldsymbol{\xi}^3$ is the outcome whereby 2 retailers order at time $t^\star - 2$, one retailer orders at time $t^\star - 3$ and, one retailer orders at time $t^\star - 4$. The intuition is that each of the outcomes $\boldsymbol{\xi}^j$ ensures that whenever a retailer is told to order at time $t^\star - j$ she will be late with positive probability if she orders instead at time $t^\star - j + 1$ and the outcome drawn was $\boldsymbol{\xi}^j$.[42] Hence, if whenever a retailer is told to order at time j and she believes that the outcome is $\boldsymbol{\xi}^j$ with sufficiently high probability, then she will never deviate by ordering any time later than j. Further, we show that in addition to this being the support of the cost minimizing correlated equilibrium, we pin down the probabilities as well. For example when $C = 20$, the cost minimizing correlated

[42] We show below that this is sufficient to prevent deviations to any time $t > t^\star - j$.

equilibrium yields a distribution over outcomes:

$$(\hat{Q}^o(\boldsymbol{\xi}^4), \hat{Q}^o(\boldsymbol{\xi}^3), \hat{Q}^o(\boldsymbol{\xi}^2)) = (\frac{20}{100}, \frac{21}{100}, \frac{59}{100})$$

where $Q^o(\boldsymbol{\xi}^j)$ is the probability of the outcome $\boldsymbol{\xi}^j$. From these probabilities we can obtain a correlated equilibrium by drawing an outcome $\mathbf{x} \in \{\boldsymbol{\xi}^2, \boldsymbol{\xi}^3, \boldsymbol{\xi}^4\}$ according to the probability $Q^o(\mathbf{x})$ and then assigning each retailer to a single order time associated with $\mathbf{x}$ uniformly at random. For example if $\boldsymbol{\xi}^3 = (0, 2, 1, 1)$ is drawn from Q^o, then the correlated equilibrium would choose a permutation π of $\{1, 2, 3, 4\}$ uniformly at random from all possible permutations and suggest retailer $\pi^{-1}(4)$ to order at time $t^\star - 4$, $\pi^{-1}(3)$ to order at time $t^\star - 3$, and retailers $\pi^{-1}(2)$ and $\pi^{-1}(1)$ to order at time $t^\star - 2$. Similarly, under $\boldsymbol{\xi}^4 = (0, 0, 3, 1)$, the correlated equilibrium would suggest retailer $\pi^{-1}(4)$ to order at time $t^\star - 4$, and retailers $\pi^{-1}(3), \pi^{-1}(2)$, and $\pi^{-1}(1)$ to order at time $t^\star - 3$. Note, that we present this format of the result due to the fact that this is the technique utilized to characterize the best correlated equilibrium as it is much easier to work with probabilities over outcomes as opposed to probabilities over strategies.

In the case where $C = 20$, we then obtain

$$\mathsf{BestCE} = \mathsf{SC}(Q^\star) = \frac{20}{100} \cdot 13 + \frac{21}{100} \cdot 11 + \frac{59}{100} \cdot 10 = 10.81$$

if instead $C = 32$ then we obtain $\mathsf{SC}(Q^\star) = \frac{1579}{150} \approx 10.53$, and finally if $C = 100$ we obtain $\mathsf{SC}(Q^\star) = 10.1524$. Further, we obtain this by utilizing the probabilities put on the three outcomes

$$(\hat{Q}^o(\boldsymbol{\xi}^4), \hat{Q}^o(\boldsymbol{\xi}^3), \hat{Q}^o(\boldsymbol{\xi}^2)) = (\frac{4}{C}, \frac{1}{C} + \frac{24}{C^2}, 1 - \frac{5}{C} - \frac{24}{C^2})$$

which we prove is optimal as a corollary to the general case below.

2.3.1 Prices

In this subsection we will present our result on the optimal premium schedule. Important to note is that charging higher prices depending on the order time will add to the cost of the retailers. If this price wasn't a concern then the wholesaler could achieve the efficient outcome by simply charging a premium for ordering t periods before the deadline of $r(t) = g \cdot (I - t)$. In this case, all retailers would be indifferent between the time that they order, conditional on not missing the deadline, as ordering at time $t^\star - t$ would result in them paying a premium equal to the cost of ordering at $t^\star - I$. What we can show though is that a different pricing scheme can actually lower the expected cost of the retailers even when we include the premium charged in the cost function.

Theorem 2.6. *Assume that I is even. Then the optimal schedule of premiums $\mathbf{r} = (r(1), ..., r(I))$ are such that $r(t) = I - t$ for all $t = \frac{I}{2} + 1, ..., I$ and $p_t = 0$ for all $t = 1, ..., \frac{I}{2}$.*

The intuition behind such a schedule is that it makes all retailers indifferent between ordering at all times $t^\star - I$ to $t^\star - k$ for some k (which is equivalent to paying an inventory cost of I). Under such a pricing schedule, any Nash equilibrium will have a single retailer order at each time $t^\star - I$,...,$t^\star - k$ and the remaining $k-1$ retailers play a Nash equilibrium of the equivalent game with $I' := k - 1$ retailers. In that case, the worst case sum of costs under such a premium schedule is given by $(I - k + 1) \cdot I + (k - 1) \cdot (k - 1)$ and the value of k that minimizes the sum of costs (when I is even) is $k^\star = \frac{I}{2} + 1$.

What is happening under such a construction is that the retailers that order the earliest pay a higher cost (equal to the worst NE cost), but allow the remaining retailers to play a more efficient equilibrium and therefore pay a lower cost. The fact that $k^\star$ is an interior solution simply comes from the tradeoff of making more retailers pay the worst NE cost (hence $k^\star > 1$) which is weighed off against the benefit of the remaining retailers paying a lower cost (hence $k^\star < I$).

2.4 General Results

We will now proceed to restate and prove the simple results above for general γ, β, and g. In order to simplify notation we will assume, without loss, that $t^\star = 0$ so ordering t periods early implies ordering at time $-t$. Before stating the main results, we will make some simple observations which we will make reference to in the remainder of the paper.

Lemma 2.7. *For the game $\Gamma = (\mathcal{I}, S, \mathbf{c}, g, C, \gamma, \beta)$, denote for any $t > 0$,*

$$\mu(t) := \gamma(t - \beta + 1).$$

and

$$\tau := \beta + \lceil \frac{I}{\gamma} \rceil - 1, \tag{2.2}$$

(1) At most $\mu(t)$ retailers can order at time $-t$ without any of them being late.

(2) Time $-\tau$ is the latest ordering time such that each retailer can guarantee their delivery will arrive on time with probability 1 when ordering at time $-\tau$, for all ordering times of the other retailers.

Proof. See appendix. □

Naturally, in the case whereby $\gamma = \beta = 1$ we have $\tau = I$ and $\mu(t) = t$ so that $-I$ is the latest *safe time* and at most t retailers can order any any given time $-t$ without any of them being late.

2.4.1 Social Optimum

Assuming that the cost C of being late is large enough, it is easy to construct a socially optimal strategy profile. Choose arbitrarily a subset of γ retailers and have them order at time $-\beta$, so that each of them incurs a cost $g \cdot \beta$. Among the remaining retailers, take another subset of γ retailers and have them order at time $-(\beta + 1)$, so that each of them incurs a cost $g \cdot (\beta + 1)$. Continue like that until all retailers are assigned a strategy. The last batch contains $I - \gamma(\lceil \frac{I}{\gamma} \rceil - 1) = I - \mu(\tau - 1)$ retailers who pay $g \cdot \tau$. In what follows we will refer to this constructed strategy as $\boldsymbol{\sigma}^{opt}$. This construction produces the following result.

Lemma 2.8. *The socially optimal cost is such that*

$$\min_{\sigma} \mathsf{SC}(\boldsymbol{\sigma}) \leq SC(\sigma^{opt}) = \gamma \sum_{j=\beta}^{\tau-1} g \cdot j + (I - \mu(\tau - 1))\, g \cdot \tau. \tag{2.3}$$

If $C \geq g \cdot (\tau + 1)$, then (2.3) holds with equality.

Proof. See appendix. □

2.4.2 Worst Nash Equilibrium Cost

The first main observation is the fact that whenever $C \geq g \cdot (\tau - \beta)$ then the game Γ has no pure Nash equilibria.

Lemma 2.9. *If $C \geq g \cdot (\tau + 1)$ and $I > 2\gamma$, then the game Γ admits no pure Nash equilibria.*

Proof. See appendix. □

Note that when $I \leq 2\gamma$ then all players either depart at time $-\beta$ or time $-(\beta + 1)$ and there exists a pure Nash equilibrium when C is large whereby γ players depart at time $-\beta$ and the remaining players depart at time $-(\beta + 1)$. For the remainder of the paper we will maintain the assumption that $I > 2\gamma$. This assumption simply states that when capacity constrained it would take longer than two periods for the wholesaler to service all retailers without any inventory arriving late.

Lemma 2.7 says that $g \cdot \tau$ is the highest cost that any retailer should pay without being late and therefore the highest cost they would be willing to incur in any equilibrium (otherwise

they would have a profitable deviation by ordering τ periods before the deadline). Hence no retailer will ever order more than τ periods before the deadline. The next result states that in the worst Nash equilibrium all players pay $g \cdot \tau$ in expectation.

Theorem 2.10. *If $C > I \cdot g \cdot \tau$ then any symmetric Nash equilibrium $\boldsymbol{\sigma}$ of the game Γ is such that $c_i(\boldsymbol{\sigma}) = g \cdot \tau$ for all $i \in \mathcal{I}$: all retailers pay the highest possible cost and therefore* $\mathsf{WorstEq} = I \cdot g \cdot \tau$.

Proof. See appendix. □

The intuition for this result is the following. Suppose that C is so large that no retailer wants to be late even with a small (yet positive) probability and consider a symmetric mixed equilibrium. If retailers do not choose τ with positive probability, then all of them order no more than $\tau - 1$ periods before the deadline, but then, due to the capacity constraint, the probability of being late is positive and bounded away from 0. Therefore, when C is large, a retailer would prefer to order at time $-\tau$ and given that strategies are symmetric and retailers are indifferent between the pure strategies they mix over in any mixed Nash equilibrium, this implies their payoff must be equal to $g \cdot \tau$. Finally, we show that the threshold for C in order for this equilibrium property to hold is precisely $C > I \cdot g \cdot \tau$.

2.4.3 Best Nash Equilibrium Cost

In this section, we consider the best Nash equilibrium and show that its social cost is not too different from the one of symmetric equilibria. Therefore, all equilibria are close to achieving the worst cost.

Theorem 2.11. *If $C > I \cdot g \cdot \tau$ then any Nash equilibrium $\boldsymbol{\sigma}$ of $\Gamma(I, \gamma, \beta, r, f)$ yields a social cost*

$$\mathsf{SC}(\boldsymbol{\sigma}) \geq (I - \mu(\tau - 1)) \cdot g \cdot \tau + \mu(\tau - 2) \cdot g \cdot (\tau - 1) \geq SC(\boldsymbol{\sigma}^{opt}) + \gamma \sum_{j=\beta}^{\tau-2} g \cdot (\tau - \beta + 1 - j)$$

$$= SC(\boldsymbol{\sigma}^{opt}) + \frac{\gamma}{2}(I - \beta - 1)(I - 3\beta + 4)$$

Proof. See appendix. □

The main intuition of this theorem comes from the fact that whenever $C > I \cdot g \cdot \tau$ then at least $I - \mu(\tau - 1)$ retailers must order at time $-\tau$ (as shown in the proof of Theorem 2.10), otherwise at least one retailer will be late for sure and therefore some retailer must pay a cost greater than τ creating an optimal deviation to time $-\tau$. Then, if any retailer mixes

over time $-\tau$ then it must be that that retailer is late with strictly positive probability if ordering instead at time $-(\tau-1)$ which can only be the case if $\mu(\tau-1)$ *other* retailers mix over time $-(\tau-1)$. In the simple case where $\gamma=\beta=1$ this implies that at least one player mixes over time $-I$ and the remaining $I-1$ players mix over time $-(I-1)$. In this case, it could potentially be that $I-\mu(\tau-1)>1$ and therefore some of the players who mix over time τ also mix over time $\tau-1$. Hence, our best lower bound on the Nash equilibrium social cost in the general case is not tight as it is in the simple case, and above we express our most notationally convenient lower bound. Note though that Theorem 2.11 implies that at least $I-\gamma$ players pay $g\cdot(\tau-1)$ hence, the improvement is not much better than the worst Nash equilibrium.

2.4.4 Premiums

We will now present our result regarding the optimal premium schedule $r^\star$ and the (worst) Nash equilibrium cost that this schedule induces in the game $\Gamma(r^\star)$.

Theorem 2.12. *Suppose $C > I\cdot g\cdot\tau$,*
(1) The premium schedule $r^\star$ that minimizes the social cost of the worst Nash equilibrium of $\Gamma(r)$ is such that $r(t)=g\cdot(\tau-t)$ for all $t=\tau-k^\star+1,...,\tau$ and $r(t)=0$ for all $t=\beta,...,\tau-k^\star$ where

$$k^\star = \underset{k\in\{\beta+1,...,\tau-1\}}{\operatorname{argmin}} \; -\mu(\tau-k)\cdot g\cdot k$$

When $\frac{I}{\gamma}$ is an integer then $k^\star=\frac{I}{2\gamma}$ and

$$\text{Premium} = \frac{3}{4}\cdot I\cdot g\cdot\tau + \frac{1}{4}\cdot I\cdot g\cdot(\beta-1)$$

2.5 Correlated equilibria

In this section, we characterize some properties of the cost minimizing correlated equilibria of the game Γ. In particular, we characterize necessary and sufficient correlated equilibrium conditions and an optimal restriction of the support of correlated equilibrium outcomes.

2.5.0.1 Correlated equilibria. Consider the game Γ. A probability measure $Q\in\Delta(S)$ over strategy profiles is a correlated equilibrium if for any $i\in\mathcal{I}$ when a profile is drawn from Q and its i-th component s_i is recommended to Retailer i, then Retailer i cannot improve their expected cost by choosing any strategy $s'\neq s_i$. In other words, for each Retailer i and

each t such that $Q(s_i = t) > 0$,

$$\sum_{\boldsymbol{s} \in S} Q(\boldsymbol{s}|s_i = t) c_i(t, \boldsymbol{s}_{-i}) \leq \sum_{\boldsymbol{s} \in S} Q(\boldsymbol{s}|s_i = t) c_i(t', \boldsymbol{s}_{-i})$$

for all $t' \in \{\beta, \ldots, \tau\}$.

Observe that any Nash equilibrium is a correlated equilibrium, and that from Theorem 2.10, for large C, we know that all symmetric equilibria yield the highest cost $g \cdot \tau$. Therefore, these are also the worst correlated equilibria. Thus, we focus now on finding the best correlated equilibria. Further, in the spirit of the exercise, we will restrict our attention to correlated equilibrium outcomes whereby no retailer is late.

2.5.0.2 Outcomes. We define an outcome of the game as the distribution of retailers over order times $t \in \{\beta, \ldots, \tau\}$. Formally, the set of outcomes of the game Γ is defined as

$$X = \left\{ \boldsymbol{x} \in \mathbb{I}^{\tau} : \sum_{j=\beta}^{\tau} x_j = n \right\},$$

where for each $t \in \{\beta, \ldots, \tau\}$, x_t denotes the number of retailers who order at time $-t$. The social cost of an outcome (assuming no retailer is late under that outcome) is

$$\mathsf{SC}^o(\boldsymbol{x}) = \sum_{j=\beta}^{\tau} x_j \cdot g \cdot j.$$

For each strategy profile $\boldsymbol{s} \in S$, denote $\boldsymbol{x}^{\boldsymbol{s}}$ the outcome induced by the pure strategy profile $\boldsymbol{s}$, that is for each $k = \beta, \ldots, \tau$, $x_k^{\boldsymbol{s}} = |\{i \in I : s_i = k\}|$. Let $S(\boldsymbol{x}) = \{\boldsymbol{s} \in S : \boldsymbol{x}^{\boldsymbol{s}} = \boldsymbol{x}\}$ be the set of pure strategies that induce the outcome $\boldsymbol{x} \in X$. A correlated distribution $Q \in \Delta(S^I)$ induces a distribution $Q^o \in \Delta(X)$ over outcomes defined for each $\boldsymbol{x}$ as

$$Q^o(\boldsymbol{x}) := \sum_{\boldsymbol{s} \in S(\boldsymbol{x})} Q(\boldsymbol{s}).$$

Observe that the social cost of a correlated distribution can be calculated from the distribution over outcomes:

$$\mathsf{SC}(Q) = \mathsf{SC}^o(Q^o) = \sum_{x} Q^o(\boldsymbol{x}) \, \mathsf{SC}^o(\boldsymbol{x}).$$

We will now characterize the set of outcomes in the support of the cost minimizing

correlated equilibrium. Define the set of outcomes whereby no retailer is late as follows:

$$Y = \left\{ \boldsymbol{x} \in X : \sum_{j=\beta}^{t} x_j \leq \mu(t), \text{ for all } t = \beta, \ldots, \tau \right\}.$$

Recall that $\mu(t) = \gamma(t - \beta + 1)$ is the number of retailers who can feasibly order between β and t without exceeding the capacity. Therefore as we will now formally state, Y is the set of outcomes whereby no retailer arrives late.

Lemma 2.13. *No retailer is late with probability 1 under the outcome $\boldsymbol{x}$ (i.e. $\mathbb{P}_x(\max_{i \in I} a_i(x) > 0) = 0$) if and only if $\boldsymbol{x} \in Y$.*

Proof. See appendix. □

Now, we define a set of outcomes which satisfy a weak version of correlated equilibrium incentive compatibility and show that this is sufficient for full incentive compatibility. Namely, we want to make sure that no retailer who is recommended to order at time $-k$, has an incentive to deviate and order one period later at time $-(k-1)$.

Let S^Y be the set of strategy profiles $\boldsymbol{s}$ such that the induced outcome $\boldsymbol{x}^{\boldsymbol{s}} \in Y$.

Definition 2.14. A pure strategy profile $\boldsymbol{s}$ enforces order k for retailer $i \in I$ if $s_i = k$ and, whenever the remaining retailers play $\boldsymbol{s}_{-i}$, if Retailer i deviated from k to $k-1$, then they would be late with positive probability.

Denote for any $i \in I$, by $S^{ik} \subset S^Y$ the set of strategy profiles $\boldsymbol{s}$ such that $s_i = k$, and by $Z^{ik} \subset S^{ik}$ the set of strategy profiles that enforce order k for Retailer i. An outcome $\boldsymbol{x}$ enforces order k if $x_k \geq 1$ and for any $i \in I$, if any pure strategy $\boldsymbol{s} \in S(\boldsymbol{x})$ is such that $s_i = k$, then $\boldsymbol{s}$ enforces order k for Retailer i.

The next proposition characterizes the set of correlated equilibria. Importantly, it states that we only need to check incentive compatibility for one period ahead deviations. Further, the only relevant strategy profiles (outcomes) are those in S^{ik} (Z^{ik}) for $i = 1, ..., I$ and $k = \beta + 1, ..., \tau$.

Lemma 2.15. *$Q \in \Delta(S^Y)$ is a correlated equilibrium of the game Γ if and only if for all $i \in I$ and all $k = \beta + 1, \ldots, \tau$,*

$$\sum_{\boldsymbol{s} \in Z^{ik}} Q(\boldsymbol{s}) \geq \frac{g \cdot (\mu(k-1) + 1)}{C} \left[\sum_{\boldsymbol{s} \in S^{ik}} Q(\boldsymbol{s}) \right]$$

Proof. See appendix. □

This result shows that it is enough to deter deviations from $-k$ to $-(k-1)$. This results in one single linear constraint on Q for each possible k. The intuition for this result is that the probability of being late when deviating from $-k$ to $-(k-1)$ is so large that it deters the one period ahead deviation. But, then when deviating to any time earlier than $-(k-1)$ this implies that the retailer is late for sure whenever he would have been late with positive probability under the one period ahead deviation (via the URP). Then, the fact that the deviation from $-k$ to $-(k-1)$ is not profitable is a sufficient condition to ensure that being late for sure when you would have been late with positive probability is never optimal. Note that we only consider $k \geq \beta + 1$ because whenever a retailer is told to order at time $-\beta$ she pays the lowest possible cost and therefore never has an incentive to deviate.

Lemma 2.16. *For all $k \in \{\beta+1, \ldots, \tau\}$,*
(1) the set of outcomes $x \in Y$ that enforce order k is

$$X^k := \{\boldsymbol{x} \in Y : x_k \geq 1, x_{k-1} = \mu(k-1)\}.$$

(2) The set of pure strategies that enforce order $k \in \{\beta+1, \ldots, \tau\}$ for Retailer i is given by $Z^{ik} = \{\boldsymbol{s} \in S^Y : s_i = k, \boldsymbol{x}^{\boldsymbol{s}} \in X^k\}$.

Finally, this allows us to state the next result which is that the only outcomes that matter for the correlated equilibrium condition are those outcomes that enforce time $k = \beta+1, ..., \tau$.

Lemma 2.17. *Let Q be a correlated equilibrium of the game Γ then*

$$\sum_{\boldsymbol{x} \in X^k} x_k Q^o(\boldsymbol{x}) \geq \frac{g \cdot (\mu(k-1)+1)}{C} \cdot \sum_{\boldsymbol{x} \in Y} x_k Q^o(\boldsymbol{x}) \tag{2.4}$$

for all $k = \beta+1, \ldots, \tau$.

Here it is clear to see that the higher the probability that any correlated equilibrium puts on outcomes $x \in Y$ with $x_k > 0$ but $x \notin X^k$, the larger the right hand side and therefore the higher the probability the correlated equilibrium must put on outcomes $x \in X^k$. Using the fact that any optimal correlated equilibrium, puts probability $1 - \sum_{x \in Y \setminus \{\boldsymbol{x}^{opt}\}} Q^o(x)$ on the social optimum $\boldsymbol{x}^{opt}$ (the least costly outcome that enforces time $\beta+1$) then implies that the optimal correlated equilibrium will only randomize over outcomes x that enforce order time k for some $k = \beta+1, ..., \tau$. This is summarized in our next result.

Theorem 2.18. *Let $Q \in \Delta^U(S^Y)$ be a cost minimizing correlated equilibrium and suppose $C \geq g(\mu(\tau-1)+1)$. Then,* $\text{supp}(Q^o) \setminus (\cup_{k=\beta+1}^{\tau} X^k) = \emptyset$.

Remark 2.19. In an earlier version of the paper we included a theorem which stated that for some $\bar{C}$ whenever $C > \bar{C}$ the optimal correlated equilibrium randomizes only over a subset of outcomes in $\cup_{k=\beta+1}^{\tau} X^k$. We removed this theorem from this version of the paper as it depended on an extremely long and convoluted proof (hence the unknown bound $\bar{C}$) which did not add much intuition to the problem: given that optimizing over the set $\cup_{k=\beta+1}^{\tau} X^k$ to satisfy the equilibrium conditions is a linear program further limiting the support does not present much practical value. In either case, our numerical solutions confirm this support is optimally restricted to these outcomes which we now present.

For each $k \in \{\beta+1, \ldots, \tau\}$ define the outcome $\boldsymbol{\xi}^k$ as follows

$$\xi_t^k = \begin{cases} 0 & \text{if } t < k-1 \\ \mu(k-1) & \text{if } t = k-1 \\ \gamma & \text{if } k \leq t < \tau \\ I - \mu(\tau - 1) & \text{if } t = \tau. \end{cases}$$

In words, $\boldsymbol{\xi}^k$ is the outcome where the maximum number of retailers are recommended to order at time $-(k-1)$ without exceeding the capacity and a single retailer is recommended to order at time $-k$. This has the property to enforce order recommendation $-k$ in the sense that if a retailer deviates from $-k$ to $-(k-1)$ then they are late with positive probability. Thus, when C is large they will not want to deviate from $-k$ to $-(k-1)$ and therefore by the last step of the previous proof they will neither want to deviate from $-k$ to $-j < -k$. Finally, the remaining retailers are allocated under the outcome $\boldsymbol{\xi}^k$ according to capacity (i.e. smoothed out) so as to get as close a possible to the social optimum with $\boldsymbol{\xi}^{\beta+1}$ equal to the socially efficient outcome. This is the logic behind which utilizing these outcomes is most efficient, yet a simple proof eludes us.

Our next result states the final conditions for the optimal correlated equilibrium distribution.

Theorem 2.20. *Let $Q \in \Delta^U(S^Y)$ be a cost minimizing correlated equilibrium. If $C > g(\mu(k-1)+1)$ then condition 2.4 of Lemma 2.17 holds with equality for all $k = \beta+2, ..., \tau$.*

Proof. See appendix. □

Finally, we will conclude by stating the relationship between the social cost of the best correlated equilibrium and the centralized optimum.

Theorem 2.21. *(1) Whenever* $C \geq 2 \cdot g \cdot \mu(\tau+1) \cdot \mu(\tau)$,

$$\mathsf{BestCE} < (1-\alpha(C)) \cdot \mathsf{Opt} + \alpha(C) \cdot \mathsf{BestEq}$$

where $\alpha(C) \in (0,1)$ *and* $\lim_{C\to\infty} \alpha(C) = 0$.

(2) Furthermore, $\alpha(C) < \frac{g \cdot \mu(\tau+1) \cdot \mu(\tau)}{\gamma \cdot C}$ *so that for any* $C > 2 \cdot g \cdot \mu(\tau+1) \cdot \mu(\tau)$ *we have* $\alpha(C) < \frac{1}{2 \cdot \gamma}$.

Note that these are weak bounds on BestCE that are obtained by utilizing the *candidate* correlated equilibrium $\tilde{Q}$ defined by the following distribution over outcomes.

$$\tilde{Q}^o(\boldsymbol{x}) := \begin{cases} \frac{2 \cdot g \cdot \mu(k)}{C} & \text{if } \boldsymbol{x} = \boldsymbol{\xi}^k \text{ for each } k = \beta+1, ..., \tau \\ 1 - \sum_{k=\beta+1}^{\tau} \tilde{Q}^o(\boldsymbol{\xi}^k) & \text{if } \boldsymbol{x} = \boldsymbol{\xi}^\beta \\ 0 & \text{if otherwise.} \end{cases} \tag{2.5}$$

Then we show that whenever $C > 2 \cdot g \cdot \mu(\tau+1) \cdot \mu(\tau)$ this distribution satisfies all of the correlated equilibrium conditions. Finally, using the fact that $\tilde{Q}$ does strictly better than $(1-\alpha(C)) \cdot \mathsf{Opt} + \alpha(C) \cdot \mathsf{BestEq}$ implies that so does BestCE.

2.6 Appendix

2.6.1 Proofs of Section 4:

2.6.1.1 Proof of Lemma 2.7

Proof. (1) Without congestion, the minimal number of periods a retailer can order before the arrival time and still arrive on time is β. Therefore, given that at most γ retailers can exit the system at any given unit of time and that there are $t - \beta + 1$ times between t and β, $\mu(t) = \gamma \cdot (t - \beta + 1)$ is the maximum number of retailers who can exit the system when simultaneously ordering at time $-t$. Therefore, if more than $\mu(t)$ retailers order at $-t$, at least one of them must be late.

(2) We claim that by ordering τ periods early, Retailer i guarantees to arrive on time, for any strategy profile $\boldsymbol{\sigma}_{-i}$ of the remaining retailers. To see this, first note that by (1) and the uniform random priority, Retailer i ordering at time $-\tau$ is late only if the number of *other* retailers $j \neq i$ who have ordered at time $-t \leq -\tau$ is greater than or equal to $\mu(\tau)$. Then, simply noting that $\mu(\tau) = \gamma(\tau - \beta + 1) = \gamma\lceil \frac{I}{\gamma} \rceil \geq I$, and the number of *other* retailers is

$I-1$, we see that no matter what strategy $\boldsymbol{\sigma}_{-i}$ is played by the remaining retailers, there can never be more than $\mu(\tau)$ other retailers who order at $-t \leq -\tau$. Hence, if retailer $i \in I$ orders at $-\tau$, then she is guaranteed to arrive before the deadline.

□

2.6.1.2 Proof of Lemma 2.8

Proof. The inequality simply says that the minimum cost is no more than the one induced by the strategy profile $\boldsymbol{\sigma}^{opt}$ constructed above. It is easy to see that this profile is constructed as minimizing the total cost subject to no retailer being late.

For small values of C, it might well be optimal to have retailers receive their orders after the deadline. Now, start from the above profile and modify it so that at least one retailer is late. This retailer's cost increases by C, and the cost saved is no more than $g \cdot (\tau - \beta)$. If $C \geq g \cdot (\tau + 1) > g \cdot (\tau - \beta)$, then any added cost of being late will strictly outweigh the benefit of a lower inventory cost, therefore increasing the cost above $\boldsymbol{\sigma}^{opt}$. □

2.6.1.3 Proof of Lemma 2.9

Proof. Let $I > 2\gamma$. By Lemma 2.7, any action $t > \tau$ is strictly dominated by τ, so every retailer will play some action $t \leq \tau$. Assume, ad absurdum, that there exists a pure Nash equilibrium $\hat{\mathbf{s}}$. Then no retailer will pay more than $g \cdot \tau$. This implies that nobody arrives late with probability 1 whenever $C > g \cdot \tau$. Now consider the retailer i such that $\hat{s}_i = t_{max} = \max_j \hat{s}_j$. In the first case, suppose that player i is not late with probability 1 under $\hat{s}$. Now, consider the deviation $s_i' = t_{max} - 1$ for player i. If this deviation is not profitable, then it must be the case that player i is late with positive probability when making this deviation which implies $\mu(t_{max} - 1)$ players depart at time $-(t_{max} - 1)$ under $\hat{\mathbf{s}}$. But, if this is the case then no player ever departs later than time $t_{max} - 1$ (otherwise some player is late for sure under $\hat{\mathbf{s}}$). Yet, in this case any of the $\mu(t_{max} - 1)$ players departing at time $-(t_{max} - 1)$ has a profitable deviation to depart at time $-\beta$ whereby they face a lower inventory cost and receive their delivery with probability 1, a contradiction.

In the second case, player i is late with positive probability under $\hat{s}$. But, in this case any player ordering after time t_{max} will be late for sure. Hence, given the construction of t_{max}, it must be the case that $\hat{s}_j = t_{max}$ for all $j \in \mathcal{I}$. Further, given that player i is late with positive probability implies that $t_{max} < \tau$ which implies that at least γ players are late when they all depart at time t_{max}. Now, consider a deviation of some player j to $s_j' = t_{max}$.

If $\hat{\mathbf{s}}$ is a Nash equilibrium then it must be the case that this deviation is not profitable:

$$g \cdot (t_{max} + 1) \geq g \cdot t_{max} + \frac{\gamma}{I} \cdot C$$

but this implies that $C \leq g \cdot \frac{I}{\gamma}$ which is a contradiction as we have assumed $C > g(\tau + 1) = g(\lceil \frac{I}{\gamma} \rceil + \beta) \geq g\frac{I}{\gamma}$.

When $I \leq 2\gamma$, any profile where $\min\{\gamma, I\}$ retailers choose action β and the remaining retailers choose $\beta + 1$ is a pure Nash equilibrium. □

2.6.1.4 Proof of Theorem 2.10

Proof. First note that by, Lemma 2.7, ordering earlier than τ is a strictly dominated strategy. Further, any retailer that orders after time β is late with probability 1, therefore whenever $C \geq g \cdot (\tau + 1)$, any equilibrium $\boldsymbol{\sigma}^*$ must have a support $\text{supp}(\sigma_i^*) \subseteq \{\beta, \ldots, \tau - 1, \tau\}$.

In order to prove the claim, we will show that whenever $C > I \cdot g \cdot \tau$ then in any equilibrium (not necessarily symmetric) it must be the case that $I - \mu(\tau - 1)$ retailers mix over time τ, which under a symmetric strategy profile implies that all retailers mix over time τ and therefore pay $g \cdot \tau$. In order to prove this, we note that Lemma 2.7 implies that if less than $I - \mu(\tau - 1)$ mix over time τ under some strategy profile $\boldsymbol{\sigma}$, then it must be the case that

$$\mathbb{P}_{\boldsymbol{\sigma}}(max_{i \in I} a_i(\boldsymbol{\sigma}) > 0)) = 1$$

which is equivalent to saying the probability that *some* retailer is late under $\boldsymbol{\sigma}$ is equal to 1. This is due to the fact that at most $\mu(\tau - 1)$ retailers can order later than τ periods early without being late. Hence, in the best case $\max\{1, I - \mu(\tau) - 1\}$ retailers mix over time τ in which case at least one person is late.

Now whenever the probability that some retailer is late under $\boldsymbol{\sigma}$ is equal to 1, then it must be the case that there exists some retailer $i \in I$ such that the probability of them being late under $\boldsymbol{\sigma}$ is greater than or equal to $\frac{1}{I}$. Therefore, it must be the case that $c_i(\boldsymbol{\sigma}) \geq \frac{1}{I} C > g \cdot \tau$ (where the last inequality comes from the fact that $C > I \cdot g \cdot \tau$) and therefore $\boldsymbol{\sigma}$ cannot be a Nash equilibrium.

□

2.6.1.5 Proof of Theorem 2.11

Proof. In the proof of Theorem 2.10 we have shown that whenever $C > I \cdot g \cdot \tau$ then any equilibrium strategy profile σ must be such that at least $I - \mu(\tau - 1)$ retailers mix over time $-\tau$. Further, if some retailer i mixes over time $-\tau$, then it must be the case that at least $\mu(\tau - 1)$ *other* retailers mix over time $-(\tau - 1)$, otherwise the probability that Retailer i is late when ordering at time $-(\tau - 1)$ is equal to zero, in which case it is not optimal for them to ever order at time $-\tau$. Therefore, whenever $C > I \cdot g \cdot \tau$ the sum of the cost of the retailers mixing over time $-\tau$ and/or $-(\tau - 1)$ must be greater than

$$(I - \mu(\tau - 1)) \cdot g \cdot \tau + (\mu(\tau - 1) - (I - \mu(\tau - 1)) \cdot g \cdot (\tau - 1)$$

which is the social cost of a strategy profile where all $I - \mu(\tau - 1)$ that mix over time $-\tau$ also mix over time $-(\tau - 1)$. Finally, we note that $I - \mu(\tau - 1) \leq \gamma$ and therefore $\mu(\tau - 1) - (I - \mu(\tau - 1)) \geq \mu(\tau - 1) - \gamma = \mu(\tau - 2)$.

For the social cost, note that

$$(I-\mu(\tau-1))\cdot g\cdot\tau+(\mu(\tau-1)-(I-\mu(\tau-1))\cdot g\cdot(\tau-1)-\mathsf{SC}(\sigma^{opt}) > \mu(\tau-2)\cdot g\cdot(\tau-1)-\sum_{j=\beta}^{\tau-2}\gamma\cdot g\cdot j$$

$$= \gamma\sum_{j=\beta}^{\tau-2}(\tau-\beta+1-j) = \frac{\gamma}{2}(I-\beta-1)(I-3\beta+4)$$

□

2.6.1.6 Proof of Theorem 2.12

Proof. (1) Suppose that the wholesaler charges the premium $r(t) \geq 0$ for each $t = \beta+1, ..., \tau$. In this case, for the same reason as before, any Nash equilibrium such that some retailer is late with probability 1 is strictly dominated by any Nash equilibrium without tolls. Namely, if some retailer is late for sure under a Nash equilibrium with tolls, then the social cost is at least $C > I \cdot g \cdot \tau$ which is higher than the worst Nash equilibrium cost of $I \cdot g \cdot \tau$.

Now, if the probability that some retailer is late is strictly less than 1, then we know that $I - \mu(\tau - 1)$ retailers must randomize over time $-\tau$. In this case, it is optimal to set $r(\tau) = 0$. Further, in order to incentivize $I - \mu(\tau - 1)$ retailers to randomize over time $-\tau$ they should not have an incentive to deviate to time $-(\tau - 1)$. This is the case if either $r(\tau - 1) \geq g \cdot \tau - g(\tau - 1) = g$ or at least $\mu(\tau - 1)$ other retailers randomize over time $-(\tau - 1)$. In the latter case, the equilibrium obtains a social cost weakly greater than the best Nash equilibrium bound (strictly greater if any $r(t) > 0$). Therefore, we consider the

case where $r(\tau-1) \geq g$.

Now if $r(\tau-1) > g$, then no retailer will ever order at time $-(\tau-1)$ in equilibrium as ordering at time $-\tau$ strictly dominates ordering at time $-(\tau-1)$. This cannot be socially optimal as it requires more retailers to mix over at time $-\tau$ to prevent some retailer from being late. If $r(\tau-1) = g$ then players are indifferent between mixing over $-\tau$ and $-(\tau-1)$ so long as less than $\mu(\tau-1)$ players mix over $-(\tau-1)$. Now, if $r(\tau-2) = 0$, then in any equilibrium it must be the case that at least $\mu(\tau-2)$ players mix over time $-(\tau-2)$. Otherwise, $r(\tau-2) = 2 \cdot g$ and all players are indifferent between mixing over times $-\tau, ... -(\tau-2)$. Continuing in this fashion, we can see that for any time k either $r(k) = (\tau-k) \cdot g$, or $\mu(k)$ players mix over time $-k$. Hence, if any Nash equilibrium of $\Gamma(r)$ yields a better social cost than the best Nash equilibrium of Γ then it sets $r(j) = g \cdot (\tau-j)$ for all $j = \tau-k+\beta+1, ..., \tau$ and $r(j) < g \cdot (\tau-j)$ for all $j = \beta, ..., \tau-k+\beta$. In this case, we can see that the worst Nash equilibrium is one such that:

(1) $I-\mu(\tau-1)$ retailers order at time $-\tau$ with probability 1.

(2) γ retailers each order at each time $j = -(\tau-k+1), ..., -(\tau-1)$ with probability 1.

(3) The remaining $\mu(\tau-k)$ retailers play the worst Nash equilibrium of the game Γ with $\mu(\tau-k)$ players.

In this case,

$$\max_{\sigma \in \mathcal{E}(r)} \mathsf{SC}(\boldsymbol{\sigma}) = (I-\mu(\tau-k)) \cdot g \cdot \tau + \mu(\tau-k) \cdot r(\tau-k) = I \cdot g \cdot \tau - \mu(\tau-k) \cdot g \cdot k$$

and therefore the regulator chooses $\boldsymbol{z}^\star$ as the above toll scheme, choosing k to satisfy

$$k^\star = \underset{k \in \{\beta, ..., \tau-1\}}{\operatorname{argmin}} \; -\mu(\tau-k) \cdot g \cdot k$$

Note that this program has an interior solution so that $\beta < k^\star < \tau$ given that whenever $k = \tau$ we are in the case where there are no tolls and therefore the worst case Nash equilibrium cost is $I \cdot g \cdot \tau$ while whenever $k = \beta$ then by construction ordering at any time $-t$ yields a cost of $g \cdot t + r(t) = g \cdot \tau$ and therefore every Nash equilibrium yields a social cost of $I \cdot g \cdot \tau$.

The last step of the proof is to show that $r(t) = 0$ for all $t < k^\star$ as opposed to $r(t) < (\tau-t) \cdot g$. It is straightforward to realize though that if $r(t) = (\tau-t) \cdot g$ for all $t > k^\star$ and $0 < r(t) < (\tau-t) \cdot g$ for some $t \leq k^\star$ then any equilibrium is such that $\mu(\tau-k^\star)$ players mix over time $-(\tau-k^\star)$, paying at least $g \cdot (\tau-k^\star)$ and the remaining $I-\mu(\tau-k^\star)$ players departing at a time $-t < -k^\star$, paying a cost of exactly $g \cdot \tau$. Further, this argument holds for any schedule of premiums with $r(t) < (\tau-t) \cdot g$, $t \leq k^\star$. Therefore, given that additional premiums only weakly add to the cost of the retailers and have no effect on the equilibrium

outcome implies that it is optimal to set $r(t) = 0$ for all $t \leq k^\star$.

Now, the optimization problem above after substituting for μ becomes

$$\max_k \ -\gamma(\lceil \frac{I}{\gamma} \rceil - k) \cdot g \cdot k$$

and whenever $\frac{I}{\gamma}$ is integer, then the solution to this problem is $k^\star = \frac{I}{2\gamma}$. Further, substituting this value of k into the social cost above and substituting for τ and $\mu(\tau - k^\star)$ yields

$$\textsf{Premium} = I \cdot g \cdot \tau - \gamma \cdot \frac{I}{2\gamma} \cdot g \cdot \frac{I}{2\gamma} = I \cdot g \cdot \tau - \frac{1}{4} I \cdot g \cdot (\tau - \beta + 1) = \frac{3}{4} \cdot I \cdot g \cdot \tau + \frac{1}{4} \cdot I \cdot g \cdot (\beta - 1)$$

□

2.6.1.7 Proof of Lemma 2.13

Proof. If $\boldsymbol{x} \in Y$, then the capacity is never exceeded and no retailer is ever late.

Conversely, take $\boldsymbol{x} \in X$ and define

$$\theta := \min \left\{ t \in \{\beta, \dots, \tau\} : \sum_{j=\beta}^{t} x_j > \mu(t) \right\}.$$

Then the total number of orders between $-\theta$ and $-\beta$ is strictly larger than $\mu(\theta)$ which, by definition, is the maximal amount of retailers that can exit the system between time $-\theta$ and $-\beta$ without being late. Therefore, whenever $x \notin Y$ the probability that at least one retailer is late is equal to 1. □

2.6.1.8 Proof of Lemma 2.15

Proof of Lemma 2.15. Suppose that $Q \in \Delta(S^Y)$ is a correlated equilibrium and that Retailer i is recommended $s_i = k$. If the drawn strategy profile is $\boldsymbol{s} \in S^{ik} \setminus Z^{ik}$, then Retailer i is not late when deviating and ordering at time $-(k-1)$, so the cost from deviating is $g \cdot (k-1)$.

If $\boldsymbol{s} \in Z^{ik}$, then $\mu(k-1)$ retailers order at time $-(k-1)$. Therefore, if Retailer i deviates from s_i by ordering at time $-(k-1)$, then she is late with probability $(\mu(k-1)+1)^{-1}$. Therefore, the condition for Retailer i to optimally order at the suggested time $-k$ instead of time $-(k-1)$ is,

$$g \cdot k \leq g \cdot (k-1) + \frac{Q(\boldsymbol{s} \in Z^{ik} | s_i = k)}{\mu(k-1)+1} \cdot C,$$

or equivalently

$$Q(\boldsymbol{s} \in Z^{ik} | s_i = k) \geq \frac{g \cdot (\mu(k-1)+1)}{C}. \tag{2.6}$$

Then, since $Z^{ik} \subset S^{ik}$,

$$Q(\boldsymbol{s} \in Z^{ik} | s_i = k) = \frac{\sum_{\boldsymbol{s} \in Z^{ik}} Q(\boldsymbol{s})}{\sum_{\boldsymbol{s} \in S^{ik}} Q(\boldsymbol{s})}$$

and rearranging we obtain our result.

Conversely, we will show that if $Q \in \Delta(S^Y)$ satisfies the above conditions, i.e., it is not profitable for Retailer i to deviate from k to $k-1$, then they also do not want to deviate to $j \in \{\beta, \ldots, \tau\}$. Note first that for any $\boldsymbol{x} \in Y$, there is no profitable deviation for a retailer who is recommended β, since ordering at $-\beta$ without being late yields the lowest cost possible.

If Retailer i is told to order at time $-k$ under $Q \in \Delta(S^Y)$, then she pays exactly $g \cdot k$ by doing so, and therefore will not want to deviate to any $j > k$. Now, we note that if Retailer i orders at time $-j > -(k-1)$, when suggested to order at time $-k$, then she is late with probability at least $Q(\boldsymbol{s} \in Z^{ik} | s_i = k)$. This is because if $\boldsymbol{s} \in Z^{ik}$ and $s_i = k$, there are exactly $\mu(k-1)$ retailers, other than Retailer i, who order at time $-(k-1)$ under $\boldsymbol{s}$. Therefore, if Retailer i orders at time $-j > -(k-1)$, then she is late for sure, as these $\mu(k-1)$ retailers all have priority over Retailer i in this case. Therefore, by deviating to time $-j > -(k-1)$ instead of ordering at time $-k$, Retailer i obtains a cost of at least

$$g \cdot j + Q(\boldsymbol{s} \in Z^{ik} | s_i = k) \cdot C \geq g \cdot j + g \cdot (\mu(k-1)+1),$$

since by condition 2.6, $Q(\boldsymbol{s} \in Z^{ik} | s_i = k) \cdot C \geq g \cdot (\mu(k-1)+1)$. Hence, the deviation to $-j$ is not profitable whenever

$$g \cdot k \geq g \cdot j + g \cdot (\mu(k-1)+1)$$

or $k - j \leq \mu(k-1)+1 = \gamma(k-\beta+1)+1$ and left hand side of this expression is maximized when $j = \beta$ and therefore given that $\gamma \geq 1$ implies that this condition is always satisfied for any k.

□

2.6.1.9 Proof of Lemma 2.16

Proof. Take $k > \beta$ and $\boldsymbol{x} \in X^k$. For any pure strategy $\boldsymbol{s} \in S^{ik} \cap S(\boldsymbol{x})$, if Retailer i orders at time $-(k-1)$ instead of $-k$, then she is late with positive probability when the remaining

retailers play $\boldsymbol{s}_{-i}$. This implies that there are at least $\mu(k-1)$ other retailers in the system at time $-(k-1)$ under x. Then, either $x_{k-1} = \mu(k-1)$, which is our desired conclusion, or at time $-k$ there are at least $\mu(k-1) - x_{k-1} + \gamma$ retailers in the system, excluding Retailer i. Hence, if $\boldsymbol{s} \in S^{ik} \cap S(\boldsymbol{x})$, then Retailer i is recommended k, which means that there are $\mu(k-1) - x_{k-1} + \gamma + 1$ retailers in the system at time k. Therefore, there are $\mu(k-1)+1$ retailers in the system at time $k-1$, a contradiction to the fact that $\boldsymbol{x} \in Y$. Finally, the fact that Z^{ik} are the set of pure strategies that enforce order time k for retailer i is a straightforward corollary. □

2.6.1.10 Proof of Lemma 2.17

Proof. Summing the inequalities of Proposition 2.15 over $i \in I$, we obtain:

$$\sum_{i=1}^{I} \sum_{\boldsymbol{s} \in Z^{ik}} Q(\boldsymbol{s}) \geq \frac{g \cdot (\mu(k-1)+1)}{C} \left[\sum_{i=1}^{I} \sum_{\boldsymbol{s} \in S^{ik}} Q(\boldsymbol{s}) \right]. \tag{2.7}$$

Then, noting that

$$\sum_{i=1}^{I} \sum_{\boldsymbol{s} \in Z^{ik}} Q(\boldsymbol{s}) = \sum_{\boldsymbol{x} \in X^k} \sum_{i=1}^{I} \sum_{\boldsymbol{s} \in Z^{ik} \cap S(\boldsymbol{x})} Q(\boldsymbol{s}) = \sum_{\boldsymbol{x} \in X^k} x_k Q^o(\boldsymbol{x})$$

and similarly,

$$\sum_{i=1}^{I} \sum_{\boldsymbol{s} \in S^{ik}} Q(\boldsymbol{s}) = \sum_{\boldsymbol{x} \in Y} \sum_{i=1}^{I} \sum_{\boldsymbol{s} \in S^{ik} \cap S(\boldsymbol{x})} Q(\boldsymbol{s}) = \sum_{\boldsymbol{x} \in Y} x_k Q^o(\boldsymbol{x})$$

we obtain our result. □

2.6.1.11 Proof of Theorem 2.18

Proof. Before proving this theorem we will give more structure to the correlated distributions we consider.

Uniform symmetric distributions: Here we will illustrate that since the game is fully symmetric, only the distributions of outcomes matter (as opposed to distributions over strategies).

Definition 2.22. $Q \in \Delta(S^Y)$ is a uniform symmetric distribution (USD) if for all $\boldsymbol{x} \in X$ and $\boldsymbol{s}, \boldsymbol{s}' \in S(\boldsymbol{x})$, we have $Q(\boldsymbol{s}) = Q(\boldsymbol{s}')$. Denote by $\Delta^U(S^Y)$ the set of all uniform symmetric distributions over outcomes in Y.

Namely, Q is a uniform symmetric distribution if all pure strategies that induce the same outcome have the same probability under Q.

Lemma 2.23. *If a uniform symmetric distribution Q satisfies (2.4) for all $k = \beta+1, \ldots, \tau$, then Q is a correlated equilibrium.*

Proof. Suppose that Q is a USD and satisfies the conditions of (2.4) for all $k = \beta+1, \ldots, \tau$. Then, as in the proof of Lemma 2.17, summing the inequalities of (2.4) over $i \in I$ we obtain

$$\sum_{i=1}^{I} \sum_{\boldsymbol{s} \in Z^{ik}} Q(\boldsymbol{s}) \geq \frac{g \cdot (\mu(k-1)+1)}{C} \left[\sum_{i=1}^{I} \sum_{\boldsymbol{s} \in S^{ik}} Q(\boldsymbol{s}) \right]. \tag{2.8}$$

Now we simply note that if Q is a USD, then

$$Q(\boldsymbol{s}) = \frac{1}{|S(\boldsymbol{x}^{\boldsymbol{s}})|} Q^o(\boldsymbol{x}^{\boldsymbol{s}}) \text{ for all } \boldsymbol{s} \in S.$$

Further, note that for all $i, j \in I$ and for all $\boldsymbol{s} \in Z^{ik}$, we have $\pi_{ij}(\boldsymbol{s}) \in Z_j^k$ where $\pi_{ij} : S \to S$ is the permutation that exchanges the strategy of Retailer i and Retailer j. Therefore, if Q is a USD,

$$\sum_{\boldsymbol{s} \in Z^{ik}} Q(\boldsymbol{s}) = \sum_{\boldsymbol{s} \in Z^{jk}} Q(\boldsymbol{s}) \text{ and } \sum_{\boldsymbol{s} \in S^{ik}} Q(\boldsymbol{s}) = \sum_{\boldsymbol{s} \in S^{jk}} Q(\boldsymbol{s})$$

for all $i, j \in I$. Using this fact, Equation (2.8) implies that for every $i \in I$ and $k = \beta+1, \ldots, \tau$

$$I \cdot \sum_{\boldsymbol{s} \in Z^{ik}} Q(\boldsymbol{s}) \geq \frac{g \cdot (\mu(k-1)+1)}{C} \left[I \cdot \sum_{\boldsymbol{s} \in S^{ik}} Q(\boldsymbol{s}) \right].$$

Dividing by I gives the result. □

Lemma 2.24. *Let $Q \in \Delta(S^Y)$ be a correlated equilibrium of the game Γ. Then, there exists a correlated equilibrium $\widehat{Q}$ such that $\widehat{Q}$ is a USD and $\mathsf{SC}(\widehat{Q}) = \mathsf{SC}(Q)$.*

Proof. Let Q be a correlated equilibrium. Define $\widehat{Q}$ by $\widehat{Q}(\boldsymbol{s}) = \frac{1}{|S(\boldsymbol{x}^{\boldsymbol{s}})|} Q^o(\boldsymbol{x}^{\boldsymbol{s}})$ for all $\boldsymbol{s} \in S^Y$. By construction, $\widehat{Q}$ is a USD. Further, $\widehat{Q}$ induces the same distribution over outcomes as Q:

$$\widehat{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{s} \in S(\boldsymbol{x})} \widehat{Q}(\boldsymbol{s}) = Q^o(\boldsymbol{x}) \sum_{\boldsymbol{s} \in S(\boldsymbol{x})} \frac{1}{|S(\boldsymbol{x})|} = Q^o(\boldsymbol{x}).$$

Therefore, it must be the case that $\widehat{Q}$ is a correlated equilibrium whenever Q is a correlated equilibrium and $\mathsf{SC}(Q) = \mathsf{SC}(\widehat{Q})$. □

Lemma 2.24 states that when searching for a cost minimizing correlated equilibrium we can restrict our attention to USD distributions. We will now prove some further results regarding enforcing outcomes and correlated equilibrium.

Lemma 2.25. *The set of outcomes X^k and X^j are disjoint for all $k = \beta+1, ..., \tau$ and $j = \beta+1, ..., \tau$ with $j \neq k$.*

Proof. With out loss assume that $k > j$. Then, $\boldsymbol{x} \in X^k$ implies that $x_k \geq 1$ and $x_{k-1} = \mu(k-1)$ and $\hat{\boldsymbol{x}} \in X^j$ implies that $\hat{x}_j \geq 1$ and $\hat{x}_{j-1} = \mu(j-1)$. But, if $x_{k-1} = \mu(k-1)$ then any retailer ordering $t < k-1$ periods early is late for sure. But $k > j$ implies that $k-1 \geq j$ and therefore given that $\hat{x}_{j-1} > 0$ we see that whenever $\boldsymbol{x} \in X^k$ and $\boldsymbol{x} \in X^j$ then $\boldsymbol{x} \notin Y$, a contradiction. □

Lemma 2.26. *Let $Q \in \Delta^U(S^Y)$, then for all $k = \beta, ..., \tau$ there exists $\boldsymbol{x} \in \text{supp}(Q^o)$ such that $x_k > 0$.*

Proof. If $Q \in \Delta^U(S^Y)$ is a correlated equilibrium, then for all $\boldsymbol{x} \in \text{supp}(Q^o)$, we know $\boldsymbol{x} \in Y$. If $\boldsymbol{x} \in Y$ then it must be the case that $x_\tau \geq 1$ as, by the definition of τ, if all retailers order less than τ periods early then at least one of them must be late. Therefore, the order τ must be enforced by some outcome $\boldsymbol{x} \in \text{supp}(Q^o)$. Now, if $\boldsymbol{x} \in \text{supp}(Q^o)$ enforces τ then this implies that $x_{\tau-1} = \mu(\tau-1)$ and therefore there must exists $\boldsymbol{x} \in \text{supp}(Q^o)$ that enforces $\tau - 1$. Continuing in this fashion we can see that for each $k = \beta, ..., \tau$ there exists $\boldsymbol{x} \in \text{supp}(Q^o)$ such that $x_k > 0$. □

We are now ready to prove the claim of Theorem 2.18.

Proof. Take any correlated equilibrium $Q \in \Delta^U(S^Y)$ such that $Q^o(\widetilde{\boldsymbol{x}}) > 0$ for some $\widetilde{\boldsymbol{x}} \notin \cup_{k=\beta+1}^{\tau} X^k$. We will now construct from Q a correlated equilibrium $\widetilde{Q}$ such that $\mathsf{SC}(\widetilde{Q}) < \mathsf{SC}(Q)$ and $\widetilde{Q}^o(\widetilde{\boldsymbol{x}}) = 0$.

Now suppose $\widetilde{\boldsymbol{x}} \notin X^{\beta+1}$. This implies that $\widetilde{x}_\beta + \widetilde{x}_{\beta+1} < \mu(\beta+1) = 2\gamma$. Let us denote by $j_1, ..., j_\alpha$ the smallest consecutive integers $\beta+1 < j_1 < j_2 < \cdots < j_\alpha$ at which the latest $2\gamma - \widetilde{x}_\beta - \widetilde{x}_{\beta+1}$ retailers order after time $\beta+1$ under $\widetilde{\boldsymbol{x}}$. Namely, $j_1, ..., j_\alpha$ are the times such that

$$\sum_{k=j_1}^{j_\alpha} \widetilde{x}_k \geq 2\gamma - \widetilde{x}_\beta - \widetilde{x}_{\beta+1}, \quad \sum_{k=j_1}^{j_\alpha - 1} \widetilde{x}_k < 2\gamma_0 - \widetilde{x}_\beta - \widetilde{x}_{\beta+1}, \quad \text{and} \quad \sum_{k=\beta+2}^{j_1-1} \widetilde{x}_k = 0$$

where the last summation holds whenever $j_1 \geq \beta+3$. Now, construct a new outcome $\hat{\boldsymbol{x}}$ from $\widetilde{\boldsymbol{x}}$ such that

$$\hat{x}_t = \begin{cases} \gamma_0 & \text{if} \quad t = \beta, \beta+1 \\ 0 & \text{if} \quad t = j_1, ..., j_{\alpha-1} \\ 2\gamma_0 - \widetilde{x}_\beta - \widetilde{x}_{\beta+1} - \sum_{l=j_1}^{j_{\alpha-1}} \widetilde{x}_l & \text{if} \quad t = j_\alpha \\ \widetilde{x}_t & \text{otherwise} \end{cases}$$

Namely, we construct $\hat{\boldsymbol{x}}$ by shifting $2\gamma_0 - \widetilde{x}_\beta - \widetilde{x}_{\beta+1}$ retailers that order under $\widetilde{\boldsymbol{x}}$ strictly before $-(\beta+1)$ to times $-\beta$ and $-(\beta+1)$ so that there are full capacity orders at these times under $\hat{\boldsymbol{x}}$. It can be easily checked by this construction that $\hat{\boldsymbol{x}} \in X^{\beta+1}$ and $\mathsf{SC}^o(\hat{\boldsymbol{x}}) < \mathsf{SC}^o(\widetilde{\boldsymbol{x}})$.

Now, given that any distribution $Q \in \Delta^U(S^Y)$ is completely determined by its corresponding distribution over outcomes Q^o by the USD property, we will now construct the distribution $\widetilde{Q} \in \Delta^U(S^Y)$ from Q via their corresponding distributions over outcomes as follows:

$$\widetilde{Q}^o(\boldsymbol{x}) = \begin{cases} 0 & \text{if} \quad \boldsymbol{x} = \widetilde{\boldsymbol{x}} \\ Q^o(\hat{\boldsymbol{x}}) + Q^o(\widetilde{\boldsymbol{x}}) & \text{if} \quad \boldsymbol{x} = \hat{\boldsymbol{x}} \\ Q^o(\boldsymbol{x}) & \text{otherwise.} \end{cases}$$

Namely, we simply shift the probability that Q^o puts on $\widetilde{\boldsymbol{x}}$ onto the probability that Q^o puts on $\hat{\boldsymbol{x}}$. The resulting distribution $\widetilde{Q}$ clearly has a higher social welfare than the distribution Q given that we are only shifting the probability on $\widetilde{\boldsymbol{x}}$ to the probability on $\hat{\boldsymbol{x}}$ and $\mathsf{SC}^o(\hat{\boldsymbol{x}}) < \mathsf{SC}^o(\widetilde{\boldsymbol{x}})$. The only thing we need to check is that $\widetilde{Q}$ is a correlated equilibrium.

To do so, first note that given $\widetilde{\boldsymbol{x}} \notin \cup_{k=\beta+1}^{\tau} X^k$ implies that

$$\sum_{\boldsymbol{x} \in X^k} x_k \widetilde{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in X^k} x_k Q^o(\boldsymbol{x})$$

for all $k = \beta+2, ..., \tau$. Therefore the left hand side of Equation (2.4) remains unchanged for all $k = \beta+2, ..., \tau$. Further, for all $k \notin \{\beta+1, j_1, ..., j_\alpha\}$ we know that the probability of being told to order at time k also remains unchanged so that conditions for $\widetilde{Q}$ to be a correlated equilibrium for each $k \notin \{\beta+1, j_1, ..., j_\alpha\}$ are the same as the conditions for Q and therefore are satisfied.

What we need to check is that the conditions of Equation (2.4) hold for each $k \in \{\beta+1, j_1, ..., j_\alpha\}$. First, we know that for each $j \in \{j_1, ..., j_\alpha\}$ the right hand side of Equation (2.4) strictly decreases by a factor of $Q^o(\widetilde{\boldsymbol{x}})$. This comes from the fact that by shifting the mass from $Q^o(\widetilde{\boldsymbol{x}})$ to $Q^o(\hat{\boldsymbol{x}})$ we decrease the probability that Retailer i is told to order j periods early. Given that we make no other changes implies that (2.4) must hold for all $k = j_1, ..., j_\alpha$. Finally, we only need to check that (2.4) holds for $k = \beta+1$. To do this, we

first note that given $\hat{x}_{\beta+1} = \gamma$, then by construction,

$$\sum_{\boldsymbol{x} \in X^{\beta+1}} x_{\beta+1} \widetilde{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in X^{\beta+1}} x_{\beta+1} Q^o(\boldsymbol{x}) + (\gamma - \widetilde{x}_{\beta+1}) Q^o(\widetilde{\boldsymbol{x}}).$$

Similarly,

$$\sum_{\boldsymbol{x} \in Y} x_{\beta+1} \widetilde{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in Y} x_{\beta+1} Q^o(\boldsymbol{x}) + (\gamma - \widetilde{x}_{\beta+1}) Q^o(\widetilde{\boldsymbol{x}}).$$

Therefore, $\widetilde{Q}$ is a correlated equilibrium whenever

$$\sum_{\boldsymbol{x} \in X^{\beta+1}} x_{\beta+1} Q^o(\boldsymbol{x}) + (\gamma - \widetilde{x}_{\beta+1}) Q^o(\widetilde{\boldsymbol{x}}) \geq \frac{g \cdot (\mu(\beta)+1)}{C} [\sum_{\boldsymbol{x} \in Y} x_{\beta+1} Q^o(\boldsymbol{x}) + (\gamma - \widetilde{x}_{\beta+1}) Q^o(\widetilde{\boldsymbol{x}})].$$

Now, given that Q is a correlated equilibrium then we know that a sufficient condition for the above inequality to hold is whenever $\frac{g \cdot (\mu(\beta)+1)}{C} \leq 1$. This is the case whenever $C \geq g \cdot (\mu(\beta)+1)$. Therefore, whenever $C \geq \delta(\beta+1)(\mu(\beta)+1)$ we have constructed a correlated equilibrium $\widetilde{Q}$ such that $\mathsf{SC}(\widetilde{Q}) < \mathsf{SC}(Q)$ and such that $\widetilde{Q}^o(\widetilde{\boldsymbol{x}}) = 0$. Finally, given that we have chosen $\widetilde{\boldsymbol{x}}$ arbitrarily, we can always iterate this process to show that for all $\widetilde{\boldsymbol{x}} \notin \cup_{k=\beta+1}^{l} X^k$ whenever $C \geq \delta(l)(\mu(l-1)+1)$, a cost minimizing correlated equilibrium never mixes over $\widetilde{\boldsymbol{x}}$ and taking $l = \tau$ we have proven our claim. □

□

2.6.1.12 Proof of Theorem 2.20

Proof. We will show that for any correlated equilibrium $Q \in \Delta^U(S^Y)$ such that (2.4) doesn't hold with equality for some $k \in \beta+2, ..., \tau$ under Q, then we can construct a new correlated equilibrium $\widetilde{Q}$ from Q such that $\mathsf{SC}(\widetilde{Q}) < \mathsf{SC}(Q)$ and inequality k of (2.4) holds with equality under $\widetilde{Q}$.

Let us first suppose that $Q \in \Delta^U(S^Y)$ is a correlated equilibrium such that there exists a $k > \beta+2$ (we will treat the $k = \beta+2$ case at the end) such that inequality k of (2.4) is strict under Q. Next, take any $\widetilde{\boldsymbol{x}} \in X^k$ such that $Q^o(\widetilde{\boldsymbol{x}}) > 0$. Now, construct the outcome $\hat{\boldsymbol{x}}$ as follows

$$\hat{x}_t = \begin{cases} \gamma & \text{if} \quad t = \beta, \beta+1 \\ \mu(k-1) - 2\gamma & \text{if} \quad t = k-1 \\ \widetilde{x}_t & \text{otherwise.} \end{cases}$$

Namely, we construct $\hat{\boldsymbol{x}}$ from $\widetilde{\boldsymbol{x}}$ by moving γ orders from $k-1$ periods early to β and $\beta+1$ periods early respectively. It is easy to check by this construction that $\hat{\boldsymbol{x}} \in X^{\beta+1}$. Now, let

us construct the distribution $\widetilde{Q}_{\epsilon(C)}$ as follows:

$$\widetilde{Q}^o_{\epsilon(C)}(\boldsymbol{x}) = \begin{cases} Q^o(\widetilde{\boldsymbol{x}}) - \epsilon(C) & \text{if} \quad \boldsymbol{x} = \widetilde{\boldsymbol{x}} \\ Q^o(\hat{\boldsymbol{x}}) + \epsilon(C) & \text{if} \quad \boldsymbol{x} = \hat{\boldsymbol{x}} \\ Q^o(\boldsymbol{x}) & \text{otherwise.} \end{cases}$$

Then, $\widetilde{Q}_{\epsilon(C)}$ clearly has a strictly lower cost for all $0 < \epsilon(C) \leq Q(\widetilde{\boldsymbol{x}})$ as $\mathsf{SC}^o(\hat{\boldsymbol{x}}) < \mathsf{SC}^o(\widetilde{\boldsymbol{x}})$. What we claim is that there exists $\epsilon(C) > 0$ such that $\widetilde{Q}_{\epsilon(C)}$ is a correlated equilibrium whenever $C > g \cdot (\mu(k-1)+1)$. In order to prove this claim, first note that, using the same logic from the proof of Lemma 2.18, the only equilibrium constraints affected by shifting $\epsilon(C)$ probability from $Q^o(\widetilde{\boldsymbol{x}})$ to $Q^o(\hat{\boldsymbol{x}})$ are the constraints $k, k-1$, and $\beta+1$. Then, we note that the left hand side of constraint $(k-1)$ for $\widetilde{Q}$ remains the same and the right hand side decreases by a factor of $\epsilon(C)$. This again comes from the fact that in moving from Q to $\widetilde{Q}$ we do not change the probability of any outcomes in X^{k-1} but we do decrease the probability that someone is told to order $k-1$ periods early. Therefore, we know that constraint $(k-1)$ for $\widetilde{Q}$ is still satisfied.

Now, we will show that Condition $(\beta+2)$ for $\widetilde{Q}$ is satisfied whenever $C \geq g \cdot (\mu(\beta)+2)$. To do this, we note that

$$\sum_{\boldsymbol{x} \in X^{\beta+2}} x_{\beta+2} \widetilde{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in X^{\beta+2}} x_{\beta+2} Q^o(\boldsymbol{x}) + \hat{x}_{\beta+2} \epsilon(C)$$

further,

$$\sum_{\boldsymbol{x} \in Y} x_{\beta+2} \widetilde{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in Y} x_{\beta+2} Q^o(\boldsymbol{x}) + \hat{x}_{\beta+2} \epsilon(C)$$

and therefore constraint $(\beta+2)$ for $\widetilde{Q}$ is satisfied if and only if

$$\sum_{\boldsymbol{x} \in X^{\beta+2}} x_{\beta+2} Q^o(\boldsymbol{x}) + \hat{x}_{\beta+2} \epsilon(C) \geq \frac{g \cdot (\mu(\beta+1)+1)}{C} [\sum_{\boldsymbol{x} \in Y} x_{\beta+2} Q^o(\boldsymbol{x}) + \hat{x}_{\beta+2} \epsilon(C)]$$

which is the case whenever $C \geq g \cdot (\mu(\beta+1)+1)$.

Now, in order to show that constraint k for $\widetilde{Q}$ can be satisfied with equality for some $\epsilon(C) > 0$ note first that

$$\sum_{\boldsymbol{x} \in X^k} x_k \widetilde{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in X^k} x_k Q^o(\boldsymbol{x}) - \widetilde{x}_k \epsilon(C) \;\text{ and }\; \sum_{\boldsymbol{x} \in Y} x_k \widetilde{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in Y} x_k Q^o(\boldsymbol{x}) - \widetilde{x}_k \epsilon(C)$$

therefore, constraint k for $\widetilde{Q}$ is satisfied if and only if

$$\sum_{\boldsymbol{x}\in X^k} Q^o(\boldsymbol{x}) - \widetilde{x}_k\epsilon(C) \geq \frac{g\cdot(\mu(k-1)+1)}{C}[\sum_{\boldsymbol{x}\in Y} x_k Q^o(\boldsymbol{x}) - \cdot\widetilde{x}_k\epsilon(C)]$$

but given that constraint k for Q is satisfied with strict inequality implies that whenever $C > g\cdot\mu(k-1)+1$ we can always find $\epsilon(C)$ such that constraint k for $\widetilde{Q}$ is *less slack* than constraint k for Q. Now, taking $\hat{\epsilon}(C)$ to be the number that makes the above inequality hold with equality, we simply note that if $\hat{\epsilon}(C) \leq Q^o(\widetilde{\boldsymbol{x}})$, then we can choose $\widetilde{Q}$ such that $\epsilon(C) = \hat{\epsilon}(C)$ and we will have constructed a CE such that Condition (k) holds with equality. Otherwise, we should iterate the process by taking another $\boldsymbol{x} \in X^k$ such that $Q^o(\boldsymbol{x}) > 0$ and applying the same transformation until Condition (k) holds with equality. Given that in each step we obtain a new correlated equilibrium with lower social cost, we have proven our claim.

The last thing to check is that we can perform the same operation for $k = \beta + 2$. Here we construct $\hat{\boldsymbol{x}}$ from any outcome $\widetilde{\boldsymbol{x}} \in X^{\beta+2}$ simply by having γ retailers who order $\beta+1$ periods early under $\widetilde{\boldsymbol{x}}$ order β periods early instead. The difference in this case is that the only constraints that are affected are $(\beta+1)$ and $(\beta+2)$. The constraint $(\beta+2)$ changes in the same way as before, so it can easily be checked that constraint $(\beta+2)$ is satisfied for $\widetilde{Q}$. Now, let us look at the effects on constraint $(\beta+1)$. There are two effects, first when moving $\epsilon(C)$ probability from $Q^o(\widetilde{\boldsymbol{x}})$ to $Q^o(\hat{\boldsymbol{x}})$ we decrease the probability of being told to order $\beta+1$ periods early. Second, we increase the probability of $\hat{\boldsymbol{x}} \in X^{\beta+1}$. These effects are represented by the fact that:

$$\sum_{\boldsymbol{x}\in X^{\beta+1}} x_{\beta+1}\widetilde{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{x}\in X^{\beta+1}} x_{\beta+1}\widetilde{Q}^o(\boldsymbol{x}) + \hat{x}_{\beta+1}\epsilon(C)$$

further,

$$\sum_{\boldsymbol{x}\in Y} x_{\beta+1}\widetilde{Q}^o(\boldsymbol{x}) = \sum_{\boldsymbol{x}\in Y} x_{\beta+1}Q^o(\boldsymbol{x}) - (\widetilde{x}_{\beta+1} - \hat{x}_{\beta+1})\epsilon(C)$$

therefore noting that $\widetilde{x}_{\beta+1} - \hat{x}_{\beta+1} = \gamma$ by construction, we can see that the left hand side of constraint $(\beta+1)$ increases for $\widetilde{Q}$ and the right hand side decreases. Thus, when $k = \beta+2$, all constraints are still satisfied.

□

2.6.1.13 Proof of Theorem 2.21

Proof. In order to prove this result we will utilize a *candidate* correlated equilibrium solution

$\tilde{Q}$ defined in 2.5 which we will show satisfies the conditions of Lemma 2.17 and obtains a social cost $\mathsf{SC}(\tilde{Q}) < (1-\alpha(C)) \cdot \mathsf{Opt} + \alpha(C) \cdot \mathsf{BestEq}$ therefore generating an upper bound on BestCE as by definition $\mathsf{BestCE} \leq \mathsf{SC}(\tilde{Q})$ for all correlated equilibria $\tilde{Q}$.

First, we will show that $\tilde{Q}^o$ satisfies the condition of Lemma 2.17 for each $k = \beta+1, ..., \tau$. First, note that this condition can be rewritten as

$$\gamma \cdot \tilde{Q}^o(\boldsymbol{\xi}^k) \geq \frac{g \cdot (\mu(k-1)+1)}{C} \cdot (\mu(k-1)\tilde{Q}^o(\boldsymbol{\xi}^{k+1}) + \gamma \sum_{j=\beta}^{k+1} \tilde{Q}^o(\boldsymbol{\xi}^j))$$

which can further be rearranged as

$$\gamma \cdot \tilde{Q}^o(\boldsymbol{\xi}^k) \geq \frac{g \cdot (\mu(k-1)+1)}{C} \cdot (\mu(k-1)\tilde{Q}^o(\boldsymbol{\xi}^{k+1}) + \gamma(1 - \sum_{j=\beta+1}^{k-1} \tilde{Q}^o(\boldsymbol{\xi}^j))$$

Therefore, a sufficient condition for $\tilde{Q}^o$ to be a correlated equilibrium is for

$$\gamma \cdot \tilde{Q}^o(\boldsymbol{\xi}^k) \geq \frac{g \cdot (\mu(k-1)+1)}{C} \cdot (\mu(k-1)\tilde{Q}^o(\boldsymbol{\xi}^{k+1}) + \gamma)$$

and after substituting the values for $\tilde{Q}^o(\boldsymbol{\xi}^k)$ and $\tilde{Q}^o(\boldsymbol{\xi}^{k+1})$ and rearranging we obtain

$$2\gamma \geq \frac{\mu(k-1)+1}{\mu(k)}(\frac{2g\mu(k+1)\mu(k-1)}{C} + \gamma)$$

Then, using the fact that $\mu(k) = \mu(k-1) + \gamma \geq \mu(k-1) + 1$, we note that this inequality is satisfied whenever $C \geq \frac{2 \cdot g \cdot \mu(k+1)\mu(k-1)}{\gamma}$ for each $k = \beta+1, ..., \tau$. Hence whenever $C \geq 2 \cdot g \cdot \mu(\tau+1)\mu(\tau)$ then $\tilde{Q}$ is a correlated equilibrium.

In order to prove the relationship between BestCE and Opt we note that the most costly outcome in the support of $\tilde{Q}^o$ is $\boldsymbol{\xi}^\tau$. Further $\mathsf{SC}^o(\boldsymbol{\xi}^\tau) = \mathsf{BestEq}$ and $\mathsf{SC}^o(\boldsymbol{\xi}^\beta) = \mathsf{Opt}$, therefore

$$\mathsf{SC}(\tilde{Q}) = (1 - \sum_{j=\beta+1}^{\tau} \tilde{Q}^o(\boldsymbol{\xi}^j)) \cdot \mathsf{Opt} + \sum_{j=\beta+1}^{\tau} \tilde{Q}^o(\boldsymbol{\xi}^j)\, \mathsf{SC}^o(\boldsymbol{\xi}^j) < (1 - \sum_{j=\beta+1}^{\tau} \tilde{Q}^o(\boldsymbol{\xi}^j)) \cdot \mathsf{Opt} + \sum_{j=\beta+1}^{\tau} \tilde{Q}^o(\boldsymbol{\xi}^j)\, \mathsf{BestCE}$$

Therefore, setting $\alpha(C) = \sum_{j=\beta+1}^{\tau} \tilde{Q}^o(\boldsymbol{\xi}^j) = \frac{2 \cdot g \cdot}{C} \sum_{j=\beta+1}^{\tau} \gamma(j-\beta+1) = \frac{g}{C}\gamma(\tau-\beta+1)(\tau-\beta+2) = \frac{g \cdot \mu(\tau) \cdot \mu(\tau+1)}{C \cdot \gamma}$. □

References

[1] Arnott, R., De Palma, A., and Lindsey R. (1990): "Economics of a bottleneck. *Journal of Urban Economics*, 27(1), 111-130.

[2] Aumann, R. J. (1974): "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics*, 1, 67-96.

[3] Beckmann, M. J. and Srinivasan, S. K. (1987): "An (s,S) inventory system with Poisson demands and exponential lead time." *OR Spectrum*, 9, 213-217.

[4] Buchanan, D. J., Love, R. F. (1985): "A (Q,R) Inventory Model with Lost Sales and Erlang Distributed Lead Times." *Naval Research Logistics Quarterly*, 32, 605-611.

[5] Chen, F. Federgruen, A. and Zheng, Yu-Sheng (2001): "Coordination Mechanisms for a Distribution System with One Supplier and Multiple Retailers." *Management Science*, 47 (5).

[6] Hendrickson, C. and Kocur, G. (1981): "Schedule delay and order time decisions in a deterministic model." *Transportation Science*, 15(1), 62-77.

[7] Hill, R. M. (1992): "Numerical analysis of a continuous-review lost-sales inventory model where two orders may be outstanding. *European Journal of Operations Research*, 62, 11-26.

[8] Hill, R.M. (1994): "Continuous review lost sales inventory models where two orders may be outstanding. *International Journal of Production Economics*.

[9] Johansen, S. G. and Thorstenson, A. (1993): Optimal and approximate (Q,r) inventory policies with lost sales and gamma-distributed lead time. *International Journal of Production Economics*, 30-31, 179-194.

[10] Karlin, S. and Scarf, H. (1958): "Inventory models of the Arrow-Harris-Marschak type with time lag. In: Arrow, K., Karlin, S., Scarf, H. (Eds.), Studies in the Mathematical Theory of Inventory and Production. Stanford University, Ch 10.

[11] Li, X. (2017): "Optimal procurement strategies from suppliers with random yield and all-or-nothing risks." *Ann Oper Res*, 257, 167-181.

[12] Mukherjee, A. S. (2008): "The Spider's Strategy: Creating Networks to Avert Crisis, Create Change, and Really Get Ahead." *FT Press*.

[13] Ravichandran, N. (1984): “Note on (s,S) Inventory Policy." *IIE Transactions*, 16, 387-390.

[14] van Donselaar, K„ Gaur, V., van Woensel, T., Broekmeulen, R. A. C. M., Fransoo, J. C. (2010): “Ordering Behavior in Retail Stores and Implications for Automated Replenishment. *Management Science*, 56(6), 766-784.

[15] Vickrey, W. S. (1969): “Congestion theory and transport investment," *American Economic Review*, 59(2), 251-260.

[16] Ziegelmeyer, A., Koessler F., Boun My, K., and Denant-Boémont, L. (2008): “Road Traffic Congestion and Public Information," *Journal of Transport Economics and Policy*, 42(1), 43-82.

3 Bank Regulation, Investment, and the Implementation of Capital Requirements

Abstract

We study the optimal design of bank capital regulations in a model where banks face adverse selection when raising capital. We show how the implementation of capital requirements is an important regulatory tool as it can help mitigate bank underinvestment by eliminating the information frictions that make raising capital costly. Specifically, the regulator can design incentive compatible requirements that induce the banks to reveal their private information to the market through their choice of capital structure. Using this insight we characterize the optimal implementation of capital requirements which induces information revelation when the banking sector is weak and pools the banks' private information otherwise.[43]

3.1 Introduction

Since the financial crisis, policy makers have worked to enhance the regulatory framework in order to prevent future crises and their associated spillover effects. While one of the most significant changes in post crisis regulations came in the form of increased bank capital requirements (via Basel III), some critics still argue that these requirements should be increased to even higher levels (see e.g. Admati and Hellwig (2013)). This begs the question of what keeps regulators from increasing capital requirements further and what is the theoretical foundation that drives such decisions?[44]

Higher capital requirements serve as a way to prevent bank failures by increasing the bank's ability to absorb unexpected losses before failing. Given that bank failures have shown to impose large negative externalities on society, this creates a well accepted rationale for higher capital requirements.[45] On the other hand, the social cost of higher capital requirements in the literature is typically taken as a black box, motivated by adverse selection and its link to underinvestment (see Myers and Majluf (1984)). In this paper we study the optimal design of capital requirements, explicitly incorporating this adverse selection prob-

[43]Special thanks to Jean Edouard Colliard, Olivier Gossner, Denis Gromb, Raphaël Levy, Eric Mengus, Adolfo de Motta, Thomas Noe, Ludovic Renou, Marco Scarsini, Tristan Tomala, and Nicolas Vieille for valuable feedback and comments.

[44]Chapter 3 of Dewatripont, Rochet, and Tirole (2010) highlights that most of the motivations for the Basel I and II accords come from political pressure on policy makers by the banking industry, first to create a regulatory framework that avoided competitive distortions, then to allow the banks to use their superior information to decide the risk weighting of assets.

[45]The loss in output due to the financial crisis is estimated to be over $75 trillion for Basel committee member countries (Basel Committee (2015)).

lem into our model, and show how utilizing this foundation generates important insights for the design of bank capital regulations. In particular, we show how the implementation of capital requirements becomes a crucial element of the regulatory design, optimally varying as a function of the strength of the banking sector, which generates a natural link between micro and macro-prudential bank regulation.

We study a model whereby banks have private information about the value of their existing assets and impose a negative externality on society when they fail. In such a setting, raising capital is costly for banks whose assets are undervalued by the market (i.e. when they have good news) which will lead them to forgo new projects when subject to high capital requirements. The regulator therefore sets capital requirements to balance a tradeoff between minimizing the social cost of bank failure and stimulating bank investment in socially valuable projects. The key insight that we develop is that the implementation of capital requirements can mitigate bank underinvestment by eliminating the bank's *information cost* of raising capital. Namely, we show how the regulator can design requirements that induce the bank to reveal its private information to the market through its choice of capital, leading the market to correctly price its shares. The need for such mechanisms was illustrated during the financial crisis and is supported by evidence showing that banks with more opaque balance sheets were more likely to recapitalize through government programs like the Troubled Asset Relief Program (TARP) in the US (see Black et. al. (2016)).[46] Interestingly, such information revelation is not always optimal and pooling the banks private information can be preferred instead.

We show the existence of three optimal regulatory regimes as a function of the strength of the banking sector and the net present value (NPV) of new investments. Under the first regime (IRB-type), the regulator resolves the underinvestment problem by designing capital requirements that induce the banks to credibly reveal their private information to the market. This type of regime is similar to the Internal Ratings Based approach (IRB) introduced in Basel II whereby banks utilize their own internal risk models to provide the regulator with key statistics of their asset returns (e.g. probability of default, loss given default, etc.) that determine the bank's capital requirement. Although it is not clear whether the IRB approach was designed to act as a way for banks to credibly signal their private information to the market[47], we show that this is precisely the merit of allowing banks to utilize their

[46] Admittance to the TARP program was conditional on meeting certain solvency requirements produced by a regulatory audit which many believe acted as a signal to markets about the quality of the participating banks' capital.

[47] Samuels et. al. (2012) survey bank investors and find that a majority lack confidence in the banks' risk weighted asset reports and believe that they should not be permitted to utilize their own internal models for the calculation of capital requirements.

own information to influence their capital requirements. In this sense, our results highlight a neglected benefit of the IRB approach whereby slightly augmenting the approach with a report specific (ex-ante) transfer can lead to a large welfare improvement.[48] That being said, the regulator must pay information rents to the banks (in the form of lower capital requirements) in order to induce them to reveal their private information under this regime which is why it is not always optimal over the underlying parameter space.

The second optimal regime (SA-type) is one whereby the regulator sets a simple *pooling* capital requirement, independent of the bank's private information. Such a regime is similar to the Standardized Approach (SA) of Basel I-III whereby the bank's capital requirements are grouped by asset type and credit rating, but independent of any additional information the bank may posses about those assets. This is precisely the mechanism under which banks with good news will optimally forgo investments when capital requirements are set too high. Hence, under the SA-type regime, capital requirements are set as high as possible subject to inducing investment by the banks with good news.

Finally, it may be the case that the cost to society of lowering capital requirements — either to induce information revelation in the IRB-type regime or to induce investment in the SA-type regime — does not outweigh the benefit of the investments that these regulations induce. In this case, the regulator utilizes a third underinvestment (UI) regime that sets high capital requirements, inducing an equilibrium whereby the banks with good news forgo all investments while the banks with bad news are recapitalized.

The main result of the paper is a characterization of the optimal regulatory mechanism which formalizes the optimal capital requirements, transfers (e.g. deposit insurance premia), and securities utilized under the optimal SA-type, IRB-type, and UI regimes and conditions under which each respective regime is optimal. The key parameter that determines the optimal regime is the proportion of banks with good news. Figure (21) illustrates the social cost of bank capital under each regime, taken as the difference between first best welfare and the welfare obtained when restricting to the optimal mechanism of each regime. As can be seen, when the proportion of good banks is high (greater than p_2), then the optimal mechanism mitigates underinvestment through the SA-type regime. This comes from the fact that, in this case, the cost of raising capital for the banks with good news is small as the securities they issue (e.g. equity) are only slightly undervalued by the market. Hence, the regulator can set high capital requirements and still induce investment. If instead, the proportion of banks with good news is low (below p_1) the optimal mechanism sets high

[48] An example of a mechanism that incentivizes credible information revelation is one whereby banks with good (bad) news face lower (higher) capital requirements on their new investments but a higher (lower) deposit insurance premium.

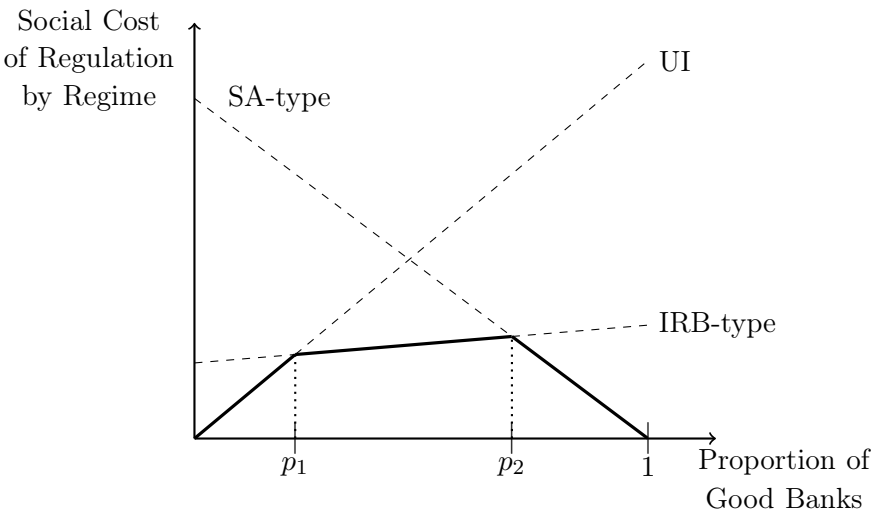


Figure 21: The Social Cost of Bank Capital Under the Optimal Regulatory Design

capital requirements under the UI regime, inducing the banks with good news to forgo the new investment in exchange for recapitalizing the large portion of bad banks. This is optimal as the cost to society of underinvestment by the good types diminishes when their proportion goes to zero.[49] Finally, if the proportion of good banks is intermediate (between p_1 and p_2) then inducing the banks to reveal their private information through the IRB-type regime is optimal. This is because the cost of inducing investment through the SA-type regime is too large as the good type's security is heavily undervalued by the market, yet the proportion of good banks is too high for underinvestment to be socially desirable through the UI regime.[50]

The results of this paper allows us to characterize how capital requirements should be adjusted with respect to the strength of the banking sector, highlighting the macro-prudential insights developed when the cost of capital is properly micro-founded. We also illustrate how the level of minimum capital requirements (as opposed to the implementation) should optimally vary with the strength of the banking sector, the economy, and the opacity of the bank's assets, allowing us to lend support to new policy measures linked to macroeconomic fundamentals such as the counter cyclical capital buffer of Basel III. Important to note is that if the regulator utilizes a static mechanism that does not adjust capital requirements with respect to these variables, then this will lead to suboptimal underinvestment or a higher (expected) social cost of bank failure. This is an important insight to be gained, especially

[49] When a bank receives bad news this implies that its assets are overvalued by the market and therefore it receives a subsidy when raising new capital. For this reason, banks with bad news will never forgo new investments.

[50] As will be seen, the conditions to induce information revelation through the IRB-type are independent of the proportion of good banks. The only (minor) variation in the IRB-type capital requirements comes from the change in weights the regulator puts on each type when calculating expected welfare.

in the context of the current regulation which, for the most part, sets capital requirements that do not adjust with the underlying fundamentals (e.g. NPV, risk, and opacity of new investments).

Another novel feature of this paper is that, to our knowledge, it is one of the first to incorporate security design into the problem of bank capital regulation. Namely, we allow the regulator not only to set capital requirements and transfers but also to restrict the set of securities that the bank can issue to meet the capital requirements (e.g. equity, subordinated debt, etc) in a very general sense. Our results show that the SA-type regime optimally restricts banks to issue securities that are the least *informationally sensitive*: securities that minimize the difference in the value of the security with respect to the bank's private information. On the other hand, the IRB-type regime optimally restricts the banks with good news to issue the least informationally sensitive security while banks with bad news are required to issue the most informationally sensitive security (equivalent to selling existing assets). This lends support for the use of contingent convertible (CoCo) bonds for the financing of regulatory capital as CoCo bonds have the ability to minimize the information sensitivity of the security, similar to debt securities, but also have the desirable property of absorbing losses before the bank fails, similar to equity. This is an important insight given the recent surge of European banks using CoCo bonds to meet *additional tier 1* capital requirements.[51]

Our results contribute to a number of policy debates on the current design of prudential bank regulations. We discuss in Section 5 when the regulator should regulate the banks under either the IRB-type or SA-type regime based on the opacity of the bank's assets. Further, we note that the current discretion that banks have to choose whether they are regulated by the SA or IRB approach under Basel III should be removed as we show that such discretion will lead banks to choose the suboptimal framework when it is allowed. We then discuss other policy implications such as how our model provides insight into the new counter cyclical capital requirement (CCyB) of Basel III and stress testing/regulatory information disclosure.

Related Literature

This paper is empirically motivated by the observation that banks decrease lending in response to regulatory capital requirements (see e.g. Peek and Rosengren (1995), Gropp, et. al (2016), Fraisse, et. al. (2017)). Further, we find empirical justification that this decrease in lending comes from the *private information cost* of raising capital as evidenced by bank equity issuance during the financial crisis. In particular, many of the largest (and most

[51] https://www.bloomberg.com/quicktake/contingent-convertible-bonds.

opaque) U.S. banks were reluctant to raise capital during the crisis, leading to government injections of equity through programs like the Troubled Asset Relief Program (TARP). Yet, over $450 billion worth of bank equity was voluntarily issued over the same time period without any government assistance (Black, et. al. (2016)). This can be explained by the fact that the bank's cost of raising equity varies with its private information, consistent with the finding in Black, et. al. (2016) that banks with more opaque assets (measured by lower turnover, higher volatility, and higher bid-ask spreads) were more likely to issue equity using government programs as opposed to issuing to private investors over this period.

In this paper we study how capital requirements can lead to underinvestment when securities are issued to a less informed market, an idea inspired by Myers and Majluf (1984). Stein (98) studies a similar adverse selection problem but instead asks how a decrease in reserves can affect bank lending in order to develop monetary policy insights. Our general security design problem and capital raising game is similar to that studied in Nachman and Noe (1994) and Noe (1988). Nachman and Noe (1994) characterize conditions on the distribution of returns under which firms prefer to finance their assets with debt as opposed to equity. In contrast to these papers, our aim is not to characterize what security maximizes the value of the firm to existing shareholders, but rather to characterize the optimal securities for the use of prudential regulation.

In our model, high capital requirements lead to credit rationing but this is not the only reason for credit rationing due to asymmetric information. Stiglitz and Weiss (1981) develop a model where banks ration credit due to the adverse selection problem that exists between the bank and its privately informed loan applicants. Thakor (1996) shows that higher capital requirements can exacerbate this credit rationing problem. Yet, in Thakor (1996) the credit rationing effect of higher capital requirements relies on the assumption that higher capital requirements lead to a higher cost of financing, justified by the Myers and Majluf (1984) insight. What we show in this paper is that the regulator has the potential to eliminate this cost of capital financing by designing capital regulations that resolve the information asymmetry between the bank and the market.

As mentioned above, the foundation of the cost of capital develops a natural link between the micro-prudential and macro-prudential objectives of the regulator. While many papers have studied issues related to macro-prudential regulation and capital requirements (see e.g. Hanson, et. al. (2011) and Repullo (2013)) the typical arguments assume that in *bad* times it is too costly to raise capital so that banks must sell their assets at fire sale prices in order to meet capital requirements. In this paper we micro-found the bank's cost of raising capital and show how the regulators mechanism can affect this cost through its ability to induce information revelation.

From a mechanism design perspective, the closest related paper is Giammarino, et al. (1993). They consider the problem of combined moral hazard and adverse selection and study the optimal design of incentive compatible capital requirements and deposit insurance premia. In their model, they assume that equity is dilutive and bears an exogenous cost driven by the investor's "preference for liquidity". While Giammarino, et. al. (1993) study incentive compatible mechanisms, as in this paper, they see no need for information revelation due to the fact that the cost of equity is driven exogenously and therefore cannot be influenced.

Morrison and White (2005) study a model of capital regulations whereby banks are formed by managers of differing skill and subject to moral hazard. One of their main results is that when the regulator has a low ability to screen managers (in order to prevent issuing banking licenses to low skill managers), then they may prefer to set high capital requirements; shrinking the banking sector to improve the quality of the remaining banks. This result bears similarity to our UI-regime which optimally tightens capital requirements when the banking sector is very weak, effectively shrinking the banking sector in exchange for recapitalizing the remaining banks (i.e. improving quality). Although these results are obtained for different reasons, they shed light on a non-standard implication whereby tightening capital requirements can be optimal, even if it reduces the size of the banking sector.

As mentioned above, the IRB-type mechanism that we propose is similar to the IRB approach introduced in Basel II. It is important to note that we do not claim that the current IRB approach of Basel III is optimal as in practice insurance premiums/taxes are not linked to IRB reports. From a theoretical perspective, strategic underreporting of bank risk via IRB has been studied in papers such as Prescott (2004), Leitner and Yilmaz (2019), and Colliard (2017).[52] In particular, Colliard (2017) shows that when the bank's internal risk estimates are private information, costly auditing leads to less risk-sensitive capital requirements in order to counteract the bank's incentive to choose risk models that underreport their true risk. Blum (2008) studies incentive compatibility issues with the IRB approach and finds that if the regulator has limited scope to sanction banks when they detect misreporting of risk ex-post, then a leverage ratio can improve welfare. The contribution of this paper to this literature is to show how, when properly designed, the IRB approach can serve to resolve information asymmetries between the bank and the market and the associated financing costs that they create.

Our results also complement the literature on the optimal disclosure of financial information. Bouvard et. al. (2011) show how, when banks are exposed to roll over risk, disclosing

[52]There is further empirical evidence that IRB is not incentive compatible along some dimensions (e.g. Plosser and Santos (2018)).

bank specific information improves financial stability in times of crisis but can have a destabilizing effect in normal times. Interestingly, we develop similar results with regards to the optimality of information disclosure in good and bad times, but for largely different reasons. Leitner and Williams (2017) show how the regulator faces a trade off between keeping its stress testing model secret to prevent gaming and revealing the model to prevent suboptimal underinvestment (the key cost of capital in our model). Goldstein and Leitner (2018) study the optimal information disclosure policy of the regulator's stress test. They show that in some cases disclosure can eliminate risk sharing opportunities for the bank but that in other cases it is necessary to facilitate such opportunities. This paper compliments this literature by studying information disclosure through the design of capital requirements. Namely, stress testing may not be necessary when the optimal capital regulations take the form of the IRB-type regime which reveals the bank's private information to the market. In contrast, in Section 5 we discuss how stress testing can complement the results of this paper when the level of opacity of the banks' existing assets is large.

Finally, our IRB-type mechanism bears some similarity to that of optimal interventions as studied by Philippon and Skreta (2012) and Tirole (2012). Both of these papers consider optimal interventions to restore lending and investment in the face of adverse selection. Philippon and Schnabl (2013) analyze the issue of recapitalizing a banking sector that restricts lending due to a debt overhang problem. In contrast, our motivation for such an *intervention* is to provide incentives for banks to voluntarily recapitalize when faced with unexpected losses and we show how this can be done without the use of government funds.

The rest of the paper is organized as follows. Section 1 presents the main model, including the mechanisms available to the regulator, the capital raising game between the bank and the market, and our equilibrium concept and refinements. Section 2 characterizes the equilibria of the capital raising game given the regulator's choice of mechanism. Section 3 characterizes the optimal SA-type (pooling) and IRB-type (separating) mechanisms. Section 4 presents our main result which characterizes when the SA-type, IRB-type, or UI regime is optimal given the proportion of banks with good news. Section 5 presents the policy implications of our results and Section 6 concludes. Section 7 is devoted to extending the main results beyond the two type case to a continuum of types. All proofs are relegated to the appendix in Section 8.

3.2 A Model of Capital Regulation Under Asymmetric Information

3.2.1 Baseline Model

The basic set up of the model is similar to Myers and Majluf (1984). The bank starts at time $t = 0$ with assets in place that generate a gross return captured by the random variable A. We assume for simplicity that A is a binary random variable whose return at time $t = 1$ is equal to a_h with probability p and a_ℓ with probability $(1 - p)$ where $a_h > a_\ell \geq 0$. The assets in place were purchased by the bank at time $t = -1$ and financed with 100% equity. We assume that at time $t = 0$ the bank receives private information regarding the time $t = 1$ return of its assets in place. In particular, we assume without loss that the bank learns its type $\theta \in \Theta = \{h, \ell\}$ and that a type θ bank knows that its time $t = 1$ return will be a_θ.[53]

After learning its type at time $t = 0$ the bank, whose manager acts in the interest of the incumbent shareholders, receives an investment opportunity that costs I and generates a net return $B \sim G$ with expected value $\hat{b} := \mathbb{E}[B] > 0$. We assume that the distribution G has a bounded support over $\mathbb{R}$, has a density g that is continuous over its support, and that g is weakly increasing for returns less than the mean and weakly decreasing for returns greater than the mean. All asset returns are generated at time $t = 1$ in which case the bank is liquidated and the funds distributed to the bank's creditors and shareholders.

3.2.2 Capital Securities

We endow the regulator with the right to set *capital requirements* which dictate that some amount of the new investment $K = \gamma \cdot I$ must be financed through the sale of a security that the regulator qualifies as a *capital security*. We assume that the fraction of the investment not financed by the sale of some capital security is financed with insured deposits which we assume are issued at the risk free rate (normalized to zero). Therefore, under the *laissez-faire* regulations (i.e. $K = 0$) all banks invest in the new project and finance themselves with 100% deposits.

We will now present our conditions for admissible capital securities. First note that a security is a mapping from the bank's return (net deposits) z to a payment $s(z)$ to the owner of the security. In what follows we will restrict attention to capital securities satisfying the following standard assumptions.

Definition 3.1. A capital security s is admissible if it satisfies the following conditions.
(1) $s(z)$ is non-decreasing in the value of the bank z.
(2) $z - s(z)$ is non-decreasing in the value of the bank z.

[53] In the extensions section we show how our results can be extended to the case where Θ is a continuum.

(3) $s(z) \geq 0$ for all $z \in \mathbb{R}$.
We denote by $\mathcal{S}$ the set of admissible capital securities.

The conditions of Definition 3.1 are commonly used assumptions when studying the design of securities (see e.g. Innes (1990) and Nachman and Noe (1994)) that prevent risk free arbitrage (conditions (1) and (2)) and account for the security holder's limited liability (condition (3)).[54] In what follows we restrict attention to general securities in $\mathcal{S}$.

The purpose of *capital* is to absorb bank losses but can be defined differently given the regulator's objective. Namely, if the bank is large and systemic then the bank's insolvency can have spillover effects on the real economy (e.g. the failure of Lehman Brothers). In this case a capital security should be defined as a security with the ability to absorb losses *before* the bank becomes insolvent (e.g. equity). If instead the bank is small and financed with deposits then the regulator may only care about protecting the deposit insurance fund, in which case, securities that absorb losses post insolvency may also qualify as capital (e.g. bail-inable/subordinated debt). In light of this discussion, we proceed throughout by assuming that equity always qualifies as capital (this will be useful to prove some of our results) but that other securities may also qualify. The only important aspect of capital securities that we model is that they are admissible and junior to deposits.

3.2.3 The Regulatory Environment

The regulator's capital requirement $K \geq 0$ dictates that the bank must raise an amount of funds (used to finance the new investment) greater than or equal to K by selling an admissible capital security $s \in \mathcal{S}$. Given that our distribution G is bounded there exists a level of capital $\bar{K}$ such that $K > \bar{K}$ provides no benefit to society.[55] Therefore, the *first best* outcome (obtained in the case of perfect information) would be one whereby the regulator imposes a capital requirement $K = \bar{K}$ and both bank types invest in the new project. We will see below how high capital requirements lead to underinvestment by the h-type banks (in the case of incomplete information), precluding this first best outcome. We further endow the regulator with the ability to impose a lump sum ex-ante tax T on the bank (e.g. a

[54] If Condition (1) is not satisfied so that $s(z) < s(z')$ for some $z > z'$, then the bank could engage in a risk free arbitrage opportunity whereby whenever its return is z', it borrows $z - z'$ and reports z as it's earnings, gaining a profit of $s(z) - s(z')$. Similarly, if Condition (2) is not satisfied, then the bank could engage in a similar arbitrage by burning money (e.g. by liquidating assets below their market value). Condition (3) represents the limited liability of the investors purchasing the security.

[55] Typically we would assume $K \leq I$ so that the banks are never required to raise more capital than the cost of their investment, but given that bad news in this model represents a devaluation of a bank's assets in place, then it also reflects a decrease in the bank's effective equity stock. Therefore, $K > I$ represents the case whereby the regulator requires the bank to recapitalize its pre-investment balance sheet before being allowed to invest in the new asset.

deposit insurance premium) and to restrict the set of securities (to a subset of $\mathcal{S}$) that the bank can use to finance the capital requirement (e.g. to equity). We assume that the bank has the right to forgo the new investment (and receive a payoff of a_θ) whenever it finds it unprofitable to meet the requirements of the regulator's mechanism.

Naturally, the requirements of the regulator can also depend on the bank's type θ so that when the bank *reports* that its type is θ then it must generate an amount of funding K_θ through the sale of a capital security in the restricted set $\mathcal{S}_\theta \subset \mathcal{S}$, and to pay a transfer T_θ. We assume throughout that the report of the bank's type is observed by the regulator but not by the market. Instead, we assume that the market observes the bank's commitment to meet the requirement K_θ and pay the transfer T_θ to ensure that the revelation principle holds.[56] Note that while the mechanism can signal the bank's type through its reported commitment (K_θ, T_θ), the bank's potential freedom to issue different securities $s \in \mathcal{S}_\theta$ (which can generate different levels of capital) may also act as an alternative signaling device in the capital raising game described below in Section 1.5. In summary, we restrict attention throughout to the following class of mechanisms.

Definition 3.2. The regulator's mechanism $\mathcal{M}$ consists of a menu $\{(K_\theta, T_\theta, \mathcal{S}_\theta)\}_{\theta \in \{h,\ell\}}$ such that the option $\theta \in \{h, \ell\}$ requires the bank to generate funds worth at least $K_\theta \in \mathbb{R}_+$ through the sale of a capital security $s \in \mathcal{S}_\theta \subset \mathcal{S}$ and to pay an ex-ante transfer $T_\theta \in \mathbb{R}_+$ to the regulator.

Note that our class of mechanisms could be potentially extended to the case whereby the regulator reports a noisy signal of the bank's type to the market. We do not model this signaling problem so that the only signaling of the bank's type through the mechanism comes from the (potential) difference in capital requirements and transfers. While generating a noisy signal regarding the bank's type may improve upon our class of mechanisms we note that it requires significant commitment power by the regulator.[57]

Another restriction of our mechanism is that we specify transfers as lump-sum and to be paid *ex-ante* in the spirit of a deposit insurance premium. In this case, the *ex-ante* transfer,

[56] This only matters when $K_h = K_\ell$ and $T_h = T_\ell$. This information can be publicly reported by the bank or regulator.

[57] Namely, if the signal the regulator sends to the market is noisy, then it must be the case that it is randomly chosen, along with different capital requirements associated with the realized signal. A simple analogy is that the regulator has to flip a coin that when lands on heads yields a high capital requirement and tails a low capital requirement regardless of bank type (although the coin for different bank types has different probabilities of heads). The issue is that the regulator then has to report truthfully to the bank and the market whether the coin has landed on heads or tails and to enforce the associated capital requirements. Given that the regulator will always prefer higher capital requirements (conditional on all bank types investing), not only does the regulator have to have significant commitment power not to always report that the coin landed on heads, but the market has to believe that the regulator will not renege on its commitment as well.

T, will effect the pricing of a given security issued by the bank as it decreases the ex-post value of the bank from z to $z - T$. We make this restriction as an ex-ante transfer can be financed through the sale of the capital security so that there are no issues with the bank's ability to pay given its limited liability nor the regulators commitment to enforce payments in bad states of the world.[58]

Given a particular mechanism $\mathcal{M}$, whenever a bank of type θ chooses the menu option $\tilde{\theta} \in \{h, \ell\}$ and issues some security $s \in \mathcal{S}_{\tilde{\theta}}$ that generates funds $P \geq K_{\tilde{\theta}} + T_{\tilde{\theta}}$ (i.e. it satisfies the requirements of the mechanism) then the bank's ex-ante expected payoff is given by

$$V_\theta(s, \tilde{\theta}; P) := \mathbb{E}[max\{a_\theta + B + P - T_{\tilde{\theta}} - s, 0\}]$$

Namely, $V_\theta(s, \tilde{\theta}; P)$ represents the post investment payoff of the type θ bank who chooses menu option $\tilde{\theta}$, net deposits $I - P$, the transfer $T_{\tilde{\theta}}$, and the security payment.[59] To clarify this expression, note that the gross return of the bank's assets after making the investment is $a_\theta + I + x$ where x is the realization of B. Further, the bank raised P through the sale of s, therefore after paying the ex-ante transfer it finances the investment with $P - T_{\tilde{\theta}}$ of *new equity* and $D = I - (P - T_{\tilde{\theta}})$ of deposits. Therefore, the bank's return (accounting for limited liability) net deposits is $\max\{a_\theta + x + P - T_{\tilde{\theta}}, 0\}$. Finally, the bank must repay the security holders according to s which we can include in the max because $s(z) = 0$ whenever $z = a_\theta + x + P - T_{\tilde{\theta}} \leq 0$. Thus we obtain our expression for $V_\theta(s, \tilde{\theta}; P)$. Note that in equilibrium the amount of funds generated, P, will be determined endogenously via the market beliefs of the bank's type given the menu option it chooses and the security it issues. In Section 2.1 we show how the payoff $V_\theta(s, \tilde{\theta}; P)$ can be decomposed into the sum of the value of the bank's existing assets, the bank's *intrinsic value* of the new investment, and the cost/subsidy the bank pays/receives due to the mis-pricing of its security.

In what follows we will differentiate between *pooling* and *separating* mechanisms which we now define.

Definition 3.3. A pooling mechanism $\mathcal{M}$ is any mechanism satisfying $K_h = K_\ell$, $T_h = T_\ell$, and $\mathcal{S}_h = \mathcal{S}_\ell$.
A separating mechanism $\mathcal{M}$ is any mechanism satisfying either $K_h \neq K_\ell$ or $T_h \neq T_\ell$.

Note that when the transfers T_θ are too large, then no bank type will ever find it profitable

[58] A more general set up would also allow for ex-post transfers dependent on the bank's realized value. We refrain from studying ex-post transfers as no such transfers currently exist in practice and the general insight can be obtained with a simpler ex-ante transfer that is inherently robust to the timing structure of the capital raising game described below and observability of returns.

[59] Given that B is the net return, the gross return is therefore $I + B$. Hence, the return on the new investment net deposits is $I + B - (I - P) = B + P$.

to invest. Therefore, we proceed by assuming without loss that T_θ is bounded above by the level of transfers that induce banks to forgo the investment. As we will see, denoting by $b_\theta(K_\theta)$ the *intrinsic value* of the new investment to the existing shareholders of a type θ bank after raising new capital worth K_θ (formally defined in Lemma 3.9 below), this implies $T_\theta \leq b_\theta(K_\theta)$ under any separating mechanism (otherwise the type θ bank will forgo the investment) and that $T \leq \min\{b_\ell(K), b_h(K)\}$ under any pooling mechanism with transfer T and capital requirement K. The latter half of this assumption will not play a role in the analysis as we will show that transfers under the optimal pooling mechanism are always set to zero.

Given the revelation principle it is without loss to restrict attention to incentive compatible mechanisms $\mathcal{M}$ such that it is optimal for the type-θ bank to *report truthfully* (i.e. choose the menu option θ). Further, whenever $\mathcal{M}$ is pooling then $\mathcal{M}$ is trivially incentive compatible given that the bank's choice of menu does not signal any information to the market. If instead $\mathcal{M}$ is a separating mechanism, then incentive compatibility is given by the following definition.

Definition 3.4. Let $\mathcal{M}$ be a separating mechanism. Then, $\mathcal{M}$ is incentive compatible if for each $\theta \in \{h, \ell\}$ there exists $s \in \mathcal{S}_\theta$ such that $\mathbb{E}_\theta[s] \geq K_\theta + T_\theta$ and

$$V_\theta(s, \theta; \mathbb{E}_\theta[s]) \geq V_\theta(\tilde{s}, \tilde{\theta}; \mathbb{E}_{\tilde{\theta}}[\tilde{s}])$$

for all $\tilde{\theta} \in \{h, \ell\}$ and $\tilde{s} \in \mathcal{S}_{\tilde{\theta}}$ such that $\mathbb{E}_{\tilde{\theta}}[\tilde{s}] \geq K_{\tilde{\theta}} + T_{\tilde{\theta}}$.

Namely, $\mathcal{M}$ is incentive compatible if whenever the market belief coincides with the bank's menu choice (i.e. whenever the bank chooses menu option θ then the market believes its type is θ), then the type θ bank prefers to issue some security $s \in \mathcal{S}_\theta$ to meet the capital requirement K_θ and pay the transfer T_θ rather than issue any other security $\tilde{s} \in \mathcal{S}_{\tilde{\theta}}$ to meet the capital requirement $K_{\tilde{\theta}}$ and pay the transfer $T_{\tilde{\theta}}$. Note that this definition of incentive compatibility assumes that the market beliefs will be correct. We show below that under a standard equilibrium refinement this will always be the case whenever $\mathcal{M}$ is a separating mechanism satisfying the conditions of Definition 3.4.

3.2.4 Welfare

We define welfare as the sum of payoffs to the bank and its creditors net the spillover costs of bank failure. Namely, given the bank's type, θ, and a level of capital, $K_1 = P - T_\theta$, generated from the sale of some capital security s, the bank fails when its losses from the new investment exceed its effective capital stock $a_\theta + K_1$. This is the case whenever the

realization x, of B, is less than $-a_\theta - K_1$. The expected loss to the bank's creditors is therefore given by

$$L_\theta(K_1) := -\mathbb{E}[min\{a_\theta + B + K_1, 0\}].$$

$L_\theta(K_1)$ is naturally independent of the type of capital security offered and is only a function of the capital K_1 that it generates. This is due to the fact that that $s(z) = 0$ whenever $a_\theta + x + K_1 \leq 0$.

Bankruptcy creates a deadweight loss to society which we assume for simplicity is proportional to the expected loss $L_\theta(K_1)$ and captured by the parameter λ.[60] Therefore, the social welfare under the mechanism $\mathcal{M} = \{K_\theta, T_\theta, \mathcal{S}_\theta\}_{\theta\in\{h,\ell\}}$ when the type θ bank invests, reports type $\tilde{\theta}$, and the funds generated by the sale of its capital security are $P \geq K_{\tilde{\theta}} + T_{\tilde{\theta}}$ (i.e. the capital generated is $K_1 = P - T_{\tilde{\theta}}$) is given by

$$W_\theta(invest|K_1) = a_\theta + b_\theta(K_1) - (1+\lambda)\cdot L_\theta(K_1) = a_\theta + \hat{b} - \lambda \cdot L_\theta(K_1)$$

Namely, the bank's profit is $V_\theta(s, \tilde{\theta}; P)$ but the buyers of the bank's capital security pay P and receive $\mathbb{E}_\theta[s]$ while the bank receives P and loses $\mathbb{E}_\theta[s]$. Further, the bank pays the regulator $T_{\tilde{\theta}}$ from the funds P generated and the regulator receives the transfer $T_{\tilde{\theta}}$.[61] Hence, after canceling out these terms from the bank's profit we obtain the first equality of the expression. The second equality comes from the fact that $b_\theta(K_1) = \hat{b} + L_\theta(K_1)$: the bank's intrinsic value of the new investment is equal to the NPV of the investment plus the value of the deposit insurance to the bank (this is formally proven in Lemma 3.9 below).

If instead the bank forgoes the investment, then the social welfare is

$$W_\theta(forgo) = a_\theta.$$

Note that the welfare only depends on the decision to invest or not, regardless of the security issued. This is due to the fact that while the security may be under/over priced with respect to the bank's private information, this discrepancy acts as a direct transfer of wealth from the bank's incumbent shareholders to the owners of the security. Hence, given that the regulator

[60] A more general functional form could be utilized as long as the deadweight loss to society is strictly decreasing in bank capital, all else being equal. One interpretation is that λ represents the deadweight loss to the bank's creditors caused by bankruptcy/liquidation proceedings. Similarly, we could also interpret λ as the deadweight loss incurred from imposing distortionary taxes on society in order to generate the funds to repay the insured deposits or the bank's creditors if the regulator cannot commit to not bailout the bank in times of distress. Finally, we can also interpret λ as the spillover effects on the real economy caused by the failure of the bank, for example due to systemic factors.

[61] Here we assume that transfers from the bank to the regulator are treated as taxes which are then redistributed to society via government expenditures. We do not assume that these transfers fund the deposit insurance fund for simplicity but the model could be easily extended in this direction.

does not weight the bank's shareholders any differently from external investors this transfer cancels out in the welfare function.

As we will see below, the relevant expected welfare (given that the ℓ-type will always invest) when the type θ raises $K_\theta = P_\theta - T_\theta$ from the sale of some capital security, is the expected welfare when both types invest

$$W(\mathcal{M}, invest) := p \cdot W_h(invest|K_h) + (1-p) \cdot W_\ell(invest|K_\ell)$$

and when only the ℓ-type invests

$$W(\mathcal{M}, forgo) := p \cdot a_h + (1-p) \cdot W_\ell(invest|K_\ell)$$

Therefore, the regulator's objective will be to choose a mechanism to maximize welfare conditional on the h-type's decision to invest or forgo given the mechanism and the equilibrium of the *capital raising game* which we describe in the following subsection.

3.2.5 The Capital Raising Game $\Gamma(\mathcal{M})$

The regulator's mechanism $\mathcal{M} = \{K_\theta, T_\theta, \mathcal{S}_\theta\}_{\theta \in \{h,\ell\}}$ induces a capital raising game $\Gamma(\mathcal{M})$ played between the bank and the market. The game $\Gamma(\mathcal{M})$, illustrated in Figure 22, proceeds as follows: at time $t = 1$ the bank of type $\theta \in \{h, \ell\}$ decides whether to forgo or invest in the new investment. If the type θ bank forgoes, the game is over and its payoff is a_θ. If instead the bank decides to invest in the new asset, it must make a report to the regulator $\tilde{\theta} \in \{h, \ell\}$ and then issue an admissible capital security $s \in \mathcal{S}_{\tilde{\theta}}$ in order to generate funds totaling $P \geq K_{\tilde{\theta}} + T_{\tilde{\theta}}$. If the bank does not meet the specified capital requirement so that the funds generated from the sale of the security P are less than then the capital requirement K_θ and the ex-ante transfer T_θ then we assume its payoff is 0. This is consistent with the bank losing its charter and therefore being nationalized by the regulator, providing the bank's existing shareholders with a payoff of 0. If the bank invests and issues security s, then the market formulates a belief $\mu(s) := Pr(\theta = h|s) \in [0, 1]$ regarding the bank's type, equivalent to the probability that the bank's type is h given the security s it issues. Note that the beliefs μ will also depend on the mechanism $\mathcal{M}$ (i.e. whether it is separating, pooling, incentive compatible, etc.) and the bank's chosen menu option. As an abuse of notation we implicitly assume that these factors are included in μ without explicitly referencing them. The market then offers a payment $P(s)$ for the security s given its beliefs $\mu(s)$. We denote by $\mathbb{E}_{\mu(s)}[s]$ the markets valuation of the security s given their beliefs $\mu(s)$ regarding the bank's type and $\mathbb{E}_\theta[s]$ the type θ bank's true valuation of the security.

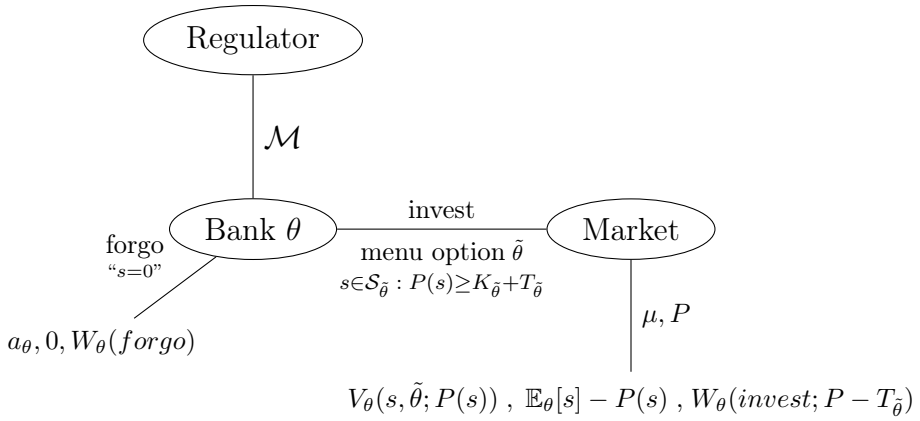


Figure 22: The capital raising game $\Gamma(\mathcal{M})$.

Given that the bank's decision to undertake the new investment is observable, we will represent the bank's decision to forgo the investment, without loss, by the issuance of the security $s = 0$ (i.e. $s(z) = 0$ for all z). In this case, whenever $s = 0$ the type θ bank's payoff is a_θ, the market payoff is 0, and welfare is $W_\theta(forgo)$. If instead, the bank reports its type is $\tilde{\theta}$ and it issues some security $s \in \mathcal{S}_{\tilde{\theta}}$ that generates funds $P(s) \geq K_{\tilde{\theta}} + T_{\tilde{\theta}}$ then the bank's payoff is $V_\theta(s, \tilde{\theta}; P(s))$, the market's payoff is $\mathbb{E}_\theta[s] - P(s)$, and welfare is $W_\theta(invest; P - T_{\tilde{\theta}})$.

The Underinvestment Problem: Given that the h-type security is always more valuable than the ℓ-type security (the h-type bank's post investment distribution of returns first order stochastically dominates the ℓ-type's) we can see that for any market beliefs μ and any security $s \in \mathcal{S}$, when transfers do not depend on type (i.e. $T_h = T_\ell = T$) then we have

$$\mathbb{E}_{\mu(s)}[s] - \mathbb{E}_h[s] \leq 0 \qquad \text{and} \qquad \mathbb{E}_{\mu(s)}[s] - \mathbb{E}_\ell[s] \geq 0$$

This states that the h-type's security is always weakly underpriced while the ℓ-type's security is always weakly overpriced (formally proven in Lemma 3.10 below). In this case, the ℓ-type will always find it profitable to invest provided that the transfer is not too large. The h-type on the other hand may find it optimal to forgo the investment (even with zero transfers) whenever the market puts a probability less than 1 on the bank being the h-type: $\mu(s) < 1$. The potential underinvestment created by this friction only exists when the value of the new investment is not too large. Namely, we can show that the h-type bank will never forgo the investment if the NPV of the new project $\hat{b} \geq a_h - a_\ell$ as in this case the value of the investment is so large that it is profitable for the h-type to invest even if the market holds the worst beliefs: $\mu(s) = 0$ for all $s \in \mathcal{S}$ (and optimally transfers will be zero). We therefore assume without loss that $\hat{b} < a_h - a_\ell$ throughout, noting that whenever this assumption does not hold then the regulator can achieve the first best outcome, inducing all banks to invest

while setting the maximum capital requirements $K_h = K_\ell = \bar{K}$ through the use of transfers $T_h = T_\ell = 0$ and securities $\mathcal{S}_h = \mathcal{S}_\ell = \mathcal{S}_{eq}$ where $\mathcal{S}_{eq}$ is the set of equity securities.

3.2.6 Equilibrium Concept and Refinements

In this subsection we will define our equilibrium concept for the game $\Gamma(\mathcal{M})$ and two refinements that we will be interested in. A strategy profile of the capital raising game $\Gamma(\mathcal{M})$ consists of a tuple (s_h, s_ℓ, μ, P) with $s_\theta \in \mathcal{S}_\theta \cup \{0\}$ the security issued by each type $\theta \in \{h, \ell\}$, $\mu : \mathcal{S} \to [0, 1]$ such that $\mu(s)$ is the market belief of the bank's type when it issues security s, and $P : \mathcal{S} \to \mathbb{R}$ such that $P(s)$ is the price offered by the market for a given security s. We will utilize the perfect Bayesian equilibrium solution concept which, in the context of $\Gamma(\mathcal{M})$, is defined as follows.

Definition 3.5. Let $\mathcal{M}$ be an incentive compatible mechanism. The strategy profile $e^\star = (s_h^\star, s_\ell^\star, \mu^\star, P^\star)$ is a perfect Bayesian equilibrium if it satisfies the following conditions:
(1) If $s_\theta^\star \neq 0$, then $P(s_\theta^\star) \geq K_\theta + T_\theta$ for each $\theta \in \{h, \ell\}$ and

$$s_\theta^\star \in \underset{s \in \mathcal{S}_\theta : P(s) \geq K_\theta + T_\theta}{\operatorname{argmax}} V_\theta(s, \theta; P(s))$$

(2) The beliefs $\mu^\star$ are consistent with the bank's type specific strategy $(s_h^\star, s_\ell^\star)$ so that $\mu^\star(s_\theta^\star)$ is computed using Bayes rule for each $\theta \in \{h, \ell\}$.
(3) The market price is competitive given the market beliefs: $P^\star(s) = \mathbb{E}_{\mu^\star(s)}[s]$ for all $s \in \mathcal{S}$.

The first two conditions represent the standard definition of perfect Bayesian equilibrium which requires that (1) if the bank invests, then the security $s_\theta^\star$ meets the capital requirement (i.e. sequential rationality of the investment decision) and the choice of security is sequentially rational with respect to the market beliefs $\mu^\star$, (2) the market beliefs are consistent with respect to the type specific strategy of the bank. Finally, condition (3) assumes that the market prices securities competitively so that the price the market offers for a security is exactly equal to the market's value of that security given its beliefs about the bank's type: $P^\star(s) = \mathbb{E}_{\mu^\star(s)}[s] = \mu^\star(s)\mathbb{E}_h[s] + (1 - \mu^\star(s))\mathbb{E}_\ell[s]$.

As is usual for signaling games, $\Gamma(\mathcal{M})$ has socially undesirable equilibria whereby, regardless of the mechanism $\mathcal{M}$, the h-type never invests. Namely, such an equilibrium outcome is supported by the beliefs $\mu(s) = 0$ for all $s \in \mathcal{S}$. We say that these equilibria are inefficient when there exists another equilibrium whereby the h-type invests, in which case $\mu(s) \neq 0$ for at least one security s. In this case, the h-type is unjustifiably excluded from the market given that if the market had beliefs that the h-type might invest ($\mu(s) > 0$) then it would

be optimal for the h-type to invest. Our first refinement will allow us to rule out any inefficient equilibria that require the market to ignore the informative signals produced by the mechanism and the bank's choice of menu option via an incentive compatible separating mechanism. To this end we will use the intuitive criterion of Cho and Kreps (1987).

Definition 3.6 (Intuitive Criterion Cho and Kreps (1987))**.** Let $e^\star = (s_h^\star, s_\ell^\star, \mu^\star, P^\star)$ be an equilibrium of the game $\Gamma(\mathcal{M})$ and let $u_\theta(s,\mu)$ be the payoff of the type-θ bank when issuing security s under beliefs μ. The equilibrium $e^\star$ satisfies the intuitive criterion if for any security $s \in \mathcal{S}$ such that for some $\theta, \theta' \in \{h, \ell\}$

$$u_\theta(s_\theta^\star, \mu^\star) < \max_\mu u_\theta(s,\mu)$$

and

$$u_{\theta'}(s,\mu)|_{\mu(s):Pr_\mu(\theta|s)=0} \geq u_{\theta'}(s_{\theta'}^\star, m^\star)$$

then $\mu^\star(s)$ is such that $Pr_\mu(\theta'|s) = 1$.

Note that this definition is simplified from the original definition of Cho and Kreps due to the fact that we are dealing only with two possible types. Namely, in the language of the general definition, whenever s is equilibrium dominated for type-θ (i.e. issuing s yields a lower off-path payoff for the type-θ than the equilibrium strategy no matter the off path beliefs) but not equilibrium dominated for type θ' then the market should not believe that the bank is type-θ when it observes security s being issued. This implies the market believes the bank is type θ' whenever there are only two types. The intuition here is that when such a condition is satisfied, then when seeing the out of equilibrium security s issued, the market should believe that the bank's type is θ' if there are no out of equilibrium beliefs that would make issuing s more profitable than $e^\star$ for type θ while the type θ' bank could profit by issuing s whenever the market believes the bank's type is θ' after s is issued.

One remaining issue is that there still exist equilibria of pooling mechanisms that satisfy the intuitive criterion but still arbitrarily deter investment by the h-type. Namely, in such an equilibrium, markets believe that only the ℓ-type will invest so that $\mu(s) = 0$ for all $s \in \mathcal{S}$. Therefore, it may be the case that the h-type will find it optimal to forgo the new investment even though it is optimal for the h-type to invest when the market believes both types invest (i.e. $\mu(s) = p$ for some $s \in \mathcal{S}$). When the optimal mechanism is of the pooling type, then these inefficient equilibria are always (welfare) dominated by any equilibrium whereby both types invest (provided such an equilibrium exists, which is the only case where the equilibria we attempt to rule out are in fact undesirable). In that case, we would like to think that the regulator's choice of a pooling mechanism should signal that both types will invest as

optimality of the pooling mechanism is publicly observable. We therefore introduce the following assumption.

Assumption 3.7. The regulator's choice of mechanism acts as a credible signal to the market of the h-type's investment decision. Namely, if the welfare of some equilibrium of a pooling mechanism where both types invest generates higher welfare than any equilibrium of any separating mechanism then the choice of the pooling mechanism credibly signals to the market that the h-type will invest.

Both Assumption 3.7 and the Intuitive Criterion refinement are not necessary if the regulator has the possibility to purchase the bank's security through a government recapitalization program. Namely, if the Intuitive Criterion or Assumption 3.7 do not hold then it may be the case that bad equilibria are coordinated on, but given the dynamic nature of security issuance this opens up a possibility for the banks to report when markets are undervaluing their securities. As proven in Lemma 3.8 below, by purchasing the bank's security in this situation the regulator can achieve a strict welfare improvement over the inefficient equilibrium outcome.

3.3 Preliminary Results and Equilibria of the Capital Raising Game

Before proceeding to characterize the equilibria of the capital raising game we will first present a few preliminary results.

3.3.1 Preliminary Results

First, we will show that the intuitive criterion and Assumption 3.7 are not necessary if the regulator has access to government recapitalizations.

Lemma 3.8. *Let $\mathcal{M}$ be a socially optimal mechanism. If there exists an equilibrium of $\Gamma(\mathcal{M})$ whereby the h-type invests and the regulator has the ability to purchase the banks' securities at their equilibrium prices, then doing so yields a strict expected welfare improvement over any equilibrium of $\Gamma(\mathcal{M})$ whereby the h-type forgoes the investment.*

Proof. See appendix Section 3.1.1.1. □

Lemma 3.8 states that if a socially optimal mechanism permits an equilibrium whereby the h-type invests, then inducing investment by the h-type must be socially optimal (otherwise the regulator could increase capital requirements or transfers to induce the bank to forgo this investment). Whenever this is the case, it is easy to show that by agreeing to

purchase the security of the bank at the price specified in the investment inducing equilibrium, the regulator strictly increases expected welfare. Namely, in this case the regulator successfully induces the h-type to invest (generating a positive expected surplus) and breaks even in expectation on the purchase of the security. This is a subtle argument but it is relevant as when the pooling mechanism is optimal it would always benefit society to allow state sponsored recapitalizations in the case that the market and the bank fail to coordinate on the efficient equilibrium.

The next lemma provides us with a more convenient expression for $V_\theta(s, \tilde{\theta}; P)$.

Lemma 3.9. *Let s be an admissible security that generates funds P and denote by $K_1 = P - T_{\tilde{\theta}}$ the capital generated from the sale of s. Then,*

$$V_\theta(s, \tilde{\theta}; P) := \mathbb{E}_\theta[max\{a_\theta + B + P - T_{\tilde{\theta}} - s, 0\}] = a_\theta + b_\theta(K_1) + K_1 - \mathbb{E}_\theta[s]$$

where

$$b_\theta(K_1) := \int_{-a_\theta - K_1}^{\infty} x dG(x) - G(-a_\theta - K_1) \cdot (a_\theta + K_1) = \hat{b} + L_\theta(K_1).$$

Furthermore, $b_\theta(K_1) = \hat{b} + L_\theta(K_1)$.

Proof. See appendix Section 3.1.1.2. □

Note here that $b_\theta(K_1)$ represents the net present value of the new investment to the bank given the newly raised capital $K_1 = P - T_{\tilde{\theta}}$ net the *contamination cost* of the risk that the new investment imposes on the bank's post investment capital stock $a_\theta + K_1$. Namely, once the bank has made the new investment, it losses its existing capital $a_\theta + K_1$ whenever the loss incurred by the new investment exceeds this value, which happens with probability $G(-a_\theta - K_1)$. The condition that $b_\theta(K_1) = \hat{b} + L_\theta(K_1)$ is equivalent to saying that the intrinsic value of the new investment to the bank is exactly equal to the investment's value under full liability plus the value of the deposit insurance (i.e. the value of the put option on the bank's assets with strike price $I - K_1$).

Finally, the next result will be useful for characterizing the equilibria of the capital raising game.

Lemma 3.10. *If $s \neq 0$ is an admissible security that generates funds P when the market believes the bank's type is $\theta \in \{h, \ell\}$ then*

$$\mathbb{E}_\theta[s] = \int_{-a_\theta - P}^{\infty} s(x + a_\theta + P) dG(x).$$

Further, whenever $T_h = T_\ell = T$ then fixing any $s \in \mathcal{S}$ and any value of P:
(1) $\mathbb{E}_h[s] > \mathbb{E}_\ell[s]$
(2) $\mathbb{E}_h[s] - \mathbb{E}_\ell[s]$ *is increasing in* a_h.

Proof. See appendix Section 3.1.2.1. □

This result states two important conditions that our admissible securities satisfy. The first is that the value of any security is always higher when the bank is the h-type (excluding the effect of transfers). Intuitively this due to the fact that the h-type's existing assets are more valuable than the ℓ-type's and they face the same new investment. The second result states that the difference in this value $\mathbb{E}_h[s] - \mathbb{E}_\ell[s]$, which we call the *information sensitivity* of s, is strictly increasing a_h (keeping a_ℓ fixed).

3.3.2 Equilibria of Pooling Mechanisms

We will now characterize the properties of perfect Bayesian equilibria of $\Gamma(\mathcal{M})$ for all pooling mechanisms $\mathcal{M}$. We show that there are effectively three types of equilibria.

Lemma 3.11. *Let $\mathcal{M}$ be a pooling mechanism with capital requirement K and transfer $T \leq \min\{b_h(K_1), b_\ell(K_1)\}$ where $K_1 \geq K$ is the capital raised, net the ex-ante transfer. Then, any equilibrium $e = (s_h, s_\ell, \mu, P)$ of $\Gamma(\mathcal{M})$ that satisfies the intuitive criterion satisfies one (and only one) of the following three properties:*
(i) $s_h = 0$, $\mathbb{E}_\ell[s_\ell] = K + T$.
(ii) $s_\ell = s_h = s$, $\mathbb{E}_p[s] \geq K + T$.
(iii) $s_\ell \neq s_h$, $\mathbb{E}_\ell[s_\ell] = K + T$, $\mathbb{E}_h[s_h] = K' + T$ *where* $K' > K$ *and* s_h *satisfy*

$$s_h \in \underset{\substack{s' \in \mathcal{S} \\ \mathbb{E}_h[s_h] = K' + T}}{\operatorname{argmin}} \; \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']$$

$$b_\ell(K) = b_\ell(K') + K' - \mathbb{E}_\ell[s_h]$$

Proof. See appendix Section 3.1.2.2. □

The first type-(i) equilibrium is the *inefficient* equilibrium discussed in Section 3.2.6. Important to note is that even under Assumption 3.7 this equilibrium will still be relevant when the underlying parameters of the model are such that the rents paid by the regulator to the bank (in the form of lower capital requirements) do not outweigh the benefit of inducing the h-type bank to invest (regardless of whether a separating or pooling mechanism is utilized). In this case, the optimal *underinvestment* (UI) mechanism sets a pooling requirement $K = \bar{K}$ and the h-type optimally forgoes the investment yielding the type-(i) equilibrium.

The second type-(ii) equilibrium will be the relevant *pooling equilibrium* whereby both types issue the same security $s \in \mathcal{S}$ and the market prices that security at its average price ($\mu(s) = p$) so that $\mathbb{E}_p[s] = p\mathbb{E}_h[s] + (1-p)\mathbb{E}_\ell[s] = K + T$. Important to note is that the security issued in any pooling equilibrium is the one that minimizes the information rents paid by the h-type to the market: $\mathbb{E}_h[s] - \mathbb{E}_\ell[s]$. In this case, this is equivalent to the banks issuing securities that minimize the *information sensitivity* of the security.

Finally, we show that there may exist a type-(iii) *separating equilibrium* whereby the h-type issues a security that generates more than the capital requirement so that the ℓ-type prefers to just meet the capital requirement and signal its type to the market than to raise the additional capital to mimic the h-type. As explained below we can effectively ignore this equilibrium as whenever it exists it can be implemented by a separating mechanism. Further, the fact that the type-(iii) equilibrium dominates the type-(ii) equilibrium implies that whenever it exists for the optimal pooling capital requirement K, then the best pooling mechanism is (weakly) dominated by the best separating mechanism.

3.3.3 Equilibria of Separating Mechanisms

The following result characterizes the properties of equilibria of $\Gamma(\mathcal{M})$ whenever $\mathcal{M}$ is an incentive compatible separating mechanism.

Lemma 3.12. *Let $\mathcal{M}$ be an incentive compatible separating mechanism with capital requirements K_ℓ and K_h. Then, any equilibrium (s_h, s_ℓ, μ, P) of $\Gamma(\mathcal{M})$ that satisfies the intuitive criterion with $s_h \neq 0$ and $s_\ell \neq 0$ is such that*
(i) $\mu(s_\ell) = 0$ and $\mu(s_h) = 1$.
(ii) $\mathbb{E}_\ell[s_\ell] = K_\ell + T_\ell$ and $\mathbb{E}_h[s_h] = K_h + T_h$.

Proof. See appendix Section 3.1.2.3. □

What this proposition states is that incentive compatibility guarantees that when the intuitive criterion is satisfied then the market beliefs always coincide with the bank's menu choice (as signaled through their choice of capital requirement and transfers). Furthermore, we show that the capital requirements will be optimally binding for both types. This is again due to the fact that the banks prefer to be as highly leveraged as possible as deposits are subsidized. Hence, the only thing that can prevent the banks from having binding capital requirements is if the market has strange beliefs that the bank that issues security s_θ but exactly meets the capital requirement (i.e. $\mathbb{E}_\theta[s_\theta] = K_\theta + T_\theta$) is not type θ. Such beliefs are ruled out by the intuitive criterion given that the mechanism is incentive compatible even when the capital requirement binds.

3.4 Optimal Mechanisms

As mentioned above, for some parameters of the model (conditions will be given below) having the h-type forgo investment in exchange for setting a high capital requirement for the ℓ-type will be socially optimal. In this case we assume, without loss, that the regulator utilizes the *underinvestment* pooling mechanism $\mathcal{M}^{\star}_{und}$ that sets $T_h = T_\ell = 0$ and $K^\star = \bar{K}$.[62]

In this section we will characterize the optimal mechanism when the regulator is restricted to the class of pooling mechanisms and then proceed to characterize the optimal mechanism when the regulator is restricted to separating mechanisms that dominate the optimal pooling mechanism. The reader can feel free to skip to the main results in Section 3.5 which is a characterization of the optimal mechanism (without restrictions), stating when the optimal pooling mechanism, optimal separating mechanism, or optimal underinvestment mechanism is preferred by the regulator as a function of the underlying parameters.

3.4.1 Optimal Pooling Mechanisms

Here we first note that we can focus without loss on type-(ii) equilibria of pooling mechanisms. Namely, given that any type-(iii) equilibrium of a pooling mechanism is payoff equivalent to an equilibrium of the separating mechanism $\mathcal{M}$ with $K_\ell = K^\star$, $K_h = K'$, $T_h = T_\ell = 0$ implies that whenever the pooling mechanism optimally sets a capital requirement $K^\star$ and permits a type-(iii) equilibrium for some $K' > K^\star$, then it is weakly dominated by the optimal separating mechanism. Therefore, in what follows we will only consider type-(ii) equilibria of pooling mechanisms as these are the relevant equilibria (under Assumption 3.7) when the pooling mechanism is not dominated.

The next proposition characterizes the optimal pooling mechanism.

Proposition 3.13. *Let $\mathcal{M}^{\star}_{pool}$ with $K_\ell = K_h = K^\star$, $T_\ell = T_h$, be the optimal pooling mechanism. Then, $T_\ell = T_h = 0$,*

$$\mathcal{S}_\ell = \mathcal{S}_h = \{s \in \mathcal{S} : s \in \underset{\substack{s\in\mathcal{S}\\ \mathbb{E}_p[s]=K^\star}}{\operatorname{argmin}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s]\}$$

and $K^\star$ is the unique value that solves

$$b_h(K^\star) = (1-p) \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_p[s]=K^\star}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s]$$

Proof. See appendix Section 3.1.3.2. □

[62] Our assumption that $\hat{b} < a_h - a_\ell$ guarantees that the h-type forgoes the investment under $\mathcal{M}^{\star}_{und}$.

Proposition 3.13 states that the optimal pooling mechanism sets transfers to zero and restricts banks to issue securities that minimize the information sensitivity. This latter characteristic is optimal as the security that minimizes the information sensitivity allows the regulator to set the highest possible capital requirement. This is due to the fact that under a pooling mechanism capital requirements are set to induce investment by the h-type and the cost the h-type pays when to investing is proportional to the information sensitivity of the security issued. Namely, when the capital requirement is K the bank invests and issues security s such that $\mathbb{E}_p[s] \geq K$ if and only if

$$b_h(K) \geq \mathbb{E}_h[s] - \mathbb{E}_p[s] = (1-p)(\mathbb{E}_h[s] - \mathbb{E}_\ell[s])$$

Hence, the regulator would like to minimize the information sensitivity of the security utilized as it allows him to weakly increase capital requirements. Therefore, the capital requirement $K^\star$ of the optimal pooling mechanism is set as high as possible to make the h-type bank indifferent between investing or not. We then show that this equation always yields an interior solution given that $b_h(K)$ is decreasing in K (banks have a preference for leverage) and the information sensitivity is increasing in the capital requirement:

$$\min_{s \in \mathcal{S}: \mathbb{E}_p[s]=K} \mathbb{E}_h[s] - \mathbb{E}_\ell[s] > \min_{s \in \mathcal{S}: \mathbb{E}_p[s]=K'} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']$$

whenever $K > K'$. Finally, whenever the bank is indifferent between investing and not investing under the capital requirement $K^\star$ then it is easy to see that there is a unique type-(ii) pooling equilibrium that induces investment whereby capital requirements bind so that $\mathbb{E}_p[s^\star] = K^\star$.

Next we characterize when it is optimal for the regulator to want to induce the h-type to invest through the optimal pooling mechanism $\mathcal{M}^\star_{pool}$ rather than setting the maximal capital requirement $\bar{K}$ and only having the ℓ-types invest through the optimal underinvestment mechansim $\mathcal{M}^\star_{und}$.

Lemma 3.14. *Let $K^\star$ be the capital requirement of the optimal pooling mechanism $\mathcal{M}^\star_{pool}$. $\mathcal{M}^\star_{pool}$ dominates the optimal underinvestment mechanism $\mathcal{M}^\star_{und}$ if and only if*

$$\hat{b} \geq \frac{\lambda}{p}(pL_h(K^\star) - (1-p)(L_\ell(K^\star) - L_\ell(\bar{K})).$$

Proof. See appendix Section 3.1.3.1. □

3.4.2 Optimal Separating Mechanisms

In this section we will proceed to characterize the optimal separating mechanisms. We will first characterize when inducing investment by the h-type in a separating mechanism is preferred to the optimal underinvestment mechanism $\mathcal{M}^{\star}_{und}$.

Lemma 3.15. *Let $\mathcal{M}^{\star}_{sep} = \{(K^{\star}_h, T^{\star}_h, \mathcal{S}^{\star}_h), (K^{\star}_\ell, T^{\star}_\ell, \mathcal{S}^{\star}_\ell)\}$ be the optimal separating mechanism. $\mathcal{M}^{\star}_{sep}$ dominates the optimal underinvestment mechanism $\mathcal{M}^{\star}_{und}$ if and only if*

$$\hat{b} \geq \frac{\lambda}{p}(p \cdot L_h(K^{\star}_h) + (1-p) \cdot (L_\ell(K^{\star}_\ell) - L_\ell(\bar{K}))).$$

Proof. See appendix Section 3.1.3.3. □

Before proceeding to characterize the optimal separating mechanism, we will first note that the incentive compatibility conditions can be written as

$$(IC_\ell) \qquad T_h - T_\ell \geq b_\ell(K_h) - b_\ell(K_\ell) + \mathbb{E}_h[s_h] - \mathbb{E}_\ell[s_h]$$

and

$$(IC_h) \qquad T_h - T_\ell \leq b_h(K_h) - b_h(K_\ell) + \mathbb{E}_h[s_\ell] - \mathbb{E}_\ell[s_\ell]$$

where s_h and s_ℓ are such that $\mathbb{E}_h[s_h] = K_h + T_h$ and $\mathbb{E}_\ell[s_\ell] = K_\ell + T_\ell$ (conditions satisfied in equilibrium). Further, under any incentive compatible separating mechanism, both bank types are indifferent between which security they issue when investing is optimal (i.e. the transfer is not too large). This is due to the fact that under any incentive compatible separating mechanism the bank's choice of capital requirement credibly reveals to the market its true type. Therefore, once the bank's type is revealed, any security that it issues is correctly priced and thus pays in expectation exactly the funding that it generates.

Proposition 3.16. *Let $\mathcal{M}^{\star}$ be an optimal separating mechanism with*

$$\mathcal{S}^{\star}_h = \{s \in \mathcal{S} : s \in \underset{\substack{s' \in \mathcal{S} \\ \mathbb{E}_h[s'] = K_h + T_h}}{\operatorname{argmin}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']\}$$

and

$$\mathcal{S}^{\star}_\ell = \{s \in \mathcal{S} : s \in \underset{\substack{s' \in \mathcal{S} \\ \mathbb{E}_\ell[s'] = K_\ell + T_\ell}}{\operatorname{argmax}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']\}$$

then $\mathcal{M}^{\star}$ weakly dominates all other separating mechanisms and strictly dominates any mechanism that sets $S_h \neq S^{\star}_h$ or $S_\ell \neq S^{\star}_\ell$ for some underlying parameters $(a_h, a_\ell, p, \hat{b})$.

Proof. See appendix Section 3.1.3.4. □

Proposition 3.16 states that any optimal separating mechanism is weakly dominated by the separating mechanism that restricts the h-type to issue the least informationally sensitive security and the ℓ-type to issue the most informationally sensitive security subject to binding capital requirements (dictated by the equilibrium conditions). This comes from the fact that restricting securities to these sets can only relax the incentive constraints (allowing for the possibility of improving welfare).

Corollary 3.17. *If the bank can sell its existing assets to finance the new investment, then* $\mathcal{S}_\ell^\star = \{s_{AS}\}$ *where* s_{AS} *is the security that represents the sale of the existing assets to the market. If* $K_\ell + T_\ell > a_\ell$ *then without loss* s_{AS} *sells a fraction of the new investment (via equity issuance) to generate the remaining funds necessary to meet the capital requirement.*

Proof. See appendix Section 3.1.3.5. □

Corollary 3.17 states that if the bank can sell its existing assets to finance the capital requirement then the optimal separating mechanism requires the ℓ-type to finance the new investment by doing so. In addition, if the sale of the bank's existing assets cannot generate enough funds to meet the capital requirement then it is without loss to have the remaining funds financed by selling a claim on the new investment by issuing equity *after* the original assets have been sold off. This latter point comes from the fact that once the original assets have been sold off, then the bank is no longer privately informed and therefore the security used to generate the additional funds is irrelevant.

Lemma 3.18. *Let* $\mathcal{M}_{sep}^\star$ *be the optimal separating mechanism. If*

$$\max_{\substack{s\in\mathcal{S}\\ \mathbb{E}_\ell[s]=\bar{K}+T_\ell}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s] > \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_h[s]=\bar{K}+T_h}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s]$$

and

$$\min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_h[s]=\bar{K}+T_h}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s] \leq \hat{b}$$

then $\mathcal{M}_{sep}^\star$ *achieves the first best:* $K_\ell = K_h = \bar{K}$*. There exists* $\hat{p}$ *such that whenever* $p > \hat{p}$*, if* $\mathcal{M}_{sep}^\star$ *achieves the first best, then so does* $\mathcal{M}_{pool}^\star$*.*

Proof. See appendix Section 3.1.3.6. □

Lemma 3.18 states conditions under which the optimal separating equilibrium leads to the first best outcome. We do not expect the conditions of Lemma 3.18 to hold in practice for sensible distributions G and we can show that they do not hold numerically (e.g. under a normal distribution). Additionally, we can show that if $\bar{K}$ is arbitrarily large, then the

second condition will fail under the assumption that $\hat{b} < a_h - a_\ell$. We proceed assuming that these conditions do not hold in order to characterize the second best separating mechanism.

We proceed with the following lemma which states that whenever the optimal separating mechanism $\mathcal{M}^\star_{sep}$ attains a higher level of welfare than the optimal pooling mechanism then the constraint IC_ℓ is always binding.

Lemma 3.19. *Let $\mathcal{M}^\star_{sep} = \{(K^\star_h, T^\star_h, \mathcal{S}^\star_h), (K^\star_\ell, T^\star_\ell, \mathcal{S}^\star_\ell)\}$ be the optimal separating mechanism. If $\mathcal{M}^\star_{sep}$ dominates $\mathcal{M}^\star_{pool}$ and does not achieve the first best outcome then,*
(i) IC_ℓ is always binding.
(ii) If $K^\star_\ell > K^\star_h$ then IC_h is not binding.

Proof. See appendix Section 3.1.3.7. □

Finally, the following proposition summarizes the optimal separating mechansim.

Proposition 3.20. *Let $\mathcal{M}^\star_{sep} = \{(K^\star_h, T^\star_h), (K^\star_\ell, T^\star_\ell)\}$ be the optimal separating mechanism. If $\mathcal{M}^\star_{sep}$ dominates the optimal pooling mechanism $\mathcal{M}^\star_{pool}$ then,*
(i) if $K^\star_\ell > K^\star_h$ then $T^\star_h = b_h(K^\star_h)$ and $T^\star_\ell = 0$.
(ii) if $K^\star_h \geq K^\star_\ell$ then $T^\star_h$ and $T^\star_\ell$ are chosen to solve the program

$$\min_{K_\ell, K_h, T_\ell, T_h} p \cdot L_h(K_h) + (1-p) \cdot L_\ell(K_\ell)$$

$$b_\ell(K_h) - b_\ell(K_\ell) + \min_{\substack{s \in \mathcal{S} \\ \mathbb{E}_h[s] = K_h + T_h}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s] = T_h - T_\ell$$

$$b_h(K_h) - b_h(K_\ell) + \max_{\substack{s \in \mathcal{S} \\ \mathbb{E}_\ell[s] = K_\ell + T_\ell}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s] \geq T_h - T_\ell$$

$$T_h \in [0, b_h(K_h)] \quad \textit{and} \quad T_\ell \in [0, b_\ell(K_\ell)]$$

Proof. See appendix Section 3.1.3.8. □

We note that whenever the optimal separating mechanism sets $K^\star_\ell > K^\star_h$, then we can determine the optimal transfers. Otherwise, it the optimal transfers in general will depend on the distribution of returns. Hence, we obtain a partial characterization in this latter case. The following lemma will prove useful when characterizing when separating is preferred to pooling and vice-versa

Lemma 3.21. *Let $\mathcal{M}^\star_{sep}$ be the optimal separating mechanism. There exists $\bar{p} \in [0,1]$ such that whenever $p < \bar{p}$ then $K^\star_\ell > K^\star_h$ and when $p > \bar{p}$ then $K^\star_h > K^\star_\ell$. Further, $\bar{p}$ is strictly increasing in $a_h - a_\ell$.*

Proof. See appendix Section 3.1.3.9. □

Example 3.22. Suppose that $B \sim \mathcal{N}(\hat{b}, \sigma^2)$, $I = 10$, $a_\ell = 0$, and $a_h = 2 \cdot I$ and let $\bar{p}_{\sigma^2}$ be the threshold of Lemma 3.21. Then,

$$\begin{aligned}\bar{p}_1 &> .9999999999\\ \bar{p}_{10} &> .9999998377\\ \bar{p}_{50} &> .9711209822\\ \bar{p}_{100} &> .8745895451\end{aligned}$$

This example will prove to be relevant in the context of our main results below. Namely, it shows that practically we expect $\bar{p}$ to be arbitrarily close to 1 when returns are normally distributed with reasonable variance.

3.5 Comparison of Optimal Mechanisms

We will now proceed to characterize under what conditions each of the mechanisms $\mathcal{M}^\star_{sep}$, $\mathcal{M}^\star_{pool}$, and $\mathcal{M}^\star_{und}$ are optimal.

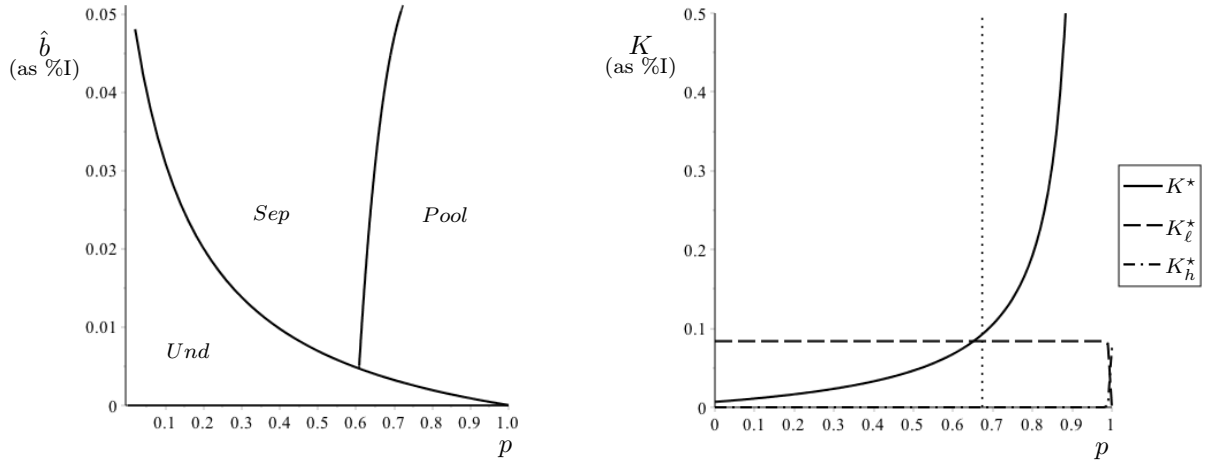
Proposition 3.23. *Let $\mathcal{M}^\star$ be the optimal regulatory mechanism. There exists $p_{pool}, p_{sep}, p_{und} \in [0,1)$ such that $p_{pool} \geq p_{sep} \geq p_{und}$ and*
(i) Whenever $p \geq p_{pool}$ then $\mathcal{M}^\star = \mathcal{M}^\star_{pool}$.
(ii) Whenever $p \in (p_{und}, p_{sep})$ then $\mathcal{M}^\star = \mathcal{M}^\star_{sep}$.
(iii) Whenever $p \leq p_{und}$ then $\mathcal{M}^\star = \mathcal{M}^\star_{und}$.
If $p_{sep} \neq p_{pool}$ then either $\mathcal{M}^\star = \mathcal{M}^\star_{pool}$ or $\mathcal{M}^\star = \mathcal{M}^\star_{sep}$ when $p \in (p_{sep}, p_{pool})$ depending on the remaining parameters and the distribution G.

Proof. See appendix Section 3.1.4.1. □

Proposition 4.4 states that the optimal mechanism is pooling for large values of p, separating for intermediate values, and underinvestment for small values of p. This idea is conveyed in Figure 21 in the introduction and Figure 23 (a) below. Namely, as the proportion of good banks goes to 1 then the market price of the equilibrium security converges to the good type bank's true valuation of the security. In that case the cost of raising capital goes to zero and therefore the regulator can set higher and higher capital requirements while still inducing investment. On the other hand, as the proportion of good banks goes to zero then the cost of underinvestment goes to zero as good banks are the only type who forgo investment when the capital requirement is too high. In that case, the benefit of setting higher capital requirements for the low type banks eventually becomes larger than the cost of underinvestment as p goes to zero. Finally, we note that given that the incentive compatibility constraints can be satisfied for positive values of K_ℓ and K_h, and these values do

not adjust with p whenever $p < \bar{p}$, then it must be the case that separation is preferred to pooling with p is not to large and separation is preferred to underinvestment with p is not to small.

Figure 23 (a) plots the regions of NPV $\hat{b}$ and proportion of good banks p whereby each mechanism $\mathcal{M}^{\star}_{sep}$, $\mathcal{M}^{\star}_{pool}$, and $\mathcal{M}^{\star}_{und}$ is optimal. Using this figure, the main result of Proposition 4.4 can be illustrated by fixing an intermediate value for $\hat{b}$. For example, taking $\hat{b} = .03 \cdot I$ (i.e. the new investment generates a 3% net return) we can see that when p is less than $\approx .1$ then the underinvestment mechanism is optimal, when is p between $\approx .1$ and $\approx .65$ then the optimal mechanism is separating, and when p is greater than $\approx .65$ then the optimal mechanism is pooling. Figure 23 (b) plots the capital requirements of the optimal pooling ($K^{\star}$) and separating ($K^{\star}_{\ell}, K^{\star}_{h}$) mechanisms as a function of p. As we can see, typically the high type capital requirement under $\mathcal{M}^{\star}_{sep}$ is zero unless p is very large, illustrating the result of Lemma 3.21. Further, provided that the incentive constraints are independent of the proportion of good banks p, the capital requirements under $\mathcal{M}^{\star}_{sep}$ are flat and only move when p is very large to reflect the decreasing weight that the regulator puts on recapitalizing the small proportion of banks with bad news. Next, we note that as p increases the capital requirement of $\mathcal{M}^{\star}_{pool}$ eventually asymptotes. This is due to the fact that as $p \to 1$, the cost of raising capital for the high type in the pooled environment converges to zero. Hence, as $p \to 1$ the regulator can set a higher and higher pooled capital requirement $K^{\star}$ while still inducing investment by the h-type bank. Finally, we have illustrated in Figure 23 (b) the value of p (given by the dotted vertical line) at which $\mathcal{M}^{\star}_{pool}$ and $\mathcal{M}^{\star}_{sep}$ generate the same welfare with $\mathcal{M}^{\star}_{pool}$ ($\mathcal{M}^{\star}_{sep}$) yielding a higher welfare when p is larger (smaller).



(a) Regions of optimal mechanisms given $(\hat{b}, p)$ with $B \sim \mathcal{N}(\hat{b}, 4)$, $\lambda = \frac{1}{10}$, $I = 10$.

(b) Optimal capital requirements as a function of p: $\hat{b} = .03 \cdot I$, $a_h - a_\ell = .5 \cdot I$.

Figure 23: Regions of optimal mechanisms (a) and optimal capital requirements (b).

One issue that can be observed in the statement of Proposition 4.4 is that it need not be the case that $p_{sep} = p_{pool}$, in which case there may be values of $p_1, p_2 \in (p_{sep}, p_{pool})$ such that $p_1 < p_2$ and pooling is optimal when $p = p_1$ yet separation is optimal when $p = p_2$. This is due to the fact that while $K^\star$ is increasing as p increases, so does $K_h^\star$ in the limit as the marginal benefit of recapitalizing the ℓ-type bank goes to zero. Hence, it is not clear whether $K_h^\star$ increases faster or slower than $K^\star$ for large values of p. The next proposition allows us to state when we have a full characterization.

Corollary 3.24. *Let p_{pool}, p_{sep}, and p_{und} be the values of Proposition 4.4. Then,*
(i) If $p_{pool} < \bar{p}$ then $p_{pool} = p_{sep}$.
(ii) There exists $\bar{a}$ such that whenever $a_h > \bar{a}$, then $p_{pool} < \bar{p}$ and therefore $p_{pool} = p_{sep}$.

Proof. See appendix Section 3.1.4.2. □

Corollary 3.24 gives us a full characterization of the optimal mechanism for all $p \in [0, 1]$. Namely, it states that (i) whenever $p_{pool} < \bar{p}$ of Lemma 3.21 then we have a full characterization and that (ii) there always exists $\bar{a}$ such that $p_{pool} > \bar{p}$ whenever $a_h > \bar{a}$. It is worth noting that although we do not have a full characterization whenever $p_{pool} < \bar{p}$, it is straightforward to extend our results to a full characterization as soon as the distribution of returns G is specified.

Finally, we note that Proposition 4.4 does not necessarily imply that $p_{sep} > p_{und}$. Namely, we do not rule out the case where $p_{sep} = p_{und}$, in which case separation is never optimal. The next corollary states that whenever $\hat{b}$ is large enough, then it must be the case that $p_{sep} > p_{und}$.

Corollary 3.25. *Let p_{pool}, p_{sep}, and p_{und} be the values of Proposition 4.4. Then,*
(i) If $b_\ell(\bar{K}) \geq b_\ell(0) - b_h(0)$ then $p_{und} = 0$.
(ii) There exists $\bar{b}$ such that whenever $\hat{b} > \bar{b}$ then $b_\ell(\bar{K}) < b_\ell(0) - b_h(0)$ and therefore $p_{und} = 0$.

Proof. See appendix Section 3.1.4.3. □

3.6 Policy Implications

In this section we will present the main policy implications of our results.

1. Internal Ratings Based v.s. Standardized Approach Regulations. From a cross sectional perspective, our results would suggest that the regulator should impose the IRB approach regulations on large and opaque banks, in line with its current use, while the SA approach should be utilized for more transparent banks. The key insight here is that transparency is an important parameter to determine the optimal regulation and therefore

regulators should work to develop accurate measures of bank transparency to utilize for regulatory purposes. This observation can potentially lend insight into why the spill over effects of the financial crisis were so large, given that banks' balance sheets had become increasingly opaque prior to the housing market crash through the widespread use of off balance sheet activities and the origination and trading of opaque assets such as mortgage backed securities.

Another point to note is that under current regulations the largest banks have discretion over which approach (IRB or SA) they use to determine their capital requirements. We note that, in our model, if the bank were to have the ability to choose the separating (IRB-type) or pooling (SA-type) mechanism before learning their type, then it is easy to show that whenever the IRB-type mechanism is socially optimal, the bank would prefer to utilize the SA-type mechanism. Similarly, whenever the SA-type mechanism is socially optimal the bank would prefer to utilize the IRB-type mechanism in most cases (whenever $p < \bar{p}$). Therefore, our results suggest that the regulator should remove the discretion of the banks to choose which approach they utilize in determining their capital requirements. Basel III has introduced a revised capital requirement *output floor* that limits the benefit banks can receive from utilizing the IRB approach which limits their capital requirement to be at least 72.5% of the SA requirement. This backstop can help to limit inefficiencies due to banks choosing the suboptimal framework but in our model would still lead to a suboptimal outcome.

2. Counter Cyclical Capital Buffer (CCyB). Basel III has introduced a counter cyclical capital buffer requiring an additional capital surcharge of 0-2.5% of core tier 1 capital to risk weighted assets. The purpose of this buffer is to allow local regulators to increase capital requirements during booms in order to prevent the excessive build up of aggregate credit and to be able to relax capital requirements during recessions in order to reduce credit rationing. This is an idea that is at the heart of this paper. Most importantly, we provide a foundation for how capital requirements lead to credit growth and rationing.

The implications of our results to the CCyB are that the regulator should only expect changes in the credit supply to come from opaque banks with good news. Therefore, the regulator can increase capital requirements on banks with transparent balance sheets, or banks that have recently been stress tested by the regulator (provided that the results of the stress test are public). Similarly, whenever the regulator utilizes the IRB-type mechanism that we propose in this paper then there will be no credit rationing so that capital requirements can be set as high as possible subject to meeting incentive compatibility of truthful reporting. Finally, we note that while local jurisdictions are encouraged to utilize the credit-to-GDP ratio in determining their CCyB, a key implication of our model is that the profitability of new investments should also influence capital requirement buffers as the more profitable the

investment is, the higher capital requirements the regulator can optimally set.

3. Government Interventions During Crisis Periods. During the financial crisis, government interventions were crucial to restore the faith in the banking system. While these interventions, such as TARP, served as a way to recapitalize banks, they also served as a way to signal information about the bank's quality to the market given that banks were only accepted to the programs after being heavily screened by the regulator. Our IRB-type mechanism is in effect a private solution to this problem. Namely, once the regulator designs capital requirements and transfers correctly, the banks will be screened into different classes (without imposing monitoring costs on the regulator), providing an informative signal to the market regarding their quality. We further note, as explained in Section 2.1, that government intervention may be necessary due to mis-coordination of the bank and market on inefficient signaling equilibria. Namely, we show that there can exist inefficient equilibria of the capital raising game whereby at certain times the market forms an extraneous belief that only the banks with bad news will invest. In such a situation, this can cause the banks with good news to forgo investments given the market's under pricing of their securities, thereby enforcing the market's belief. We show in Lemma 3.8 that the regulator can resolve this issue through a government recapitalization program such as TARP by agreeing to purchase the bank's security at the efficient equilibrium price and that this is strictly welfare improving with respect to the inefficient equilibrium outcome.

4. Stress Testing. We have yet to discuss stress testing of banks, a highly utilized regulatory practice since the crisis. Stress testing would complement the mechanisms in our model provided that the results of the stress test are made public and reveal credible information about the bank's asset quality. This lends to the debate regarding whether the results of regulatory supervision should be disclosed to the market by highlighting how doing so will help to resolve the adverse selection problem that raising capital creates.[63] In this sense, it would be most appropriate to utilize stress tests when the level of bank opacity is large. Namely, while the regulator can utilize our IRB-type mechanism to reveal the bank's private information, such an approach requires paying information rents in the form of lower capital requirements in order to credibly induce this information revelation. Further, these information rents are strictly increasing in the opacity of the bank's assets. Hence, when the level of bank opacity is large, the benefit of information revelation through stress testing will outweigh the cost of performing the test. This comes from the fact that once the regulator reveals the information gathered during the bank's stress test, then that bank's capital security will be more accurately priced allowing the regulator to set a higher

[63]Note that this relates potentially more to regulatory supervision of bank solvency rather than stress testing.

capital requirements (through either mechanism) without inducing underinvestment. These insights complement the current literature on stress testing and information disclosure (e.g. Leitner and Williams (2017) and Goldstein and Leitner (2018)).

5. Capital Security Design. Finally, we would like to mention our results on security design. In current regulations equity is considered the *highest quality* capital instrument. This is due to the fact that equity allows the bank to absorb maximal losses before becoming insolvent in comparison to other securities such as subordinated debt that only absorbs losses after the bank fails. Yet, this begs the question of whether the regulator should be concerned with absorbing losses pre-insolvency or post-insolvency. While for large and systemic banks it is clear that pre-insolvency loss absorption provides a much larger benefit to society, this may not be the case for smaller, less-systemic banks. What we show in this paper is that in the latter case the regulator may want to consider the use of less *informationally sensitive* securities for capital regulation (e.g. subordinated debt). Similarly, given the current interest in hybrid debt securities such as contingent convertible bonds (see e.g. Squam Lake (2010)), our paper states that, barring any potential pricing or other issues that these new securities may impose, the use of these instruments can allow the regulator to set higher capital requirements without inducing underinvestment and yet still maintaining the same level of pre-insolvency loss absorption. Finally, we show how under the optimal IRB-type mechanism the regulator should force the banks with bad news to sell their existing assets in order to finance the new investment, something that we saw done in practice through the use of the TARP program during the crisis.

3.7 Conclusion

In this paper we have analyzed how capital requirements should optimally be set when banks must issue new securities to meet regulatory capital requirements. Adverse selection builds the natural link between higher capital requirements and underinvestment. We then proceed to characterize the problem of designing the optimal mechanism in this environment and show that three regulatory frameworks may be optimal over the underlying parameter space.

The first type of mechanism bypasses the investment incentives of the firms by inducing them to truthfully reveal their private information to the market. Namely, we show that under such a mechanism the bank's securities are correctly priced by the market and therefore all banks optimally invest regardless of the capital requirement. That being said, the regulator is restricted to set capital requirements to ensure that it is incentive compatible for the banks to truthfully reveal their private information, thereby paying information rents to

induce truthful revelation. The second type of mechanism instead pools the information of the banks by setting a single capital requirement. In this case the capital requirement is set as high as possible subject to inducing investment by the banks with good news. Finally, we show that it may also be optimal for the regulator to set capital requirements very high, purposefully inducing the banks with good news to forgo the new investment. We characterize under what conditions each of these three mechanisms is optimal given the underlying parameters of the model and the resulting policy implications. Given the tendency for the literature to take the cost of capital as exogenously given (motivated by information frictions), we hope that this model and its insights will prove to be useful for studying more complex issues of banking regulation in future research.

References

[1] Basel Committee on Banking Supervision (2015): "Finalising post-crisis reforms: an update: A report to G20 leaders" *Bank for International Settlements*, November 2015.

[2] Black, L., Floros, I., and Sengupta, R. (2016): "Raising Capital When the Going Gets Tough: U.S. Bank Equity Issuance from 2001 to 2014." Federal Reserve Bank of Kansas City Working Paper No. 16-05.

[3] Blum, J, M. (2008): "Why 'Basel II' may need a leverage ratio restriction" *Journal of Banking & Finance*, 32(8), 1699-1707.

[4] Bouvard, M., Chaigneau, P., and de Motta, A. (2015): "Transparency in the Financial System: Rollover Risk and Crises." *Journal of Finance*, 70(4).

[5] Cho, I. and Kreps, D. (1987): "Signaling games and stable equilibria," *Quarterly Journal of Economics*, 102, pp. 179-221.

[6] Colliard, J-E. (2018): "Strategic selection of risk models and bank capital regulation." *Management Science*, Articles in Advance, 1-16.

[7] Dewatripont, M., Rochet, J.-C., and Tirole, J. (2010):"Balancing the Banks: Global Lessons from the Financial Crisis." *Princeton University Press.*

[8] Fraisse, H., Lé, M., and Thesmar, D. (2017):"The Real Effects of Bank Capital Requirements" *ESRB Working Paper Series*, 47, European Systemic Risk Board.

[9] Giammarino, R. M., Lewis, T., and Sappington, D. E. M. (1993): "An incentive approach to banking regulation" *Journal of Finance*, 48, 1523-1542.

[10] Goldstein, I., and Leitner, Y. (2018): "Stress tests and information disclosure" *Journal of Economic Theory*, 177, 34-69.

[11] Gropp, R., Mosk, T., Ongena, S., and Wix, C. (2016):"Bank response to higher capital requirements: Evidence from a quasi-natural experiment" *SAFE Working Paper Series*, No. 156, Sustainable Architecture for Finance in Europe.

[12] Hanson, S. G., Kashyap, A. K., and Stein, J. C. (2011): "A Macroprudential Approach to Financial Regulation," *Journal of Economic Perspectives*, 25(1), 3-28.

[13] Innes, R. D. (1990): "Limited Liability and Incentive Contracting with Ex-ante Action Choices" *Journal of Economic Theory*, 52, 45-67.

[14] King, R. M. (2009): "The cost of equity for global banks: a CAPM perspective from 1990 to 2009." *BIS Quarterly Review*.

[15] Leitner, Y. (2014): "Should Regulators Reveal Information About Banks? Federal Reserve Bank of Philadelphia *Business Review*, Third Quarter.

[16] Leitner, Y. and Williams, B. (2018): "Model Secrecy and Stress Tests" *FRB of Philadelphia Working Paper*, No. 17-41: https://ssrn.com/abstract=3077205

[17] Leitner, Y., and Yilmaz, B. (2019): "Regulating a model," *Journal of Financial Economics*, 131 (2), 251-268.

[18] Morris, A. D. and White, L. (2005): "Crises and Capital Requirements in Banking" *The American Economic Review*, 95 (5), 1548-1572.

[19] Myers, S. C. and Majluf, N. S. (1984): "Corporate financing and investment decisions when firms have information that investors do not have." *Journal of Financial Economics*, 13(2), 187-221.

[20] Nachman, D. C. and Noe, T. H. (1994): "Optimal Design of Securities Under Asymmetric Information" *The Review of Financial Studies*, 7(1), pp. 44.

[21] Noe, T. H. (1988): "Capital Structure and Signaling Game Equilibria," *Review of Financial Studies*, 1, 331-355.

[22] Peek, J. and Rosengren, E. (1995): "Bank regulation and the credit crunch." *Journal of Banking & Finance*, 19, 679-692.

[23] Philippon, T. and Schnabl, P. (2013): "Efficient recapitalization." *The Journal of Finance*, 68(1).

[24] Philippon, T. and Skreta, V. (2012): "Optimal interventions in markets with adverse selection." *The American Economic Review*, 102(1), 1-28.

[25] Plosser, M. and Santos, J. A. C. (2018): "Banks' Incentives and Inconsistent Risk Models" *The Review of Financial Studies*, 31(6), 2080-2112.

[26] Prescott, E. S. (2004): "Auditing and bank capital regulation." *Federal Reserve Bank Richmond Economic Quarterly*, 90(4), 47-63.

[27] Repullo, R. (2013): "Cyclical adjustment of capital requirements: a simple framework." *Journal of Financial Intermediation*, 22(4), 608-626.

[28] Samuels, S., Harrison, M., and Rajkotia, N. (2012): "Bye Bye Basel? Making Basel More Relevant" *Barclays Equity Research* (London).

[29] Squam Lake Working Group on Financial Regulation (2010): "An expedited resolution mechanism for distressed financial firms: regulatory hybrid securities." Squam Lake working group, Council on Foreign Relations, April.

[30] Stein, J. C. (1998): "An adverse-selection model of bank asset and liability management with implications for the transmission of monetary policy." *RAND Journal of Economics*, 29(3), 466-486.

[31] Stiglitz, J. E., Weiss, A. (1981): "Credit rationing in markets with imperfect information." *American Economic Review*, 71(3), 393-410.

[32] Thakor, A. V. (1996): "Capital requirements, monetary policy, and aggregate bank lending: theory and empirical evidence." *The Journal of Finance*, 51(1), pp. 279-324.

[33] Tirole, J. (2012): "How public intervention can restore market functioning," *The American Economic Review*, 102(1), 29-59.

3.8 Extensions

3.8.1 Continuum of Types

In this section we will show that our main results extend to the case where the bank's private information is the updated value of its assets in place a which falls in some interval $[\underline{a}, \bar{a}]$. In this case, we assume the market and the regulator have a prior belief $p \in \Delta([\underline{a}, \bar{a}])$ over $[\underline{a}, \bar{a}]$. In this case, we will parameterize the asymmetric information problem by $\mathbb{E}_p[a] := \hat{a} \in [\underline{a}, \bar{a}]$, the market expectation of the bank's assets in place with respect to the prior p. In this

sense, as $\hat{a}$ increases this is equivalent to saying that p puts a higher probability on good news types.

3.8.1.1 Pooling

In a pooling mechanism, the regulator sets a single capital requirement K and the bank's type specific decision is given by $d_a(K) \in \{0, 1\}$ where $d_a(K) = 1$ implies that the bank issues a security $s \in \mathcal{S}$ and makes the investment when its type is a while $d_a(K) = 0$ implies the bank forgoes the investment. Note that without loss we can focus on pooling equilibria of the pooling mechanism as if there exist some semi-separating equilibria that dominate the pooling mechanism then the regulator can implement these equilibria using a separating mechanism and therefore the pooling mechanism is dominated. Furthermore, the regulator can rule out any other semi separating equilibria from being coordinated on by restricting $\mathcal{S}_a = \mathcal{S}_{a'} = \{s\}$ to be a single security s for all $a, a' \in [\underline{a}, \bar{a}]$ thereby removing the possibility of the bank's security signaling its type.

Now, given the nature of the problem we know that for any K, if all bank types invest, then

$$b_{\bar{a}}(K) \geq \min_{\substack{s \in \mathcal{S} \\ \mathbb{E}_{pool}[s]=K}} \mathbb{E}_{\bar{a}}[s] - \mathbb{E}_{pool}[s]$$

where

$$\mathbb{E}_{pool}[s] := \int_{\underline{a}}^{\bar{a}} \mathbb{E}_a[s] p(a) da.$$

Otherwise, for each K, there exists a unique threshold $\tau(K)$ such that for all $a > \tau(K)$ the bank forgoes the project ($d_a(K) = 0$) and for all $a < \tau(K)$ the bank undergoes the investment $d_a(K) = 1$. Of course, in this case $\hat{a}$ is determined by this threshold. Therefore we denote by

$$\hat{a}(K) = \int_{\underline{a}}^{\tau(K)} a p(a) da \tag{3.1}$$

the market expectation when all banks $a > \tau(K)$ forgo the investment. In that case, for any K, $\tau(K)$ is determined by 3.1 and

$$b_{\tau(K)}(K) = \min_{\substack{s \in \mathcal{S} \\ \mathbb{E}_{p_\tau}[s]=K}} \mathbb{E}_{\bar{a}}[s] - \mathbb{E}_{p_\tau}[s] \tag{3.2}$$

where

$$\mathbb{E}_{p_\tau}[s] := \int_{\underline{a}}^{\tau(K)} \mathbb{E}_a[s] p_\tau(a) da.$$

and $p_\tau(a) := p(a|a \geq \tau(K))$.

Therefore, denoting by $L_a(K)$ the liability of the a-type bank, the regulator chooses the optimal pooling mechanism to solve the program

$$\max_K \int_{\underline{a}}^{\tau(K)} (\hat{b} - \lambda \cdot L_a(K)) p(a) da$$

where $\tau(K)$ solves 3.1 and 3.2. It should be straightforward to see that as $\hat{a} \to \bar{a}$ then $\tau(K) \to \bar{a}$ and $K \to +\infty$.

3.8.1.2 Separating

Now, we have characterized the optimal pooling mechanism we can proceed to characterize the optimal separating mechanism. Note that we will assume here that full separation is optimal which may not always be the case. It should be straightforward to extend our characterization to the case of semi-separation. We further assume that p is single peaked so that p is weakly increasing for all $a \in [\underline{a}, \hat{a})$ and weakly decreasing for all $a \in (\hat{a}, \bar{a}]$.

Now, denoting by K_a the capital requirement of type a and T_a the transfer to be paid by type a, incentive compatibility requires that for any $a, a' \in [\underline{a}, \bar{a}]$ we have

$$T_{a'} - T_a \geq b_a(K_{a'}) - b_a(K_a) + \mathbb{E}_{a'}[s_{a'}] - \mathbb{E}_a[s_{a'}]$$

and

$$T_{a'} - T_a \leq b_{a'}(K_{a'}) - b_{a'}(K_a) + \mathbb{E}_{a'}[s_a] - \mathbb{E}_a[s_a]$$

While we can say more about the optimal design of the separating mechanism it is not necessary to present the extension of our results to a continuum of types.

3.8.1.3 Extension of Proposition 4.4 to the Continuum Case

Now, before presenting the analog to Proposition 4.4 we first note that the no investment mechanisms are more complicated in the continuum setting as the regulator can set capital requirements to induce investment from all types $a < \tilde{a}$ for any threshold $\tilde{a}$. Denote by $\mathcal{M}_{no(\tilde{a})}$ the optimal pooling mechanism with capital requirement $K^\star$ such that $\tau(K^\star) = \tilde{K}$. In this case, the optimal *full investment* pooling mechanism is simply $\mathcal{M}^\star_{pool} = \mathcal{M}^\star_{no(\bar{a})}$. In this case we can see that there is some added insight as $\hat{a} \to 0$ then it must be the case that $\mathcal{M}^\star_{pool}$ is dominated by all mechanisms $\mathcal{M}^\star_{no(\tilde{a})}$ with $\tilde{a} < \bar{a}$.

Proposition 3.26. *Let $\mathcal{M}^\star$ be the optimal regulatory mechanism. There exists a_{pool}, a_{sep}, $a_{no} \in [\underline{a}, \bar{a})$ such that $a_{pool} \geq a_{sep}$ and*

(i) Whenever $\hat{a} > a_{pool}$ *then* $\mathcal{M}^\star = \mathcal{M}^\star_{pool}$.
(ii) Whenever $\hat{a} \in (a_{no}, a_{sep})$ *then* $\mathcal{M}^\star = \mathcal{M}^\star_{sep}$.
(iii) Whenever $\hat{a} < a_{no}$ *then* $\mathcal{M}^\star = \mathcal{M}^\star_{no(\underline{a})}$.

Proof. First, we note that as $\hat{a} \to \bar{a}$ then it must be the case that $\mathcal{M}^\star_{pool}$ dominates both $\mathcal{M}^\star_{sep}$ and $\mathcal{M}^\star_{no(\tilde{a})}$ for all $\tilde{a} < \bar{a}$. The latter case is trivial given that as $\hat{a} \to \bar{a}$ the regulator puts approximately probability 1 on the bank's type being $\bar{a}$. In that case it cannot be that $\mathcal{M}^\star_{no(\tilde{a})}$ dominates $\mathcal{M}^\star_{pool}$ for some $\tilde{a} < \bar{a}$ as the capital requirement of $\mathcal{M}^\star_{pool}$ is such that $K^\star \to +\infty$ as $\hat{a} \to \bar{a}$.

To prove that there exists a_{pool} such that whenever $\hat{a} > a_{pool}$ then $\mathcal{M}^\star_{pool}$ dominates $\mathcal{M}^\star_{sep}$ we simply note that this is the case whenever there are only two types $\underline{a}$ and $\bar{a}$ as shown in the proof of Proposition 4.4. Given that satisfying the incentive compatibility conditions for types $a \in (\underline{a}, \bar{a})$ must weakly decrease the optimal capital requirements $K_{\underline{a}}$ and $K_{\bar{a}}$ then it must be the case that if $\mathcal{M}^\star_{pool}$ dominates $\mathcal{M}^\star_{sep}$ assuming only two types the it must further dominate $\mathcal{M}^\star_{sep}$ when there are more types as $\mathcal{M}^\star_{pool}$ sets the same capital requirement when there are two types $\underline{a}$ and $\bar{a}$ as well as when there is a continuum of types $[\underline{a}, \bar{a}]$.

Now to prove that there exists a_{no} such that $\mathcal{M}^\star_{no(\underline{a})}$ dominates both $\mathcal{M}^\star_{sep}$ and $\mathcal{M}^\star_{pool}$ whenever $\hat{a} < a_{no}$ we note that as $\hat{a} \to \underline{a}$ then K is strictly decreasing (assuming that $K < \bar{K}$ for all $\hat{a}$ for some arbitrarily large $\bar{K}$) as

$$\min_{\substack{s \in \mathcal{S} \\ \mathbb{E}_{pool}[s] = K}} \mathbb{E}_{\bar{a}}[s] - \mathbb{E}_{pool}[s]$$

is strictly decreasing in $\hat{a}$. Hence, at some point it must be the case that in the limit $K < \bar{K}$ and therefore $\mathcal{M}^\star_{no(\underline{a})}$, which sets $K^\star_{no} = \bar{K}$, dominates $\mathcal{M}^\star_{no(\underline{a})}$.

Finally given the statement of this proposition, we note that if for all $\hat{a} < a_{pool}$ it is the case that $\mathcal{M}^\star_{sep}$ dominates $\mathcal{M}^\star_{no(\underline{a})}$ then $p_{no} = 0$. Similarly, if $\mathcal{M}^\star_{no(\underline{a})}$ dominates $\mathcal{M}^\star_{sep}$ for all $a < a_{pool}$, then $a_{sep} = a_{no}$. Otherwise, there exists $a_{sep} > a_{no}$ such that $\mathcal{M}^\star_{sep}$ is optimal whenever $\hat{a} \in (a_{sep}, a_{no})$ and $\mathcal{M}^\star_{no(\underline{a})}$ is optimal whenever $\hat{a} < a_{no}$. □

3.9 Appendix

3.1.1 Proofs of Section 2

3.1.1.1 Proof of Lemma 3.8

Proof. First note that it is without loss to assume that inducing investment by the h-type is socially optimal if $\mathcal{M}$ is socially optimal and separating. Namely, if a separating mechanism

induces the h-type to forgo the investment then the equivalent outcome can be implemented by a pooling mechanism that sets capital requirements, security restrictions, and transfers equal to the ℓ-type's in the separating mechanism.

In this case, if the market coordinates on an equilibrium whereby the h-type does not invest because the market believes the h-type will never invest then the regulator can purchase the h-types security s_h at a price $\mathbb{E}_h[s_h]$. The expected gain from purchasing this security is 0 as given that the mechanism is incentive compatible only the h-type will ask the regulator to purchase its security (while meeting the terms of the h-type menu option). In that case the regulator breaks even on the h-type's security but induces the h-type to invest over the alternative equilibrium yielding a strict expected welfare improvement of $p \cdot \hat{b}$.

If instead the optimal mechanism is pooling, then if there exists an equilibrium whereby the h-type invests, it must be the case that investment is optimal for society. Namely, under our assumption $\hat{b} < a_h - a_\ell$ implies that if the market prices the bank's security at the pooling average so that $\mu(s) = p$, then the h-type will forgo the new investment if the pooling capital requirement K is too large. Therefore, if inducing investment by the h-type is not socially optimal then the regulator should increase the capital requirement, contradicting the fact that the mechanism is socially optimal.

Now suppose $\mathcal{M}$ is a socially optimal pooling mechanism and that there exists an equilibrium of $\mathcal{M}$ that induces the h-type to invest whereby both banks issue the same security s. In that case, by purchasing the bank's security at the price $\mathbb{E}_p[s]$ the regulator breaks even on the security. This is due to the fact that when the bank is the ℓ-type the regulator loses $\mathbb{E}_p[s] - \mathbb{E}_\ell[s]$ but when the bank is the h-type the regulator gains $\mathbb{E}_h[s] - \mathbb{E}_p[s]$. Therefore, given that $\mathbb{E}_p[s] = p\mathbb{E}_h[s] + (1-p)\mathbb{E}_\ell[s]$ the regulator breaks even in expectation on the purchase of s. Hence, the regulator obtains a strict expected welfare gain over the equilibrium whereby the h-type doesn't invest equal to $p \cdot \hat{b}$.

Finally, if $\mathcal{M}$ is a socially optimal pooling mechanism with an equilibrium whereby the two bank types issue different securities, then this equilibrium could be implemented through the use of a separating mechanism, in which case we know that agreeing to purchase the bank's security at the appropriate price is strictly welfare improving over any equilibrium whereby the h-type forgoes investment. □

3.1.1.2 Proof of Lemma 3.9

Proof. To prove this we simply use the definitions to obtain

$$\begin{aligned} V_\theta(s,\tilde{\theta};P) &= \mathbb{E}_\theta[max\{a_\theta + B + P - T_{\tilde{\theta}} - s, 0\}] \\ &= \int_{-a_\theta - P + T_{\tilde{\theta}}}^{\infty} (x + a_\theta + P - T_{\tilde{\theta}} - s) dG(x) \\ &= \int_{-a_\theta - P + T_{\tilde{\theta}}}^{\infty} (x - s) dG(x) + (1 - G(-a_\theta - P + T_{\tilde{\theta}}))(a_\theta + P - T_{\tilde{\theta}}) \\ &= \int_{-a_\theta - P + T_{\tilde{\theta}}}^{\infty} x dG(x) - \mathbb{E}_\theta[s] + (1 - G(-a_\theta - P + T_{\tilde{\theta}}))(a_\theta + P - T_{\tilde{\theta}}) \\ &= a_\theta + b_\theta(P - T_{\tilde{\theta}}) + P - \mathbb{E}_\theta[s] - T_{\tilde{\theta}} \end{aligned}$$

Finally, we substitute $K_1 = P - T_{\tilde{\theta}}$ to obtain the result. Note that the fourth equality is valid due to the fact that the bank's limited liability implies that $z - s(z) \geq 0$ for all $z \in \mathbb{R}$ so that $x + a_\theta + P - T_{\tilde{\theta}} - s \geq 0$ if and only if $x + a_\theta + P - T_{\tilde{\theta}} \geq 0$ which is the case whenever $x \geq -a_\theta - P + T_{\tilde{\theta}}$.

To show that $b_\theta(K_1) = \hat{b} + L_\theta(K_1)$ we note that

$$\hat{b} + L_\theta(K_1) = \int_{-\infty}^{\infty} x dG(x) - \int_{-\infty}^{-a_\theta - K_1} (x + a_\theta + K_1) dG(x) = b_\theta(K_1)$$

□

3.1.2 Proofs of Section 3

3.1.2.1 Proof of Lemma 3.10

Proof. The first expression for $\mathbb{E}_\theta[s]$ comes from the fact that if s is admissible, then $s(z) = 0$ whenever $z \leq 0$ and therefore given $z = x + a_\theta + P - T_\theta$ then $s(x + a_\theta + P - T_\theta) = 0$ whenever $x < -a_\theta - P + T_\theta$.

To prove that $\mathbb{E}_h[s] > \mathbb{E}_\ell[s]$ when $T_h = T_\ell = T$ for all admissible securities we note that in this case denoting $K_1 = P - T$ then

$$\mathbb{E}_h[s] = \int_{-a_h - K_1}^{\infty} s(x + a_h + K_1) dG(x) = \int_{-a_\ell - K_1}^{\infty} s(x + a_h + K_1) dG(x) + \int_{-a_h - K_1}^{a_\ell - K_1} s(x + a_h + K_1) dG(x)$$

Now, given that $s(z)$ is monotone in z we know that $s(x + a_h + K_1) \geq s(x + a_\ell + K_1)$ for all $x \geq -a_\ell - K_1$. Thus, if there exists some measurable set $C \subset [-a_\ell - K_1, +\infty)$ such that $s(x + a_h + K_1) > s(x + a_\ell + K_1)$ for all $x \in C$ then

$$\int_{-a_\ell - K_1}^{\infty} s(x + a_h + K_1) dG(x) > \int_{-a_\ell - K_1}^{\infty} s(x + a_\ell + K_1) dG(x) = \mathbb{E}_\ell[s]$$

and therefore $\mathbb{E}_h[s] > \mathbb{E}_\ell[s]$.

Otherwise, $s(x + a_h + K_1) = s(x + a_\ell + K_1)$ for all measurable sets $C \subset [-a_\ell - K_1, +\infty)$. In this case, suppose that $s(z) = s(z') = d > 0$ for all measurable sets $C \subset [z_0, +\infty)$ with

$z, z' > z_0 \geq d$, then

$$\mathbb{E}_h[s] = (1 - G(z_0 - a_h - K_1)) \cdot d + \int_{-a_h - K_1}^{z_0} s(x + a_h + K_1) dG(x)$$

while

$$\mathbb{E}_\ell[s] = (1 - G(z_0 - a_\ell - K_1)) \cdot d + \int_{-a_\ell - K_1}^{z_0} s(x + a_\ell + K_1) dG(x)$$

Now given that

$$\int_{-a_h - K_1}^{z_0} s(x + a_h + K_1) dG(x) = \int_{-a_\ell - K_1}^{z_0} s(x + a_h + K_1) dG(x) + \int_{-a_h - K_1}^{-a_\ell - K_1} s(x + a_h + K_1) dG(x)$$

we can see, again by monotonicity, of s that

$$\int_{-a_\ell - K_1}^{z_0} s(x + a_h + K_1) dG(x) \geq \int_{-a_\ell - K_1}^{z_0} s(x + a_\ell + K_1) dG(x)$$

and given that $G(z_0 - a_\ell - K_1) > G(z_0 - a_h - K_1)$ for all z_0 implies again that $\mathbb{E}_h[s] > \mathbb{E}_\ell[s]$.

Now, based on the above proof the only difference between the type ℓ and type h banks is that $a_h > a_\ell$. In this sense, we could always introduce a third type h' such that $a_{h'} > a_h$ and reproduce the same proof to obtain that $\mathbb{E}_{h'}[s] > \mathbb{E}_h[s] > \mathbb{E}_\ell[s]$. Therefore, it must be the case that $\mathbb{E}_h[s] - \mathbb{E}_\ell[s]$ is strictly increasing in a_h for any fixed security s that generates funds P. □

3.1.2.2 Proof of Lemma 3.11

Proof. We will prove the Proposition in the following steps.

Claim (1): There exist no equilibria of $\Gamma(\mathcal{M})$ *with* $s_\ell = 0$. Denoting by $P_{\mu(s)}$ the market price of security s under beliefs $\mu(s)$ we can see

$$\min_{\mu} \; a_\ell + b_\ell(P_{\mu(s)}) + P_{\mu(s)} - \mathbb{E}_\ell[s] - T > a_\ell$$

coming from the fact that $P_{\mu(s)} - \mathbb{E}_\ell[s] = \mathbb{E}_{\mu(s)}[s] - \mathbb{E}_\ell[s] \geq 0$ for all beliefs μ and $b_\ell(K_1) - T > 0$ for all equilibrium prices $P_{\mu(s)}$ given that $T \leq \min\{b_h(K_1), b_\ell(K_1)\}$ where $K_1 = P_{\mu(s)} - T$. Hence, the ℓ-type always finds it profitable to invest. □

Claim (1) states that all equilibria of $\Gamma(\mathcal{M})$ are such that $s_\ell \neq 0$. Therefore the condition that either $s_h = 0$, $s_h = s_\ell$, or $s_h \neq s_\ell$ is trivial. What is left to prove are the remaining conditions on points (i) and (iii) (the condition on (ii) that $\mathbb{E}_p[s] \geq K + T$ comes from the

sequential rationality of the investment decision).

Claim (2): Any equilibrium of $\Gamma(\mathcal{M})$ *satisfying the intuitive criterion and* $s_h = 0$ *is such that* $\mathbb{E}_\ell[s_\ell] = K+T$. In order to prove this, we first note that in any equilibrium with $s_h = 0$, it must be the case that $\mu(s_\ell) = 0$ and therefore the payoff to the ℓ-type is $a_\ell + b_\ell(\mathbb{E}_\ell[s_\ell] - T)$. Now suppose that in some equilibrium with $s_h = 0$ it is the case that $\mathbb{E}_\ell[s_\ell] > K + T$. Then, given that $b_\ell(K_1)$ is strictly decreasing in K_1 as $\frac{\partial}{\partial K_1} b_\ell(K_1) = -G(-a_\ell - K_1) < 0$, we know that if the ℓ-type issues a security s such that $\mathbb{E}_\ell[s] = K + T$, then it achieves a strictly higher payoff, regardless of the market beliefs (beliefs can only improve the ℓ-types payoff when switching securities). Hence, any sequentially rational strategy s_ℓ must satisfy $\mathbb{E}_\ell[s_\ell] = K + T$.

Claim (3): Any equilibrium with $s_\ell \neq s_h$ *must satisfy* $\mathbb{E}_h[s_h] = K' + T > K + T = \mathbb{E}_\ell[s_\ell]$

$$s_h \in \underset{s' \in \mathcal{S}: \mathbb{E}_h[s'] = K' + T}{\operatorname{argmin}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s'] \quad \text{with } K' \text{ satisfying} \quad b_\ell(K) = b_\ell(K') + \mathbb{E}_h[s_h] - \mathbb{E}_\ell[s_h]$$

In order to prove this, we first note that $s_\ell \neq s_h$ implies $\mu(s_\ell) = 0$ and $\mu(s_h) = 1$. Therefore, as concluded from the proof of the previous claim it must be the case that $\mathbb{E}_\ell[s_\ell] = K + T$. Furthermore, if there is no profitable deviation for the ℓ-type to mimic the h type, then it must be the case that

$$a_\ell + b_\ell(K) \geq a_\ell + b_\ell(K') + \mathbb{E}_h[s_h] - \mathbb{E}_\ell[s_h]$$

which implies

$$b_\ell(K) - b_\ell(K') \geq \mathbb{E}_h[s_h] - \mathbb{E}_\ell[s_h] \tag{3.3}$$

and given that $\mathbb{E}_h[s_h] - \mathbb{E}_\ell[s_h] > 0$ and $b_\ell(K)$ is decreasing in K implies that it must be the case that $K' > K$.

Now, the h-type should always issue a security that minimizes $\mathbb{E}_h[s_h]$ while still satisfying (3.3) given that the intuitive criterion states that any security s_h that satisfies (3.3) must have equilibrium beliefs $\mu(s_h) = 1$. In this case the security that satisfies this condition is the security s_h that minimizes $\mathbb{E}_h[s_h] - \mathbb{E}_\ell[s_h]$ and sets (3.3) to equality. Hence, in any equilibrium with $s_h \neq s_\ell$, we have $\mathbb{E}_\ell[s_\ell] = K + T < K' + T = \mathbb{E}_h[s_h]$,

$$s_h \in \underset{s' \in \mathcal{S}: \mathbb{E}_h[s'] = K' + T}{\operatorname{argmin}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']$$

and K' such that

$$b_\ell(K) = b_\ell(K') + \mathbb{E}_h[s_h] - \mathbb{E}_\ell[s_h]$$

□

3.1.2.3 Proof of Lemma 3.12

Proof. If $\mathcal{M}$ is incentive compatible then it must be the case that the h-type is weakly better off choosing the h-option of the menu and the ℓ-type weakly better off choosing the ℓ-option of the menu. Therefore, given that we always assume without loss that the banks choose their own menu when they are indifferent and that the market correctly believes this, then it must be the case that $\mu(s_h) = 1$ and $\mu(s_\ell) = 0$ for any equilibrium satisfying the intuitive criterion.

To show that capital requirements are binding we note that under an incentive compatible mechanism the bank of type $\theta \in \{h, \ell\}$ receives a payoff of $a_\theta + b_\theta(K_1)$ where $K_1 = P - T \geq K_\theta$ is the capital generated by the sale of the security s_θ. In that case, given that the mechanism is incentive compatible when the capital requirements are binding and the bank's equilibrium payoff is strictly decreasing in the capital generated K_1 (coming from $\frac{\partial}{\partial P} b_\theta(K_1) = -G(-a_\theta - K_1) < 0$), sequential rationality of the bank's strategy implies that it must be the case that both types generate exactly the capital required so that $\mathbb{E}_\theta[s_\theta] = K_\theta + T_\theta$ for each $\theta \in \{h, \ell\}$. □

3.1.3 Proofs of Section 4

3.1.3.1 Proof of Lemma 3.14

Proof. If investment is not socially desirable, then the regulator will optimally set $K = \bar{K}$ and only the ℓ-type bank will invest. Therefore, investment is socially desirable under the pooling equilibrium whenever the welfare of both banks investing with pooling requirement $K^\star$ is greater than the welfare of just the ℓ-type investing with capital requirement $\bar{K}$: $W(K^\star|invest) \geq W(I|forgo)$. Further, using the fact that $\hat{b} = \hat{b}(K^\star) + \hat{L}(K^\star)$ we can see that, after rearranging, $W(K^\star|invest) \geq W(\bar{K}|forgo)$ if and only if $\hat{b} \geq \frac{\lambda}{p}(pL_h(K^\star) - (1-p)(L_\ell(K^\star) - L_\ell(\bar{K}))$. □

3.1.3.2 Proof of Proposition 3.13

Proof. First note that $T_h = T_\ell = 0$ is optimal under any mechanism with $K_\ell = K_h$ given that transfers cancel out in the welfare function and $T_h = T_\ell > 0$ will require lower capital requirements in order to induce the h-type to invest. Hence, optimally $T_h = T_\ell = 0$. Further, we know that under any type-(ii) equilibrium both types issue a security s such

that $\mathbb{E}_p[s] \geq K$. Further, this equilibrium exists only if the h-type bank prefers investment and selling an underpriced security as opposed to forgoing the investment. The first step is to show that the regulator should optimally restrict securities to the set

$$\mathcal{S}' := \{s \in \mathcal{S} : s \in \underset{\mathbb{E}_p[s] \geq K}{\operatorname{argmin}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s]\}$$

In order to prove this, we first note that investment by the h-type is optimal only if

$$a_h + b_h(K') + K' - \mathbb{E}_h[s] \geq a_h$$

where $K' = \mathbb{E}_p[s] \geq K$. This can equivalently be expressed as

$$b_h(K') \geq \mathbb{E}_h[s] - K' = (1-p)(\mathbb{E}_h[s] - \mathbb{E}_\ell[s])$$

Now, given that $b_h(K)$ is decreasing in K, then this expression tells us that it is weakly optimal for the regulator to restrict securities to the set $\mathcal{S}'$. This comes from that fact that if $s \notin \mathcal{S}'$ then the regulator could restrict securities to $\mathcal{S}'$ and strictly increase the capital requirement which weakly improves welfare (strictly if the capital requirement is binding in the pooling equilibrium).

Next we will show that

$$\min_{\substack{s \in \mathcal{S} \\ \mathbb{E}_{pool}[s] = K}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s] \tag{3.4}$$

is increasing in K. In order to do so, consider $K' > K$ and denote by $s, s' \in \mathcal{S}$ two securities satisfying

$$s \in \underset{\substack{s \in \mathcal{S} \\ \mathbb{E}_{pool}[s] = K}}{\operatorname{argmin}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s] \quad \text{and} \quad s' \in \underset{\substack{s \in \mathcal{S} \\ \mathbb{E}_{pool}[s] = K'}}{\operatorname{argmin}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s].$$

We claim that no matter the values of K' and K, as long as $K' > K$, then $\mathbb{E}_h[s'] - \mathbb{E}_\ell[s'] > \mathbb{E}_h[s] - \mathbb{E}_\ell[s]$. To prove this, we simply note that there exists $\phi \in (0,1)$ such that if we let $\tilde{s} = \phi s'$, then $\mathbb{E}_{pool}[\tilde{s}] = K$ and

$$\mathbb{E}_h[s'] - \mathbb{E}_\ell[s'] > \phi(\mathbb{E}_h[s'] - \mathbb{E}_\ell[s']) = \mathbb{E}_h[\tilde{s}] - \mathbb{E}_\ell[\tilde{s}] \geq \mathbb{E}_h[s] - \mathbb{E}_\ell[s]$$

where the last inequality comes from the fact that s minimizes $\mathbb{E}_h[s] - \mathbb{E}_\ell[s]$ among all securities such that $\mathbb{E}_{pool}[s] = K$. Hence, we have proven our claim that (3.4) is strictly increasing in K.

Now, the regulator would like to increase K as large as possible just until the h-type bank is indifferent between investing or not. If there exists a pooling equilibrium under the

capital requirement K such that $\mathbb{E}_p[s] = K' > K$ then the regulators mechanism cannot be optimal as it implies that

$$b_h(K) > (1-p) \min_{\substack{s \in \mathcal{S} \\ \mathbb{E}_p[s]=K}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s]$$

which implies that there exist equilibria where the banks raise $K < K'$ which is strictly worse than the equilibrium whereby the banks raise exactly K'. If instead the capital requirement K is set so that

$$b_h(K) = (1-p) \min_{\substack{s \in \mathcal{S} \\ \mathbb{E}_p[s]=K}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s]$$

then the bank and the market are always guaranteed to coordinate on the (unique) pooling equilibrium that generates the highest possible level of capital. □

3.1.3.3 Proof of Lemma 3.15

Proof. If investment by the h-type is not socially desirable, then the regulator will optimally set $K = \bar{K}$ and only the ℓ-type bank will invest. Therefore, investment is socially desirable under the pooling equilibrium whenever the welfare of both banks investing with separating requirements $K_h^\star$ and $K_\ell^\star$ is greater than the welfare of just the ℓ-type investing with capital requirement $\bar{K}$: $W(K_\ell^\star, K_h^\star | invest) \geq W(I | forgo)$. This is the case whenever

$$\hat{b} - \lambda(p \cdot L_h(K_h^\star) + (1-p) \cdot L_\ell(K_\ell^\star)) \geq (1-p)(\hat{b} - \lambda \cdot L_\ell(\bar{K}))$$

and after rearranging we obtain our result. □

3.1.3.4 Proof of Proposition 3.16

Proof. First assume that the first best $K_\ell = K_h = \bar{K}$ is not possible under any separating mechanism. This implies that for any separating mechanism, one of the incentive compatibility constraints is binding. Now, let $\mathcal{M} = \{K_\theta, T_\theta, S_\theta\}_{\theta \in \{h,\ell\}}$ be a mechanism with $\tilde{s} \in S_h$ such that

$$\mathbb{E}_h[\tilde{s}] - \mathbb{E}_\ell[\tilde{s}] > \min_{\substack{s' \in \mathcal{S} \\ \mathbb{E}_h[s'] \geq K_h + T_h}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']$$

We claim that $\mathcal{M}$ is weakly dominated by a mechanism $\mathcal{M}' = \{K'_\theta, T'_\theta, S'_\theta\}_{\theta\in\{h,\ell\}}$ that sets

$$\mathcal{S}'_h = \{s \in \mathcal{S} : s \in \underset{\substack{s'\in\mathcal{S}\\ \mathbb{E}_h[s']\geq K_h+T_h}}{\operatorname{argmin}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']\}.$$

In order to prove this claim, suppose the IC_ℓ constraint is binding under $\mathcal{M}$ and assume $\tilde{s} \notin \mathcal{S}'_h$. Then, given that IC_ℓ must hold for all $s \in \mathcal{S}_h$ implies that

$$T_h - T_\ell = b_\ell(K_h) - b_\ell(K_\ell) + \mathbb{E}_h[\tilde{s}] - \mathbb{E}_\ell[\tilde{s}] > b_\ell(K_h) - b_\ell(K_\ell) + \min_{\substack{s'\in\mathcal{S}\\ \mathbb{E}_h[s']\geq K_h}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s'] \quad (3.5)$$

Now, if IC_h is not binding, then there exists $\hat{K} > K_\ell$ such that $\mathcal{M}'$ is incentive compatible with $K'_\ell = \hat{K}$ and $K'_h = K_h$ implying that $\mathcal{M}'$ strictly dominates $\mathcal{M}$. To show that this is the case, we simply note that the strict inequality of (3.5) implies that one can increase K_ℓ by a small amount ϵ without violating the incentive compatibility constraint IC_ℓ whenever the regulator restricts securities to $\mathcal{S}'_h$. Further, given that IC_h is not binding, we can always find an $\bar{\epsilon}$ such that for all $\epsilon < \bar{\epsilon}$ setting $K'_\ell = K_\ell + \epsilon$ produces a mechanism that satisfies both incentive compatibility constraints. If instead, IC_h is also binding so that we cannot make such a welfare improvement then $\mathcal{M}'$ can achieve the same welfare as $\mathcal{M}$ even when restricting securities to the set $\mathcal{S}'_h$ by setting $K'_\theta = K_\theta$ and $T'_\theta = T_\theta$ for each $\theta \in \{h, \ell\}$ as this restriction only relaxes IC_ℓ.

Next consider the case where IC_ℓ is not binding. In this case IC_h should be binding otherwise $\mathcal{M}$ is not optimal. Further, given that IC_h is binding and is independent of the security s_h, then $\mathcal{M}'$ generates the same welfare as $\mathcal{M}$ whenever the capital requirements and transfers are set equal. Therefore, we have shown that restricting securities to $\mathcal{S}'_h$ weakly improves welfare.

To conclude the first part of the proof we will show that restricting to $\mathcal{S}^\star_h$ over $\mathcal{S}'_h$ is also without loss. Namely, we will show that

$$\min_{\substack{s'\in\mathcal{S}\\ \mathbb{E}_h[s']\geq K_h+T_h}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s'] = \min_{\substack{s'\in\mathcal{S}\\ \mathbb{E}_h[s']= K_h+T_h}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']$$

To do so we will show that for any s such that $\mathbb{E}_h[s] > K + T$ there exists s' such that $\mathbb{E}_h[s'] = K + T$ and $\mathbb{E}_h[s'] - \mathbb{E}_h[s'] < \mathbb{E}_h[s] - \mathbb{E}_h[s]$. Namely, consider $s'(z) = \phi \cdot s(z)$ for all $z \in \mathbb{R}$. Then, letting $\phi = \frac{K+T}{\mathbb{E}_h[s]} < 1$ we can see that $\mathbb{E}_h[s'] = \mathbb{E}_h[\phi \cdot s] = \phi \cdot \mathbb{E}_h[s] = K + T$. Further, this implies that

$$\mathbb{E}_h[s'] - \mathbb{E}_h[s'] = \phi \cdot (\mathbb{E}_h[s] - \mathbb{E}_h[s]) < \mathbb{E}_h[s] - \mathbb{E}_h[s]$$

and we have proven our claim.

We will now prove that it is weakly optimal for the regulator to restrict the ℓ-type to issue securities in the set

$$S'_\ell = \{s \in \mathcal{S} : s \in \underset{\substack{s' \in \mathcal{S} \\ \mathbb{E}_\ell[s'] \geq K_\ell + T_\ell}}{\operatorname{argmax}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']\}$$

suppose in a similar vein that $\mathcal{M}$ is a mechanism such that there exists $\tilde{s} \in S_\ell$ with

$$\mathbb{E}_h[\tilde{s}] - \mathbb{E}_\ell[\tilde{s}] < \max_{\substack{s' \in \mathcal{S} \\ \mathbb{E}_h[s'] \geq K_\ell + T_\ell}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']$$

Now, if IC_h is binding then

$$T_h - T_\ell = b_h(K_h) - b_h(K_\ell) + \mathbb{E}_h[\tilde{s}] - \mathbb{E}_\ell[\tilde{s}] < b_h(K_h) - b_h(K_\ell) + \max_{\substack{s' \in \mathcal{S} \\ \mathbb{E}_h[s'] \geq K_\ell + T_\ell}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s'].$$

Therefore, if IC_ℓ is not binding then the regulator can strictly increase K_h by a positive amount when restricting securities to $\mathcal{S}'_\ell$. If instead IC_ℓ is binding, making such a restriction yields the same welfare when setting the same capital requirements and transfers as $\mathcal{M}$. Finally, we note that when choosing from a security $s \in \mathcal{S}_\ell$, the ℓ-type bank will optimally choose a security such that $\mathbb{E}_\ell[s_\ell] = K_\ell + T_\ell$ in any equilibrium and therefore it is without loss to restrict

$$S_\ell = \{s \in \mathcal{S} : s \in \underset{\substack{s' \in \mathcal{S} \\ \mathbb{E}_\ell[s'] = K_\ell + T_\ell}}{\operatorname{argmax}} \mathbb{E}_h[s'] - \mathbb{E}_\ell[s']\}$$

In order to conclude the proof we note that we have just shown that restricting securities to $\mathcal{S}^\star_\theta$ can only improve welfare under the optimal capital requirements and transfers. Therefore, if there exists a separating mechanism that achieves the first best (so that no incentive constraints are binding) then it is without loss to restrict the securities of that mechanism to $\mathcal{S}^\star_\theta$. □

3.1.3.5 Proof of Corollary 3.17

Proof. In order to prove this first note that if s_{AS} represents the sale of the existing asset, then

$$\mathbb{E}_h[s_{AS}] - \mathbb{E}_\ell[s_{AS}] = a_h - a_\ell$$

This is obvious if the necessary funds to meet the capital requirement are less than a_ℓ. If

instead $K_\ell + T_\ell > a_\ell$ then

$$\mathbb{E}_\theta[s_{AS}] = a_\theta + \eta \cdot b_0(K_\ell)$$

where $b_0(K_\ell) = \int_{-K_\ell}^{\infty} xf(x)dx$ and η is chosen to satisfy $\mathbb{E}_\ell[s_{AS}] = a_\ell + \eta \cdot b_0(K_\ell) = K_\ell + T_\ell$. This expression comes from the fact that given that the only uncertainty is regarding the existing assets, then once those assets are sold the value of any claims on the bank are independent of θ.

Now, what is left to prove is that for all $s \in \mathcal{S}$, $\mathbb{E}_h[s] - \mathbb{E}_\ell[s] \leq a_h - a_\ell$. In order to show this, consider a security s such that

$$\mathbb{E}_h[s] - \mathbb{E}_\ell[s] > a_h - a_\ell.$$

If this is the case, then it must be that there exists some set X with positive measure under G such that $x \in X$ implies

$$s(x + a_h + K_\ell) - s(x + a_\ell + K_\ell) > a_h - a_\ell \tag{3.6}$$

for all $a_h > a_\ell$. Otherwise it would be the case that

$$\mathbb{E}_h[s] - \mathbb{E}_\ell[s] = \int_{-a_h - K_\ell}^{+\infty} (s(x + a_h + K_\ell) - s(x + a_\ell + K_\ell))dG(x) < a_h - a_\ell$$

where the first equality comes from the fact that if $x \in [-a_h - K_\ell, -a_\ell - K_\ell]$ then $s(x + a_\ell + K_\ell) = 0$.

Now, we can rewrite equation 3.6 as

$$\frac{s(x + a_\ell + (a_h - a_\ell) + K_\ell) - s(x + a_\ell + K_\ell)}{a_h - a_\ell} > 1$$

and taking the limit as $a_h - a_\ell \to 0$ implies that

$$s'(x + a_\ell + K_\ell) > 1$$

which implies that $s \notin \mathcal{S}$ as it violates Condition (2) of Definition 3.1. Hence, we have shown that s_{AS} maximizes $\mathbb{E}_h[s] - \mathbb{E}_\ell[s]$ among all $s \in \mathcal{S}$. □

3.1.3.6 Proof of Lemma 3.18

Proof. The conditions given in the lemma are precisely the conditions necessary to achieve

incentive compatibility when $K_\ell = K_h = \bar{K}$. In this case, incentive compatibility, conditional on choosing the optimal securities, is achieved if and only if

$$\max_{\substack{s\in\mathcal{S}\\ \mathbb{E}_\ell[s]=\bar{K}+T_\ell}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s] \geq T_h - T_\ell \geq \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_h[s]=\bar{K}+T_h}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s]$$

but inducing investment requires $T_h < b_h(\bar{K}) = \hat{b}$. Therefore, whenever

$$\hat{b} \geq \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_h[s]=\bar{K}+T_h}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s]$$

the regulator can achieve the first best by setting $T_\ell = 0$ and $T_h = T \leq \hat{b}$ achieving incentive compatibility of the first best level of capital requirements $K_\ell = K_h = \bar{K}$.

For the second part of the proof we note that there always exists $\hat{p}$ such that

$$(1-\hat{p}) \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_{pool}[s]=\bar{K}}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s] = \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_h[s]=\bar{K}+T_h}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s] \leq \hat{b}$$

which implies that the regulator can implement the first best through the optimal pooling mechanism whenever $p > \hat{p}$. □

3.1.3.7 Proof of Lemma 3.19

Proof. First, suppose that $K_\ell > K_h$. We will show that when $\mathcal{M}^\star_{sep}$ dominates $\mathcal{M}^\star_{pool}$ then no matter the choice of $K_\ell > K_h$ the constraint IC_h is never binding and therefore it must be the case that IC_ℓ is binding. To do so, note that based on the characterization of the optimal pooling capital requirement $K^\star$, we note that if $\mathcal{M}_{sep}$ dominates $\mathcal{M}_{pool}$ then it must be the case that $K_\ell > K^\star$. This implies that

$$b_h(K_\ell) < (1-p) \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_{pool}[s]=K_\ell}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s] < \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_{pool}[s]=K_\ell+T_\ell}} \mathbb{E}_h[s]-\mathbb{E}_\ell[s]$$

where the second inequality comes from the fact that securities that generate more funds always increase the minimum information sensitivity (see the proof of Proposition 3.16) and we can drop the $(1-p)$ as these values are strictly positive. This implies that

$$b_h(K_h) - b_h(K_\ell) + \max_{\substack{s_\ell\in\mathcal{S}\\ \mathbb{E}_\ell[s_\ell]=K_\ell+T_\ell}} (\mathbb{E}_h[s_\ell]-\mathbb{E}_\ell[s_\ell]) >$$

$$b_h(K_h) - \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_{pool}[s]=K_\ell+T_\ell}} (\mathbb{E}_h[s]-\mathbb{E}_\ell[s]) + \max_{\substack{s_\ell\in\mathcal{S}\\ \mathbb{E}_\ell[s_\ell]=K_\ell}} (\mathbb{E}_h[s_\ell]-\mathbb{E}_\ell[s_\ell])$$

Further, given that $T_h - T_\ell < b_h(K_h)$ (coming from the fact that the bank will optimally forgo investment if $T_h > b_h(K_h)$) implies that IC_h is never binding whenever

$$\min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_{pool}[s]=K_\ell+T_\ell}} (\mathbb{E}_h[s]-\mathbb{E}_\ell[s]) < \max_{\substack{s_\ell\in\mathcal{S}\\ \mathbb{E}_\ell[s_\ell]=K_\ell+T_\ell}} (\mathbb{E}_h[s_\ell]-\mathbb{E}_\ell[s_\ell])$$

We will show that this inequality holds when s_ℓ is equity. Namely denoting by s^1_{eq} the equity security satisfying $\mathbb{E}_{pool}[s^1_{eq}] = K_\ell + T_\ell$ then denoting by $V_\theta(K_\ell)$ the value of the firm type θ after issuing equity worth $K_\ell + T_\ell$ and making the investment we obtain

$$\min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_{pool}[s]=K_\ell+T_\ell}} (\mathbb{E}_h[s]-\mathbb{E}_\ell[s]) \leq \mathbb{E}_h[s^1_{eq}] - \mathbb{E}_\ell[s^1_{eq}] = \frac{V_h(K_\ell)-V_\ell(K_\ell)}{\hat{V}(K_\ell)}K_\ell$$

where $\hat{V}(K_\ell) = pV_h(K_\ell)+(1-p)V_\ell(K_\ell)$. Now if s^2_{eq} is the equity security satisfying $\mathbb{E}_\ell[s^2_{eq}] = K_\ell + T_\ell$ then

$$\max_{\substack{s_\ell\in\mathcal{S}\\ \mathbb{E}_\ell[s_\ell]=K_\ell+T_\ell}} (\mathbb{E}_h[s_\ell]-\mathbb{E}_\ell[s_\ell]) \geq \mathbb{E}_h[s^2_{eq}] - \mathbb{E}_\ell[s^2_{eq}] = \frac{V_h(K_\ell)-V_\ell(K_\ell)}{V_\ell(K_\ell)}K_\ell$$

and therefore, given that $\hat{V}(K_\ell) > V_\ell(K_\ell)$ implies

$$\frac{V_h(K_\ell)-V_\ell(K_\ell)}{\hat{V}(K_\ell)}K_\ell < \frac{V_h(K_\ell)-V_\ell(K_\ell)}{V_\ell(K_\ell)}K_\ell$$

and we obtain our result.

Now we turn to the case where $\mathcal{M}^\star_{sep}$ sets $K_h > K_\ell$. In this case, suppose by contraposition that IC_ℓ is not binding. Then, it must be the case that IC_h is binding so that

$$T_h - T_\ell = b_h(K_h) - b_h(K_\ell) + \max_{\substack{s\in\mathcal{S}\\ \mathbb{E}_\ell[s]=K_\ell+T_\ell}} \mathbb{E}_h[s] - \mathbb{E}_h[s].$$

Now let

$$s_\ell \in \operatorname*{argmax}_{\substack{s\in\mathcal{S}\\ \mathbb{E}_\ell[s]=K_\ell+T_\ell}} \mathbb{E}_h[s] - \mathbb{E}_h[s]$$

be the chosen security of the ℓ-type. Then,

$$\frac{\partial}{\partial K_\ell}[b_h(K_h) - b_h(K_\ell) + \mathbb{E}_h[s_\ell] - K_\ell] = G(-a_h - K_\ell) + \frac{\partial}{\partial K_\ell}(\mathbb{E}_h[s_\ell] - K_\ell))$$

implies that if $\mathbb{E}_h[s_\ell] - \mathbb{E}_\ell[s_\ell]$ is strictly increasing in K_ℓ then the regulator could do strictly better by increasing K_ℓ by some positive amount without violating IC_h and therefore IC_ℓ which is assumed to be non-binding. In order to prove this, we will show that for any value of $K'_\ell > K_\ell$, if

$$s_\ell \in \underset{\substack{s\in\mathcal{S}\\ \mathbb{E}_\ell[s]=K_\ell+T_\ell}}{\operatorname{argmax}} \ \mathbb{E}_h[s] - \mathbb{E}_h[s] \qquad \text{and} \qquad s'_\ell \in \underset{\substack{s\in\mathcal{S}\\ \mathbb{E}_\ell[s]=K'_\ell+T_\ell}}{\operatorname{argmax}} \ \mathbb{E}_h[s] - \mathbb{E}_h[s]$$

then $\mathbb{E}_h[s_\ell] - \mathbb{E}_\ell[s_\ell] < \mathbb{E}_h[s'_\ell] - \mathbb{E}_\ell[s'_\ell]$.

In order to prove this, simply note that s'_ℓ can always be constructed as $s'_\ell(z) = s_\ell(z) + s_0(z)$ for some appropriately constructed $s_0(z)$ such that $s_0(z) \in [0, \max\{z - s_\ell(z), 0\}]$ and $s_0(z) \in \mathcal{S}$. Namely, s'_ℓ pays the same as s_ℓ plus an additional residual s_0 which in expectation is worth $K'_\ell - K_\ell$. In that case given that $\mathbb{E}_h[s] - \mathbb{E}_\ell[s] > 0$ for all $s \in \mathcal{S}$ implies

$$\max_{\substack{s\in\mathcal{S}\\ \mathbb{E}_\ell[s]=K'_\ell+T_\ell}} \ \mathbb{E}_h[s] - \mathbb{E}_h[s] \geq \mathbb{E}_h[s'_\ell] - \mathbb{E}_\ell[s'_\ell] = \mathbb{E}_h[s_\ell] - \mathbb{E}_\ell[s_\ell] + \mathbb{E}_h[s_0] - \mathbb{E}_\ell[s_0] > \mathbb{E}_h[s_\ell] - \mathbb{E}_\ell[s_\ell]$$

and we have proven our claim.

Therefore, we have just shown that if IC_h is binding and IC_ℓ is not binding, then the regulator can increase $K^\star_\ell$ by a small amount increasing the RHS of IC_h. If this increase in $K^\star_\ell$ increases the RHS of IC_h by more than it increases the RHS of IC_ℓ then the regulator would increase $K^\star_\ell$ until $K^\star_\ell \geq K^\star_h$ in which case we are no longer in this case. Otherwise, the regulator will increase $K^\star_\ell$ until IC_ℓ is binding.

□

3.1.3.8 Proof of Proposition 3.20

Proof. By Lemma 3.19 we know that the constraint IC_h is never binding whenever the separating mechanism is optimal and chooses $K^\star_\ell > K^\star_h$. Hence, the relevant binding constraint is IC_ℓ

$$T_h - T_\ell = b_\ell(K_h) - b_\ell(K_\ell) + \min_{\substack{s_h\in\mathcal{S}\\ \mathbb{E}_h[s_h]=K_h+T_h}} \ \mathbb{E}_h[s_h] - \mathbb{E}_\ell[s_h]$$

Now, given that no term on the RHS of IC_ℓ depends on T_ℓ implies that optimally $T_\ell = 0$ whenever the optimal separating equilibrium dominates the optimal pooling mechanism. Namely, if $T_\ell > 0$ then the regulator can decrease T_ℓ which relaxes the IC_ℓ constraint allowing for an increase in capital requirements. If IC_h binds in this case then either we contradict the fact that the optimal separating equilibrium dominates the optimal pooling

mechanism or the fact that $K_\ell^\star > K_h^\star$.

In order to prove that optimally $T_h = b_h(K_h^\star)$ we note that if IC_ℓ is binding with $T_h < b_h(K_h^\star)$, then it is not binding when $T_h = b_h(K_h^\star)$. In order to prove this we note that as T_h increases, $\mathbb{E}_h[s_h]$ increases identically no matter the security chosen as $\mathbb{E}_h[s_h] = K_h + T_h$. This implies that by increasing T_h the security s_h must yield a higher expected payment and therefore $\mathbb{E}_\ell[s_h]$ must weakly increase in value (e.g. if s_h is equity or standard debt then $\mathbb{E}_\ell[s_h]$ will strictly increase in value as T_h increases). If $\mathbb{E}_\ell[s_h]$ strictly increases in value then the increase in T_h is larger than the increase in the RHS of IC_ℓ and therefore setting $T_h = b_h(K_h^\star)$ is strictly optimal as it relaxes the IC_ℓ constraint. Otherwise setting $T_h = b_h(K_h^\star)$ it is weakly optimal. While it can be shown that the security that minimizes the information sensitivity will always yield a strict increase in $\mathbb{E}_\ell[s_h]$ we exclude the proof as this shorter proof suffices.

Now, if instead the optimal mechanism sets $K_h^\star > K_\ell^\star$, then we know that IC_ℓ is still binding and equal to $T_h - T_\ell$. Therefore, $T_h - T_\ell$ should be chosen so that $K_\ell^\star$ and $K_h^\star$ maximize the welfare given investment which is equivalent to minimizing the expected liability subject to incentive compatibility, $T_h \in [0, b_h(K_h)]$, and $T_\ell \in [0, b_\ell(K_\ell)]$. □

3.1.3.9 Proof of Lemma 3.21

Proof. Note that the regulator's objective is to maximize welfare which, when the mechanism is incentive compatible, is equivalent to minimizing the expected liability given by $p \cdot L_h(K_h) + (1-p) \cdot L_\ell(K_\ell)$. Therefore, there always exists a value of p_1 large such that when $p > p_1$ the increase in welfare from increasing K_h by any amount Δ is larger than the decrease in welfare from decreasing K_ℓ to zero. Similarly, there exists p_2 such that when $p < p_2$ then the benefit of increasing K_ℓ by any amount Δ is larger than the decrease in welfare from decreasing K_h to zero. Further, optimally, K_ℓ is weakly decreasing in p while K_h is weakly increasing. Therefore, there are three cases: (i) $K_\ell^\star > K_h^\star$ for all $p > p_2$, in which case $\bar{p} = 1$, (ii) $K_h^\star > K_\ell^\star$ for all $p < p_1$ in which case $\bar{p} = 0$, and (iii) there exists $p_0 \in (p_2, p_1)$ such that $K_\ell^\star > K_h^\star$ whenever $p < p_0$ and $K_h^\star > K_\ell^\star$ when $p > p_0$ in which case $\bar{p} = p_0$.

To show that $\bar{p}$ is strictly increasing in a_h we simply note that the marginal benefit of increasing K_h is given by $p \cdot G(-a_h - K_h)$ which goes to zero as $a_h \to +\infty$. Therefore, as the marginal benefit of increasing K_h decreases, p must be larger to induce the regulator to set $K_h > K_\ell$. □

3.1.4 Proofs of Section 5

3.1.4.1 Proof of Proposition 4.4

Proof. We will proceed to prove this proposition in steps.

Claim 3.27. *There exists p_{sep} such that whenever $p < p_{sep}$ then $\mathcal{M}^\star_{sep}$ dominates $\mathcal{M}^\star_{pool}$.*

Proof. Denote by s_{pool} and s_{sep} the securities such that

$$s_{pool}(K) \in \underset{\substack{s\in\mathcal{S}\\ \mathbb{E}_{pool}[s]=K}}{\operatorname{argmin}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s] \quad \text{and} \quad s_{sep}(K) \in \underset{\substack{s\in\mathcal{S}\\ \mathbb{E}_h[s]=K+T_h}}{\operatorname{argmin}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s]$$

then, we know that the capital requirement $K^\star$ of $\mathcal{M}^\star_{pool}$ is chosen to solve

$$b_h(K^\star) = (1-p)\cdot(\mathbb{E}_h[s_{pool}(K^\star)] - \mathbb{E}_\ell[s_{pool}(K^\star)])$$

but from the definition of s_{pool} we know that $p\cdot\mathbb{E}_h[s_{pool}(K^\star)] + (1-p)\cdot\mathbb{E}_\ell[s_{pool}(K^\star)] = K^\star$ which implies that

$$\mathbb{E}_h[s_{pool}(K^\star)] - \mathbb{E}_\ell[s_{pool}(K^\star)] = \frac{1}{p}(K^\star - \mathbb{E}_\ell[s_{pool}(K^\star)])$$

and therefore, $K^\star$ is chosen to solve

$$b_h(K^\star) = \frac{1-p}{p}(K^\star - \mathbb{E}_\ell[s_{pool}(K^\star)])$$

which shows that $K^\star$ is increasing in p. Further, as $p \to 0$ it must be the case that $K^\star - \mathbb{E}_\ell[s_{pool}(K^\star)] \to 0$ given that $b_h(K^\star)$ is positive and strictly greater than 0 for all $K^\star$. Further, $K^\star - \mathbb{E}_\ell[s_{pool}(K^\star)] \to 0$ only if $K^\star \to 0$ coming from the fact that $\mathbb{E}_h[s_{pool}(K)] - \mathbb{E}_\ell[s_{pool}(K)] > 0$ and strictly increasing in $K^\star$ for all $s \in \mathcal{S}$ (see the proof of Proposition 3.13).

Now, we know that there exists $\bar{p}$ such that whenever $p < \bar{p}$ then optimally $K^\star_\ell > K^\star_h$ and therefore $K^\star_\ell$ and $K^\star_h$ must satisfy

$$b_h(K^\star_h) = b_\ell(K^\star_h) - b_\ell(K^\star_\ell) + \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_h[s]=K^\star_h+T^\star_h}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s]$$

Hence, we simply note that if we set $K^\star_\ell = K^\star_h = 0$ then we obtain

$$b_h(0) > \min_{\substack{s\in\mathcal{S}\\ \mathbb{E}_h[s]=b_h(0)}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s]$$

where the inequality comes from the fact that the security that minimizes the information sensitivity satisfies $\mathbb{E}_h[s_h] = b_h(0)$ and therefore as long as $\mathbb{E}_\ell[s_h] > 0$ then we obtain our result. Further, once this result holds we know that $K_\ell > 0$ and $K_h = 0$ is incentive compatible and therefore there exists p_{sep} such that whenever $p < p_{sep}$, $\mathcal{M}^\star_{sep}$ dominates $\mathcal{M}^\star_{pool}$.

To prove that $\mathbb{E}_\ell[s_h] > 0$ whenever $\mathbb{E}_h[s_h] = b_h(0)$, we note that the only case where $\mathbb{E}_h[s_h] = b_h(0)$ and $\mathbb{E}_\ell[s_h] = 0$ is if s_h is such that $s_h(z) > 0$ if and only if $z \leq (a_h - a_\ell)$: the ℓ-type pays 0 under s_h whenever the h-type pays a positive amount under s_h. But in that case, this implies that $s_h(z) = 0$ for large values of z and $s_h(z) > 0$ for small values of z (this must be the case as $\mathbb{E}_h[s_h] > 0$) contradicting the fact that $s(z)$ is non-decreasing in z.

□

Claim 3.28. *There exists p_{pool} such that whenever $p > p_{pool}$ then $\mathcal{M}^\star_{pool}$ dominates $\mathcal{M}^\star_{sep}$.*

Proof. In order to prove this claim, we note that as $p \to 1$ then $K^\star \to \bar{K}$. Namely, denoting by $s_e(K^\star)$ the equity security such that $\mathbb{E}_{pool}[s_e(K^\star)] = K^\star$, then we know

$$\frac{1-p}{p}(K^\star - \mathbb{E}_\ell[s_{pool}(K)]) \leq \frac{1-p}{p}(K^\star - \mathbb{E}_\ell[s_e(K)]) =$$

$$\frac{1-p}{p} \cdot K^\star(1 - \frac{a_\ell + b_\ell(K^\star) + K^\star}{p \cdot (a_h + b_h(K^\star)) + (1-p)(a_\ell + b_\ell(K^\star)) + K}) =$$

$$\frac{1-p}{p} \cdot K^\star \cdot \frac{p \cdot (a_h + b_h(K^\star) - a_\ell - b_\ell(K^\star))}{p \cdot (a_h + b_h(K^\star)) + (1-p)(a_\ell + b_\ell(K^\star)) + K} \leq (1-p) \cdot (a_h - a_\ell)$$

where the last inequality comes from the fact that the LHS is strictly increasing in $K^\star$ and the RHS is obtained by taking the limit as $K^\star \to +\infty$. Therefore, as $p \to 1$ we know that

$$\frac{1-p}{p}(K^\star - \mathbb{E}_\ell[s_{pool}(K)]) \to 0$$

and therefore it must be the case that $K^\star \to \bar{K}$.

Given this, the only way $\mathcal{M}^\star_{pool}$ does not dominate $\mathcal{M}^\star_{sep}$ as $p \to 1$ is if both $K^\star_\ell \to \bar{K}$ and $K^\star_h \to \bar{K}$ as $p \to 1$. Now, we always know that for any values of $K^\star_h$ and $K^\star_\ell$ it must be the case that IC_ℓ is binding. Therefore, using the fact that $b_h(K_h) \geq T_h - T_\ell$ we can see that under the optimal separating mechanism

$$b_h(K^\star_h) \geq T_h - T_\ell = b_\ell(K^\star_h) - b_\ell(K^\star_\ell) + \min_{\substack{s \in \mathcal{S} \\ \mathbb{E}_h[s] = K^\star_h + T^\star_h}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s]$$

Further, as $p \to 1$ we know that $K^\star_h > K^\star_\ell$. Therefore, if $\mathcal{M}^\star_{sep}$ dominates $\mathcal{M}^\star_{pool}$ then it must

be the case that $K_h^\star > K_{pool}^\star$ which implies

$$b_h(K_h^\star) < (1-p) \min_{\substack{s\in\mathcal{S} \\ \mathbb{E}_{pool}[s]=K_h^\star}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s] \leq (1-p)(a_h - a_\ell)$$

which implies that if $\mathcal{M}_{sep}^\star$ dominates $\mathcal{M}_{pool}^\star$ for all $p > \bar{p}$ then

$$(1-p)(a_h - a_\ell) > b_h(\bar{K}) \geq b_\ell(\bar{K}) - b_\ell(K_\ell^\star) + \min_{\substack{s\in\mathcal{S} \\ \mathbb{E}_h[s]=\bar{K}+T_h^\star}} \mathbb{E}_h[s] - \mathbb{E}_\ell[s].$$

Which can only be the case if $K_\ell^\star < \bar{K}$ as the information sensitivity is always positive and therefore as p approaches 1 this inequality can only be satisfied if $b_\ell(K_\ell^\star) - b_\ell(\bar{K}) > 0$ which implies $K_\ell^\star < \bar{K}$. Hence, there must exist a level p_{pool} such that whenever $p > p_{pool}$, $\mathcal{M}_{pool}^\star$ dominates $\mathcal{M}_{sep}^\star$. □

Claim 3.29. *There exists $p_{und} \in [0,1)$ such that whenever $p < p_{und}$ then $\mathcal{M}_{und}^\star$ dominates both $\mathcal{M}_{pool}^\star$ and $\mathcal{M}_{sep}^\star$.*

Proof. By lemma's 3.14 and 3.15 we know that $\mathcal{M}_{und}^\star$ dominates both $\mathcal{M}_{pool}^\star$ and $\mathcal{M}_{sep}^\star$ whenever

$$b < \min\{\frac{\lambda}{p}(p \cdot L_h(K^\star) + (1-p)(L_\ell(K^\star) - L_\ell(\bar{K})), \frac{\lambda}{p}(p \cdot L_h(K_h) + (1-p)(L_\ell(K_\ell) + L_\ell(\bar{K}))\}$$

further, we know that $K^\star \to 0$ as $p \to 0$ and therefore

$$\frac{\lambda}{p}(p \cdot L_h(K^\star) + (1-p)(L_\ell(K^\star) - L_\ell(\bar{K})) \to +\infty$$

Thus, there exists $p_{und} > 0$ such that $\mathcal{M}_{und}^\star$ dominates $\mathcal{M}_{pool}^\star$ whenever $p < p_{und}$.

Finally, the only way that $\mathcal{M}_{und}^\star$ dominates $\mathcal{M}_{sep}^\star$ is if $K_\ell^\star < \bar{K}$ for all p close to zero. This is the case if and only if

$$b_h(0) < b_\ell(0) - b_\ell(\bar{K})$$

or equivalently

$$b_\ell(\bar{K}) > b_\ell(0) - b_h(0)$$

Namely, this condition states that $K_\ell^\star = \bar{K}$ and $K_h^\star = 0$ is not incentive compatible. When it is satisfied then there exists $p_{und} > 0$ such that $p < p_{und}$ implies $\mathcal{M}_{und}^\star$ dominates $\mathcal{M}_{sep}^\star$. Whenever this condition is not satisfied then it implies that setting $K_\ell^\star = \bar{K}$ and $K_h^\star = 0$ is incentive compatible for all $p > 0$ and therefore $\mathcal{M}_{sep}^\star$ weakly dominates $\mathcal{M}_{und}^\star$ for all $p > 0$ in which case $p_{und} = 0$. □

Now, note that Claim 3.28 and Claim 3.29 imply together that there exists p_{und} and p_{sep} such that $p < p_{und}$ implies $\mathcal{M}^\star = \mathcal{M}^\star_{und}$ and $p \in (p_{und}, p_{sep})$ implies $\mathcal{M}^\star = \mathcal{M}^\star_{sep}$ whenever $p_{sep} > p_{und}$. This comes from the fact that we can always take p_{und} to be the largest value of p such that whenever $p < p_{und}$ then $\mathcal{M}^\star = \mathcal{M}^\star_{und}$.

Claim 3.30. *There exists p_{pool} such that whenever $p > p_{pool}$, then $\mathcal{M}^\star = \mathcal{M}_{pool}$*

Proof. First note that if $p_{sep} > p_{und}$ then $p_{pool} \geq p_{sep} \geq p_{und}$ and therefore by Claim 3.27 and the definition of p_{und} we know that $\mathcal{M}^\star_{pool}$ dominates $\mathcal{M}^\star_{sep}$ which in turn dominates $\mathcal{M}^\star_{und}$.

Now, if instead $p_{sep} < p_{und}$ then $\mathcal{M}^\star_{sep}$ is never optimal. Further, in that case we know that as $p \to 1$ then $K^\star \to \bar{K}$, in which case there exists p_{pool} such that $\mathcal{M}^\star = \mathcal{M}_{pool}$ whenever $p > p_{pool}$. □

□

3.1.4.2 Proof of Proposition 3.24

Proof. (i) First, note that it must be the case that $\mathcal{M}_{pool}$ dominates $\mathcal{M}_{sep}$ for all $p \in (p_{sep}, \bar{p})$. Namely, as p increases above p_{sep}, then $K^\star_\ell$ is weakly decreasing as the marginal benefit of higher capital for the ℓ-type decreases given that the probability of the ℓ-type decreases. Further, we have shown that $K^\star$ is strictly increasing in p. Therefore, as long as $K^\star_\ell > K^\star_h$ then $K^\star > K^\star_\ell > K^\star_h$ and therefore $\mathcal{M}_{pool}$ strictly dominates $\mathcal{M}_{sep}$. Hence, the only way that $\mathcal{M}_{sep}$ can dominate $\mathcal{M}_{pool}$ when $p \in (p_{sep}, p_{pool})$ is if $p > \bar{p}$ and therefore $K^\star_h > K^\star_\ell$.

(ii) We know that for all p' there exists $\bar{a}$ such that $\bar{p} > p'$ whenever $a_h > \bar{a}$. This comes from the fact that the marginal benefit of increasing K_h is $p \cdot G(-a_h - K_h)$ which goes to zero as $a_h \to +\infty$. What is left to prove is that p_{pool} is bounded away from 1 as $a_h \to +\infty$. In order to show this, we note that if s_{pool} were equity, then denoting K_e the optimal capital requirement when the banks are restricted to issuing equity then we know that $K^\star \geq K_e$. Further, K_e is determined by

$$b_h(K_e) = \frac{1-p}{p} \frac{a_h - a_\ell + b_h(K_e) - b_\ell(K_e)}{a_h + b_h(K_e) + K_e} \cdot K_e$$

or equivalently

$$K_e = \frac{p}{1-p} \cdot \frac{a_h + b_h(K_e) + K_e}{a_h - a_\ell + b_h(K^\star) - b_\ell(K_e)} b_h(K_e)$$

Next, note that $b_h(K_e)$ is strictly decreasing in a_h and as $a_h \to +\infty$ it is the case that $b_h(K_e) \to \hat{b}$ for all values of K_e. Further,

$$\lim_{a_h \to +\infty} \frac{a_h + b_h(K_e) + K_e}{a_h - a_\ell + b_h(K_e) - b_\ell(K_e)} = 1$$

and this expression is strictly decreasing in a_h. Therefore,

$$K_e \to \frac{p}{1-p} \cdot \hat{b}$$

as $a_h \to +\infty$. Furthermore, we know that K_e is decreasing in a_h and therefore

$$K^\star > \frac{p}{1-p} \cdot \hat{b}$$

for all $a_h > a_\ell$.

Hence, we have just shown that $K^\star \to +\infty$ as $p \to 1$ for all $a_h > a_\ell$ and therefore there exists $\tilde{p} < 1$ such that $p_{pool} < \tilde{p}$ for all a_h. Finally, given that we know there exists $\bar{a}$ such that whenever $a_h > \bar{a}$ then $\bar{p} > \tilde{p}$ for any $\tilde{p} < 1$, we have proven our claim. □

3.1.4.3 Proof of Corollary 3.25

Proof. Part (i) simply states the condition for $K_\ell^\star = \bar{K}$ and $K_h^\star = 0$ to be incentive compatible under the optimal separating mechanism. Therefore, if $b_\ell(\bar{K}) \geq b_\ell(0) - b_h(0)$ then the regulator could always implement $K_\ell^\star = \bar{K}$ and $K_h^\star = 0$ regardless of the value of p and therefore $\mathcal{M}_{und}^\star$ never dominates $\mathcal{M}_{sep}^\star$ so that $p_{und} = 0$.

(ii) Suppose that $b_\ell(\bar{K}) < b_\ell(0) - b_h(0)$ so that $p_{und} > 0$. Then, given that $b_\theta(K) = \hat{b} + L_\theta(K)$ we note that $p_{und} > 0$ only if

$$\hat{b} < L_\ell(0) - L_h(0) - L_\ell(\bar{K}) \tag{3.7}$$

and noting that as $\hat{b} \to +\infty$ then $L_\theta(K) \to 0$ as $Pr(x < -a_\theta - K) \to 0$. Hence, RHS of 3.7 goes to 0 as $\hat{b} \to +\infty$ and therefore there exists $\bar{b}$ such that $\hat{b} > \bar{b}$ implies that $b_\ell(\bar{K}) \geq b_\ell(0) - b_h(0)$. □

4 Robust Regulation of Bank Risk: Reporting and Risk Aversion

Abstract

We study the design of bank capital regulations when banks have private information regarding the riskiness of their assets. Under any risk sensitive capital requirement scheme, a higher capital requirement signals to the market that the bank is riskier leading in many cases to (1) a decrease in the underpricing of the bank's equity and (2) an increase in the required return on equity when investors are risk averse. The interaction between these two terms crucially depends on the level of investor risk aversion and pins down the bank's incentive to truthfully report risk. We study the ability to design a risk sensitive capital requirement scheme that is robust to small perturbations of the level of risk aversion. We show that for any perturbation of the initial level of risk aversion, there exists a capital requirement scheme that is robust to that perturbation (i.e. maintains truthful reporting) only if the level of risk aversion is sufficiently small or sufficiently large.

Financial innovations provide many benefits to society, yet we have learned from the 2007-09 financial crisis that they can also create some unforeseen costs. For example, most if not all US banks were regarded as well capitalized by regulators prior to the 2007 housing market crash which revealed many of the undetectable risks accumulated through the use of complex financial instruments.[64] This leads to a natural question; how should bank's be regulated given that financial innovations create ever more opportunities for banks to hide undetectable risks? [65]

Suppose a bank privately learns that its existing assets have become riskier. Under the internal ratings based approach of Basel III, the bank would be required to report a higher level of risk weighted assets which maps to a higher capital requirement for the bank. If this increase in risk weighted assets would require the bank to raise additional capital, then this would substantially affect the bank's incentive to report this information as failure to do so would not be detected for some period of time. Yet, the regulator may be able to provide incentives to the bank's to reveal this information, particularly when this is good news for equity holders (i.e. when more risk implies a higher value of equity). In this paper we show

[64] It is still not clear whether those same financial innovations were designed to help repackage and share risk or to arbitrage financial regulations (see e.g. Jones (2000), Rajan (2005), Yorulmazer (2013), and Acharya et. al. (2013)).

[65] Further, the recent inability of credit ratings to accurately rate certain asset classes (see e.g. Ashcraft et. al. (2011)) and the ability of banks to game the internal ratings based approach (see e.g. Begley et. al. (2017), Behn et. al. (2014), and Plosser and Santos (2014)) creates a strong motivation for understanding what regulators can do to help control unobservable bank risk.

how the proper design of such incentives crucially relies on the underlying level of investor risk aversion and study the feasibility of designing risk based capital requirements that are robust to small changes in investor risk appetite.

Risk aversion is a crucial element of financial markets and asset prices. It has been well documented that investor risk aversion varies over time and with other economic fundamentals (see e.g. the survey by Cochrane (2017)). Yet, modern literature on banking regulation assumes risk neutrality of all agents and postulates that adding risk aversion to the model would not substantially alter the qualitative results and only lead to quantitative differences in the resulting variables.[66] In this paper we aim to alter this discourse by studying a model of bank regulation whereby banks, funded with insured deposits, have private information about the riskiness of their assets, measured as the spread of the distribution of returns. The regulator would like to design a mechanism consisting of report specific capital requirements and transfers to induce the bank to reveal its risk and appropriately raise more capital when necessary. Yet, an issue that has been overlooked by the banking literature is that such revelation will have substantial effects on the cost of financing as the market will also learn from the banks actions (i.e. raising capital or reporting higher RWAs) that the bank is riskier. Importantly, we document here how the change in the cost of raising capital from this revelation of information crucially depends on the level of market risk aversion.

Given the difficulty to measure the level of investor risk aversion for both the bank and the regulator, we take risk aversion to be an exogenous variable that is estimated by the regulator and learned by the bank, but not incorporated into the design of the capital requirements. While this approach is not without loss, it provides a benchmark for understanding how well existing regulations — which do not rely on the bank's reported perception of the level of risk aversion — fare when this information is not utilized. Further, given the many calls for a decrease in the complexity of banking regulations (see. e.g. Haldane and Madouros (2012)) it is hard to imagine that such a variable will be incorporated into the calculation of capital requirements anytime soon. For this reason, we focus on the issue of providing incentives that are robust to small perturbations in risk attitudes.[67]

The first result of this paper is to show that for any risk sensitive capital requirement scheme that maps higher reported levels of risk to higher capital requirements, there always exists an intermediate level of risk aversion which renders that scheme no longer incentive compatible. What this implies is that under such a scheme, when the level of risk aversion

[66] While some of the older literature considered bank regulation from a portfolio maximization perspective, in which case risk aversion plays a role, it also excluded the possibility that banks have private information.

[67] Madarász and Prat (2017) follow a similar line of reasoning when studying robustness issues in a setting with a seller who has a misspecified model of the buyers preferences and designs a mechanism that does not elicit those preferences.

falls in the relevant region, either riskier banks will have an incentive to report that they are less risky (to avoid having to raise more capital) or less risky banks will have an incentive to report they are more risky (to avoid having to pay a larger transfer). Hence, when regulators find it optimal for riskier banks to raise more capital, this break down in incentives creates an inefficiency whereby a subset of banks are either under or over capitalized.

Next we suppose that regulators have the ability to observe the initial level of investor risk aversion but can only do so imprecisely or infrequently. In this case, the regulator would like to design a capital requirement scheme that remains *robust* to small perturbations of the initial level of risk aversion. The main result is to show that for any risk sensitive capital requirement scheme there always exists an interval of intermediate values of the initial level of risk aversion such that that scheme is not robust to *any* perturbations around that level. A corollary of that result is that the regulator can provide robust incentives for higher risk banks to raise more capital if and only if the level of risk aversion is either sufficiently small or sufficiently large. Important to note is that these results continue to hold whether or not the regulator finds it optimal for riskier banks to raise more capital as opposed to resolving the adverse selection problem that occurs when banks retain this private information (see e.g. Rivera (2019) for a detailed discussion of this case).[68]

When attempting to provide incentives for banks to reveal private information about risk, the most important incentive effects come from the fact that this type of disclosure will affect the bank's costs of raising capital, an effect ignored by the majority of the literature that takes the cost of raising capital as exogenous.[69] In particular, there will be two countervailing effects when a bank reveals that its portfolio is riskier than the market's expectation: First, bank's with a larger spread on their distribution of returns, keeping the expected return constant, will generate higher returns for their shareholders through the increased value of the deposit insurance put option (Merton 1977). Hence, revealing this information to the market will lead to a decrease in the the cost of capital financing. This is a standard Myers & Majluf (1984) argument as when the market believes that the bank can take less risk, then it believes that the return on the bank's equity is lower and therefore requires a larger share of the returns to break even on the investment (i.e. the market underprices the bank's equity). If this were the only effect then there would always be scope for the regulator to design incentive compatible regulations whereby riskier banks are subject to higher capital requirements (see Rivera (2019)). But, in contrast to the Myers & Majluf (1984) case,

[68] As illustrated in Rivera (2019), when bank's have private information about the profitability of their existing assets then capital requirements will lead the more profitable banks to forgo positive NPV projects unless the capital regulations provide a channel for incentive compatible information revelation to the market.

[69] All effects described with respect to issuing equity would apply as well for issuing debt and other hybrid securities.

because of the fact that such a mechanism signals not only the profitability of the bank but also the riskiness of the bank's returns, this creates another effect: when investors are risk averse they demand a higher return on equity when they learn that the bank's assets are riskier. We call these two effects the *dilution* effect — whereby revealing more risk increases the bank's share price — and *safety* effect — whereby revealing more risk increases the risk premium on the bank's equity and therefore decreases the bank's share price.

In order to illustrate our main results, first suppose that bank's all face the same capital requirement and the regulator's only goal is to provide incentives for the banks to truthfully reveal their privately known level of risk to the market. Then, for two levels of risk (e.g. variance), one high and one low, there must exist a level of risk aversion such that bank is indifferent between a portfolio with the high and low level of risk, keeping the expected return the same. In this case, the dilution effect perfectly offsets the safety effect when reporting the high v.s. low level of risk. It is precisely around this level of risk aversion that incentive compatibility breaks down when capital requirements are risk sensitive. Namely, if the bank is indifferent between reporting it is high risk or low risk because the pricing of the bank's equity will be the same, then the bank must also be indifferent between holding the high level of capital and the low level when they differ. This, of course, will never be the case when deposits are insured (i.e. cheap). Therefore, for any two different capital requirements there exists a region around this indifference point such that whenever the level of risk aversion falls in that region then both types prefer to raise less capital as the benefit of doing so outweighs the benefit of signaling a higher level of (risk neutral) profitability.

It is worth discussing why incentive compatibility becomes infeasible for intermediate values of risk aversion. In the case of low levels of risk aversion, incentive compatibility can be achieved because the high risk banks are willing to pay more than the low risk banks to signal their profitability (i.e. the dilution effect largely outweighs the safety effect). Hence the high risk banks can be charged a higher capital requirement while still maintaining incentive compatibility. Similarly, when the level of risk aversion is very high, the low risk banks now benefit from revealing to the market that they are less risky (i.e. the safety effect largely outweighs the dilution effect) and in a similar fashion are willing to pay more to reveal this information than the high risk types and therefore are willing to pay a higher transfer. Hence, in this case the regulator can set a high enough transfer for banks that report low risk in order to generate incentive compatibility while still charging the high risk types a higher capital requirement.

Extending this logic to the aim of providing robust incentives, we note that if the regulator's only objective was to get banks to truthfully report their private information to the market (i.e. capital requirements are independent of risk reports) then there would always

exist an incentive compatible mechanism. The issue is that this mechanism relies solely on the transfer in order to produce the right incentives for truthful reporting. Further, when the level of risk aversion is low the high risk transfer must be larger than the low risk transfer to keep the low risk type from benefiting from selling over priced shares when mimicking the high risk type. Similarly, when the level of risk aversion is high then the low risk transfer must be higher than the high risk transfer as in that case the high risk type will sell overpriced shares when mimicking the low risk type (when risk aversion is high, low risk banks generate a higher risk adjusted return than high risk banks). Finally at the aforementioned indifference point whereby the bank is indifferent between owning the high risk or low risk portfolio, the transfers must be equal. Hence, robustness around this indifference point requires the difference in transfers to be both positive, equal, and negative. This contradiction implies that the initial level of risk aversion needs to be sufficiently far away from this indifference point in order to guarantee robustness to any perturbation of the initial level.

Finally, we note that for high levels of risk aversion it is clear that bank risk taking is not optimal. In fact, once the level of risk aversion exceeds the indifference level described above, then the high risk bank will optimally choose the level of risk of the low risk bank (or less). In this case, our results still hold, only with a smaller upper bound on the interval whereby robust incentives fail to exist. Namely, incentive compatibility becomes trivial when the level of risk aversion is large, but still remains an issue when the level of risk aversion is small to intermediate. On the other hand, if banks vary in their ability to both increase and decrease the riskiness of their portfolios then banks will decrease their risk as much as possible when the level of risk aversion is large, but there will still be a role for risk sensitive capital requirements and therefore a robustness issue.

4.1 Literature Review

Our motivation for the regulation of bank risk taking begins with the introduction of deposit insurance and the subsequent implementation of Basel I and Basel II capital requirements. Many papers since (e.g., Koehn and Santomero (1980), Kahane (1977), and Gennotte and Pyle (1991), Blum (1999)) have shown how inefficiently priced deposit insurance can lead to higher incentives for bank risk taking and how the introduction of a leverage ratio can potentially exacerbate these incentives. Kim and Santomero (1988) and Rochet (1992) show that for this reason capital requirements should be weighted by the risk of the bank's assets and construct the *theoretically optimal* risk weights under differing assumptions. In line with this reasoning, the Standardized Approach of Basel I-III defines capital requirements by associating with each asset a risk weight and then determining the banks capital requirements

as a percentage of *risk weighted assets*. In light of this, Chan et. al. (1992) show that when depository institutions are perfectly competitive, then a fairly priced incentive compatible deposit insurance pricing scheme may fail to exist. Similarly, Giammarino et. al. (1993) extend the results of Chan et. al. (1992) to show that in general the regulator can discriminate among banks on the basis of their level of risk, but that any mechanism that does so will give banks an incentive to lower their asset quality. In this paper we differ from this literature in that we provide a micro-foundation of the bank's cost of raising capital, combining the effects of asymmetric information on the mispricing of equity (Myers and Majluf (1984)) with the natural fact that risk aversion leads to an increased required rate of return on equity for riskier banks. Typically, either one or both of these effects have been ignored or taken as exogenously given in the previous studies of bank regulation.

Another motivating factor for this paper is the large consensus that bank's risk weighted assets to not accurately reflect their true risk (see e.g. Vallascas and Hagendorff (2013) or Ferri and Pesic (2017) and the summary of existing studies they provide). There is related line of literature that aims to provide evidence that the mismatch of risk weights and market based risk estimates is a strategic choice of the banks. Of the empirical papers, Plosser and Santos (2018) show how banks with less capital report lower risk for the same loans within loan syndicates. Similarly, Begley et. al. (2017) show how value at risk violations are negatively correlated with bank capital. Mariathasan and Merrouche (2012) represents the closest empirical verification of our results. Namely, they ask whether the unweighted leverage ratio or the risk weighted asset ratio are better predictors of bank failure, showing that the leverage ratio performs better when the risk of crisis is high. This supports our results which state that changes from low to high levels of risk aversion — which can be argued as either a cause or effect of financial crises (see e.g. Coudert and Gex (2008)) — leads to a breakdown in incentives for truthful reporting of risk.

From the theoretical perspective, strategic underreporting has been studied in papers such as Leitner and Yilmaz (2019) and Colliard (2017). These papers study a situation whereby the regulator relies on the bank's private information produced by their models to determine their risk, but bank's can produce alternative *regulatory* models that the regulator can only detect through auditing. Leitner and Yilmaz (2019) show how in such a situation, more auditing leads to less information production while Colliard (2017) show how costly auditing optimally leads to less risk sensitive capital requirements. Finally, Blum (2008) studies incentive compatibility issues with the Internal Ratings Based Approach of Basel II and shows that if the regulator has a limited ability to punish banks ex-post for misreporting risk than a leverage ratio can improve welfare.

The closest related papers on the mechanism design aspect of bank regulation are Gi-

ammarino, Lewis, and Sappington (1993) and Rivera (2019). Giammarino, Lewis, and Sappington (1993) study a model of bank regulation with moral hazard and adverse selection with risk neutral agents and a fixed cost of raising capital. The main contribution is to introduce a social welfare function and to highlight the effects of bank failure on the deposit insurance fund. While they characterize the optimal mechanism consisting of a menu of capital requirements and transfers they implicitly assume that investors ignore the differences in bank deposit insurance premia and capital requirements (given that the cost of raising capital does not depend on these variables). Namely, under their mechanism riskier banks will face a higher capital requirement but the same cost of raising that capital as less risky banks. Rivera (2019) endogenizes the cost of raising capital when banks are privately informed about the quality of their assets and shows how the assumptions of Giammarino, Lewis, and Sappington (1993) are not without loss. Yet, Rivera (2019) maintains two assumptions (1) that bank's only vary in their private information about asset quality (as opposed to risk) and (2) that all agents are risk neutral.

Rochet (1992) looks at the case where banks are portfolio maximizers with concave utility (although assuming no possibility of raising capital) and show that if the Arrow-Prat coefficient of relative risk aversion is decreasing (increasing) then the default probability of an unregulated bank is an increasing (decreasing) function of its risk adjusted net worth. Hence, while Rochet (1992) makes many strong assumptions that are relaxed in this paper it is still the closest related paper in terms of the study of risk aversion on bank behavior in the context of banking regulation.

We believe that this is one of the first papers to study the issue of providing robust incentives for reporting risk in the context of bank capital regulation. That being said, robustness concerns have been highlighted in the banking regulation literature (see e.g. Acharya et. al. (2011)), the macroeconomics literature (see e.g. Hansen and Sargent (2001)), as well as in contract theory and mechanism design (see e.g. Carroll (2015) or Bergemann and Morris (2005)). Carroll (2018) surveys the literature on robust mechanism design and contracting and shows how much of the focus in the existing literature is on robustness to large changes in the underlying environment. Yet, there a few papers that look at robustness in applications whereby small mis-specifications of the environment can lead to large changes in payoffs (see e.g. Masarász and Prat (2017)). This paper falls into this line of literature, showing how incentive compatibility can break down entirely among a range of levels of risk aversion and therefore robustness of incentive compatibility cannot be obtained with respect to any level of risk aversion nearby to those values even when allowing for small levels of mis-specification and fundamental changes to the mechanism (i.e. a decrease in the level of risk sensitivity of the capital requirements).

4.2 Model

The bank starts with initial capital K_0 and existing assets with net returns distributed according to the probability distribution function f_0. The bank, then receives an investment opportunity with a fixed cost I and returns such that the combined portfolio of existing assets and the new investment results in the *observable* distribution of net returns f_1. One can think of this new investment as the most profitable opportunity available to the bank at this time given the risk that it carries. The bank must raise external funds for the new investment which we assume will consist of insured (subsidized) deposits D and newly raised equity K_1 (which we will also refer to as capital). We assume the regulator sets a capital requirement so that $K_1 \geq \rho \cdot I$ for some $\rho \in [0, 1]$ and will simply refer to $K_1 \in [0, I]$ as the bank's capital requirement.

Prior to raising new capital K_1 we assume that the bank learns private information about the new investment opportunity which gives them the ability to adjust the distribution f_1 by increasing or decreasing the *spread* of f_1 without detection. Their ability to make such adjustments is parameterized by $\theta \in [0, \overline{\theta}]$ with the bank's ex-post (unobserved) distribution denoted by f_θ. Practically, we can think of the new investment as a bundle of loans, in which case this *risk shifting* is equivalent to the bank privately learning that some of its new loans are more or less risky via *soft information* and then subsequently deciding which loan applications to accept or reject. We say that the bank increases (decreases) risk if f_θ (resp. f_1) is a spread of f_1 (resp. f_θ): f_1 second order stochastically dominates (is dominated by) f_θ. Further, in order to pin down the risk taking technology (and rule out certain technical difficulties) we assume that banks continuously increase or decrease the spread of their assets represented by a single risk parameter σ. Hence, if $\sigma_\theta > \sigma_{\theta'}$ then f_θ stochastically dominates $f_{\theta'}$ in a strict sense so that $\int_{-\infty}^{x} F_\theta(t) - F_{\theta'}(t)dt > 0$ for almost all $x \in \mathbb{R}$. Assuming the distribution of returns f_1 generates a level of risk σ_1 (e.g. variance), then for simplicity we will assume that the distribution f_θ generates risk $\sigma_\theta := \sigma_1 + \theta$. In order to focus on the reporting of risk, we will assume that the only difference between f_θ and f_1 is the spread so that $\mathbb{E}_{f_1}[X] = \mathbb{E}_{f_\theta}[X]$ for all $\theta \in [0, \bar{\theta}]$. As we will show below, provided that the level of risk aversion is not too large, the bank's owners will find it optimal to increase risk even in this case where such an increase in risk does not come with an increase in the expected return. Therefore, the incentive to take risk will be stronger if the expect return also increases but we maintain this assumption in order to isolate the effect of increased risk on the cost of raising equity and therefore the incentives to report that risk.[70]

[70]Naturally, any technology that produces more risk for lower returns should be such that the decrease in returns is small in order for the bank to find it optimal to engage in that form of risk shifting. Hence, the results with this form of technology will not vary substantially from the case we study.

Another interpretation of this problem is that the bank receives an initial shock to its existing assets to which the market imperfectly observes the resulting distribution $f_{\tilde{\theta}}$. In this case, the market would require the bank to decrease the risk on its balance sheet by returning the spread to its initial level (resulting in a distribution f_1) for example through certain debt covenants or regulatory mandates. We assume though that the bank has the ability to convince the market that it has done so (e.g. through a recalibration of internal model risk weights) while actually only lowering the spread to f_θ with $\theta \leq \tilde{\theta}$, with a larger θ representing the bank's better ability to hide risk. This interpretation could be practically seen as banks receiving an external shock to their existing assets resulting in an increase in their risk weighted assets (RWAs) under the internal ratings based approach of Basel III which would require an increase in capital or sale of existing assets. Yet, as has been observed in practice, we assume the bank could instead adjust its internal models to effectively lower its RWAs. Further, we assume that banks have differing ability to perform this adjustment without detection so that some banks (depending on the size of the shock) could lower their RWAs without detection by a larger level than others.

We assume that the risk neutral regulator sets the capital requirement to maximize welfare consisting of the economic surplus (NPV) of the new investments minus the deadweight loss of bank failure. In what follows we will assume $K_0 = 0$ to ease notation but note that this will not affect the main results.[71] Denoting by X the random variable representing the net return on the bank's assets, then social welfare when the bank has assets f and a capital stock K_1 can be given by:

$$W(f, K_1) := \mathbb{E}_f[X] + \lambda \cdot \int_{-\infty}^{-K} (x + K) f(x) dx - c(K_1) := \mathbb{E}_f[X] - \lambda \cdot L(f, K_1) - c(K_1)$$

where $L(f, K_1)$ denotes the expected loss to the deposit insurance fund and λ is a parameter which represents the deadweight loss of bank failure. Note that without a deadweight loss of bank failure (or equivalently deadweight loss of taxation used to repay depositors), the regulator would be indifferent to the bank's level of capital as losses to depositors is just an effective transfer from the depositors to the bank's shareholders and therefore washes out of the welfare function. Using this logic, we can also refer to $L(f, K)$ as the value of the deposit insurance put option in the sense of Merton (1977). Further note that there is a social cost of raising capital $c(K_1)$. We think of this cost as the expected underinvestment caused by the capital requirement K_1 due to adverse selection on the value of the bank's existing assets.

[71] Including existing equity (e.g. in the form of retained earnings) would adjust the distribution of returns f_θ but given that we make no assumptions on the functional form of these functions other than Assumption 4.3 below, our main proofs will still be valid.

While we abstract away from this consideration, it does not matter for the main results which are a statement regarding the incentives of banks to reveal their private information when that information maps to differing capital requirements and market beliefs.[72] The main purpose that this cost serves in this paper is to generate an outcome whereby the regulator would like to charge different capital requirements to different banks; without this cost the regulator would optimally set $K_1 = I$ independent of the bank's returns f. Instead, when including this cost, the first best capital requirement K_1^{FB} satisfies:

$$\frac{\partial}{\partial K_1} L(f, K_1^{FB}) = F(-K_1^{FB}) = \frac{\partial}{\partial K_1} c(K_1^{FB})$$

The main issue for the regulator is that higher σ_θ implies a higher first best capital requirement K_θ yet the variable θ is unobservable. Hence setting a single capital requirement would result in some bank's undercapitalized and some overcapitalized creating an inefficiency. Hence, the motivation of the regulator to induce the bank's to truthfully reveal their private information so that they can pair the bank's risk, measured by θ, with a risk sensitive capital requirement K_θ. We further assume that the regulator has the ability to impose an ex-ante tax T_θ on the bank (also linked to the bank's report) which can be interpreted as a deposit insurance premium or the implicit cost of higher regulatory supervision and auditing of the bank's risk management practices. Thus, without loss the regulator will choose a menu consisting of a pair (K_θ, T_θ) for each $\theta \in [0, \bar{\theta}]$. In what follows we will restrict attention to mechanisms $\mathcal{M} = \{(K_\theta, T_\theta)\}_{\theta \in [0,\bar{\theta}]}$ such that $\theta > \theta'$ implies $K_\theta \geq K_{\theta'}$ as the focus of this paper is on understanding how to provide proper incentives for banks to reveal risk, especially when that risk will be paired with a higher capital requirement. Other than this assumed *optimality* condition, the results of this paper will only have to do with the *ability* to provide incentives, taking as given that doing so is optimal (or may be in certain cases).

It is important to note here that we are making a restriction by focusing solely on the class of mechanisms $\mathcal{M} = \{K_\theta, T_\theta\}_{\theta \in [0,\bar{\theta}]}$. Namely, more generally we could require the bank to report the level of risk aversion to the mechanism. We restrict from focusing on this form of mechanism as from a practical stand point it does not seem feasible that regulators would continually probe banks for estimates of risk aversion and pair those estimates with capital requirements and transfers (e.g. deposit insurance premia). Namely, if we assume that the bank does not learn the level of risk aversion until it raises capital, then the regulator must provide incentives to the bank to reveal its information based on its belief of the level of market risk aversion. Hence, we would be studying a situation whereby the bank's belief of the level of market risk aversion is continuously reported and tied to the bank's

[72] The full treatment of this problem is available in Rivera (2019).

capital requirements and transfers. Not only would this make capital regulations extremely complex — something regulators hope to move away from (see e.g. Basel (2013) or Haldane and Madouros (2012)) — but such a mechanism would interfere with the principal aim of the paper which is to provide to the market useful information about the bank's risk. Practically speaking, if risk maps to capital requirements in a strictly monotone way, then it will be very simple for the market to discern the bank's level of risk as implied by higher levels of risk weighted assets. On the other hand, if capital requirements map from the level of risk *and* the bank's estimated level of market risk aversion (a much more difficult moving target), then in pricing the bank's equity investors will have to inverse this mapping which may not be possible (e.g. if low risk and low risk aversion map to the same capital requirement as high risk and high risk aversion). Finally, another potentially more important purpose for studying this restricted class of mechanisms is to understand how the current regulations (which do not vary with the market level of risk aversion) will hold up in this moving environment, and when we can expect them to display certain robustness properties.

In what follows we will be interested in differing levels of risk aversion of the agents in our model and how it will affect the regulator's ability to control bank risk taking and/or induce bank's to reveal their private information. For simplicity we will parameterize risk preferences with a single variable γ (e.g. the constant of relative/absolute risk aversion) shared by all agents in the economy including the bank's existing shareholders and their prospective investors. This is without loss given that bank's control a single risk taking variable σ. We thus assume that all agents share the same utility function u_γ parameterized so that when $\gamma = 0$ agents are risk neutral and therefore the value of the bank of size I with distribution of returns f is:

$$\mathbb{E}_f[u_0(I + X)] = I + \mathbb{E}_f[X]$$

On the other hand, $\gamma > 0$ implies that agents are risk averse so that

$$\mathbb{E}_f[u_\gamma(I + X)] < I + \mathbb{E}_f[X] \tag{4.1}$$

and $\mathbb{E}_f[u_\gamma(I + X)] < \mathbb{E}_f[u_{\gamma'}(I + X)]$ for all $\gamma > \gamma'$ and f with positive variance.[73] Further, this also implies that $\mathbb{E}_f[u_\gamma(I + X)] < \mathbb{E}_{f'}[u_\gamma(I + X)]$ for any $\gamma > 0$ whenever the variance of f is strictly greater than the variance of f'.

We will make the following further assumptions regarding the class of utility functions under consideration:

Assumption 4.1. The class of utility functions $\{u_\gamma\}_{\gamma \geq 0}$ satisfy

[73] We do not rule out the case where $\gamma < 0$ so that banks are risk loving, but our main focus will be on the case where $\gamma \geq 0$.

(1) $\mathbb{E}_{\delta_x}[u_\gamma(I+X)] = I + x$ where δ_x is the distribution that returns x with certainty.
(2) $\mathbb{E}_f[u_\gamma(\alpha(I+X)))] = \alpha \cdot \mathbb{E}_f[u_\gamma(I+X)]$ for all $\alpha \in [0,1]$.

In particular, in our numerical examples we will use the class of power utility functions such that:

$$\mathbb{E}_f[u_\gamma(I+X)] = (\int (I+x)^{1-\gamma} f(x)dx)^{\frac{1}{1-\gamma}}$$

As can be seen when $\gamma = 0$ this corresponds to the risk neutral expectation whereas whenever $\gamma > 0$ this utility function penalizes variance in returns. This class of utility functions is typically utilized to isolate risk aversion in inter-temporal consumption models (see e.g. Epstein & Zin (1989)). Focusing on this class of utility functions allows us to focus on the incentives to communicate risks across different levels of risk aversion (1) without imposing that the utility of certain outcomes is decreasing in risk aversion (as is the case for CARA and CRRA utility functions) and (2) ignoring issues of scale. Note that both of these assumptions are primarily made to ease the exposition and proofs.

Given that, other than for examples, we will be utilizing a general utility function to capture risk preferences, we should be sure that this class of utility functions displays the intuitive features of risk aversion. This leads to the following assumption.

Assumption 4.2. Assuming f_θ has variance $\sigma_1 + \theta$, then for any K and $\gamma > 0$,
(1) $\frac{\partial}{\partial \gamma} V_\gamma(f_\theta, K) < 0$.
(2) $\frac{\partial^2}{\partial \theta \partial \gamma} V_\gamma(f_\theta, K) < 0$ with $\lim_{\gamma \to 1} \frac{\partial}{\partial \theta} V_\gamma(f_\theta, K) = -\infty$.
(3) $\frac{\partial^2}{\partial K \partial \gamma} V_\gamma(f_\theta, K) < 0$.

The first condition states that when investors are more risk averse, they value the bank's equity (i.e. a risky investment) less. The second condition first states that this decrease in the valuation of the bank's equity due to an increase in risk aversion is larger the riskier the bank is or equivalently that the marginal value of taking risk is strictly decreasing in the level of risk aversion. The next condition of (2) when combined with the first states that risk taking becomes prohibitively expensive as investors become *infinitely risk averse*. Finally, the third condition states that the marginal benefit of raising equity is decreasing in the level of risk aversion. This is a condition that we always expect to hold whenever inside shareholders do not find it optimal to raise equity. Namely, raising equity decreases the value of the deposit insurance put option but also decreases the risk premium the bank must pay by making it safer. Yet, as γ increases the decrease in the value of the deposit insurance put option is the same, but the decrease in the risk premium is less now that investors are more risk averse. Hence, we expect condition (3) to hold for any sensible class of utility functions whenever raising equity is costly for the bank's inside shareholders.

A bank of type θ has the objective of maximizing the risk adjusted return of its existing (inside) shareholders, subject to meeting the capital requirement K. We normalize the risk free rate required by depositors to zero, in which case the only potential cost the bank faces is in raising new (outside) equity K. To this extent we assume that equity markets are competitive so that outside investors are willing to pay K for a share of the firm α that generates an expected risk adjusted return equal to K. Note that the post investment value of the bank's equity (inside plus outside) is given by

$$V_\gamma(f, K) := \int_{-K}^{\infty} u_\gamma(K + X) f(x) dx$$

with the truncation coming from limited liability. Further, this implies that in the perfect information case, the break even condition of the outside equity holders given a level of risk θ is

$$\alpha_\theta \cdot V_\gamma(f_\theta, K) = K$$

and therefore for a generic capital requirement K, the objective of the bank of type θ is:

$$\max_{\tau \in [0,\theta]} \ (1 - \alpha_\tau) \cdot V_\gamma(f_\tau, K) = V_\gamma(f_\tau, K) - K \tag{4.2}$$

We will now make a key assumption regarding the preferences of the bank.

Assumption 4.3. The marginal cost of raising capital $c'(K)$ and the class of distributions $\{f_\theta\}_{\theta \in [0,\bar{\theta}]}$, are such that there exists $\overline{K}$ such that $K_\theta > \overline{K}$ is never socially desirable for any $\theta \in [0, \bar{\theta}]$, and

$$\frac{\partial}{\partial K}[V_\gamma(f_\theta, K) - K]|_{K=\overline{K}} < 0 \tag{4.3}$$

$$\frac{\partial}{\partial K}[V_\gamma(f_\theta, K) - V_\gamma(f_{\theta'}, K)]|_{K=\overline{K}} < 0 \tag{4.4}$$

for all $K < \overline{K}$, $\theta > \theta'$, and all $\gamma \geq 0$.

The first point of this assumption states that optimal capital requirements are bounded for the class of return distributions under consideration. Inequality (4.3) states that for all relevant bank types and levels of risk aversion, the existing shareholders *dislike* raising capital. The reason why such a condition needs to be stated this way is that when banks are sufficiently capitalized then they have enough skin in the game to want to raise even more capital. Hence, capital requirements will never bind and a capital requirement is redundant. Given that we do not model any friction that would give banks a reason to not want to be 100% equity financed in this case we think that proceeding with such an assumption is reasonable. Another way to phrase this assumption is to simply state that socially desirable

or politically feasible capital requirements are not too high. Inequality (4.4) on the other hand states that the marginal increase in the value of total equity (as opposed to existing shareholder equity) due to an increase in capital is less for risker banks. This is a natural assumption that should hold in practice and holds in the risk neutral case, the only issue is whether this result holds for large values of γ which we assume here is the case.[74]

In the appendix we show how the bank's problem can be reformulated in terms of risk premium $r_{\gamma,\theta}(K)$ (measured in per dollar of equity terms) so that the risk adjusted value of equity can be represented as $V_\gamma(f_\theta, K) = V_0(f_\theta, K) - r_{\gamma,\theta}(K) \cdot K$. We will utilize this formulation in order to generate intuition at times below.

4.3 Main Results

As a benchmark case we will assume that the bank of type θ knows that its distribution of returns is f_θ but cannot make changes to f_θ and then proceed in the next section to the case whereby the bank can potentially reduce the level of risk to $\theta' < \theta$ (which will be a simple extension of the benchmark case).

In order to understand the incentives of the bank, suppose that the regulatory mechanism consists of a menu $\{(K_\theta, T_\theta)\}_{\theta \in [0,\bar{\theta}]}$. Note that we drop the dependence on the level of risk aversion γ for the moment. Then, in order for the bank of type θ to find it optimal to report truthfully instead of reporting it is some type $\theta' < \theta$ it must be the case that

$$V_\gamma(f_\theta, K_\theta) - K_\theta - T_\theta \geq V_\gamma(f_\theta, K_{\theta'}) - K_{\theta'} \cdot \frac{V_\gamma(f_\theta, K_{\theta'}) - T_{\theta'}}{V_\gamma(f_{\theta'}, K_{\theta'}) - T_{\theta'}} - T_{\theta'} \tag{4.5}$$

and

$$V_\gamma(f_{\theta'}, K_{\theta'}) - K_{\theta'} - T_{\theta'} \geq V_\gamma(f_{\theta'}, K_\theta) - K_\theta \cdot \frac{V_{\gamma'}(f_{\theta'}, K_\theta) - T_\theta}{V_\gamma(f_\theta, K_\theta) - T_\theta} - T_\theta \tag{4.6}$$

Namely, on the left hand side of (4.5) (resp. (4.6)) we have the payoff to the bank if the outside investor's perfectly knew their type θ (resp. θ') — and therefore their equity is correctly priced — and when they face a capital requirement of K_θ (resp. $K_{\theta'}$). On the other hand, on the right hand side we have the profit of the bank who faces a capital requirement $K_{\theta'}$ (resp. K_θ) but who raises capital which is priced as if it was the θ' (resp. θ) type bank. Therefore, whenever the level of risk aversion is low the riskier bank is more valuable so that $V_\gamma(f_\theta, K) > V_\gamma(f_{\theta'}, K)$ for all K. Hence, whenever this is the case and the θ type reports as the θ' type then their shares are underpriced so that in order to raise capital $K_{\theta'}$ they must give away a fraction of the firm worth $\frac{V_\gamma(f_\theta, K_{\theta'}) - T_{\theta'}}{V_\gamma(f_{\theta'}, K_{\theta'}) - T_{\theta'}} K_{\theta'} > K_{\theta'}$. Similarly, in this case when

[74] For large values of γ this need not be true and therefore we are implicitly assuming that the parameters of the model are such that the relevant range of γ is below this threshold.

the θ' type reports it is the θ type, then its shares are overpriced and therefore in order to raise capital K_θ they give away a fraction of the firm worth $\frac{V_\gamma(f_{\theta'},K_\theta)-T_\theta}{V_\gamma(f_\theta,K_\theta)-T_\theta}K_\theta < K_\theta$.

Another way to express these two conditions is as follows:

$$IC_{\theta\to\theta'}(\gamma,K_\theta,K_{\theta'}) := V_\gamma(f_\theta,K_\theta)-V_\gamma(f_\theta,K_{\theta'})+K_{\theta'}\cdot\frac{V_\gamma(f_\theta,K_{\theta'})-T_{\theta'}}{V_\gamma(f_{\theta'},K_{\theta'})-T_{\theta'}}-K_\theta \geq T_\theta - T_{\theta'} \quad (4.7)$$

and

$$T_\theta - T_{\theta'} \geq V_\gamma(f_{\theta'},K_\theta)-V_\gamma(f_{\theta'},K_{\theta'})+K_{\theta'}-K_\theta\frac{V_\gamma(f_{\theta'},K_\theta)-T_\theta}{V_\gamma(f_\theta,K_\theta)-T_\theta} =: IC_{\theta'\to\theta}(\gamma,K_{\theta'},K_\theta) \quad (4.8)$$

Finally, denoting by $\mathcal{M}=\{K_\theta,T_\theta\}_{\theta\in[0,\bar\theta]}$ the regulatory mechanism, then for every θ and θ' truthful reporting requires the conditions (4.7) and (4.8) to be satisfied. We will use the notation that

$$\mathcal{IC}(\gamma,\theta,\theta'|\mathcal{M}) := \{IC_{\theta\to\theta'}(\gamma,K_\theta,K_{\theta'})\geq T_\theta-T_{\theta'}, T_\theta-T_{\theta'}\geq IC_{\theta'\to\theta}(\gamma,K_{\theta'},K_\theta)\}$$

Thus, incentive compatibility of the mechanism $\mathcal{M}$ requires that $\mathcal{IC}(\gamma,\theta,\theta'|\mathcal{M})$ be satisfied for all $\theta,\theta'\in[0,\bar\theta]$.

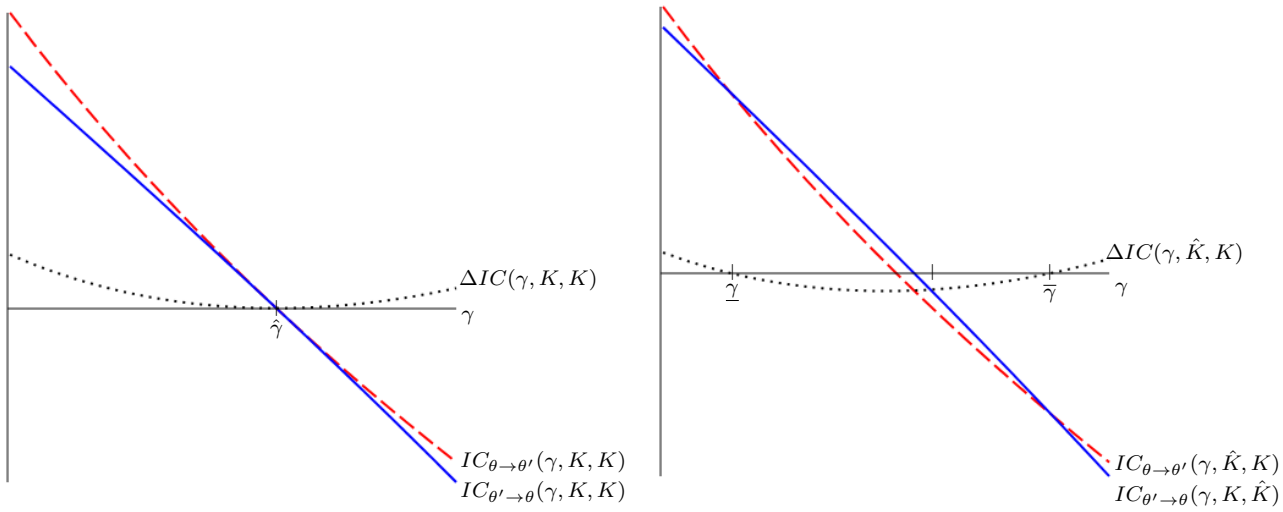
Now, note that $\mathcal{IC}(\gamma,\theta,\theta'|\mathcal{M})$ provides us with a simple necessary condition for incentive compatibility. Namely, given that the regulator would like to set higher capital requirements for the riskier type $\theta>\theta'$ then it must be the case that

$$IC_{\theta\to\theta'}(\gamma,K,K)-IC_{\theta'\to\theta}(\gamma,K,K)>0 \quad (4.9)$$

This allows us to highlight the incentives that arise for information revelation and how risk aversion will affect those incentives. Namely, when $K_\theta=K_{\theta'}=K$ and the type θ bank reports that it is less risky by claiming to be type θ', then there are two contradictory effects. First, the market offers a lower risk premium to the bank: $r_{\gamma,\theta'}(K)<r_{\gamma,\theta}(K)$. Second, when risk shifting is optimal then the value of type θ equity is greater than type θ' equity, hence the market will underprice the type θ bank's equity when it pretends to be the θ' type. We will call these the *safety* and *dilution* effects respectively, and note that the combination of these two effects will determine whether raising equity is more or less costly when imitating a safer bank. Naturally though, given that raising capital is costly (i.e. it diminishes the value of the deposit insurance put option), we can see how (4.9) is a necessary condition for incentive compatibility as when reporting truthfully the θ type will have to incur an even higher cost whenever $K_\theta>K_{\theta'}=K$ and therefore must generate a strictly higher benefit of truthful reporting whenever $K_\theta=K_{\theta'}=K$. This brings us to our first main result.

Proposition 4.4. *Consider any mechanism* $\mathcal{M} := (K_\theta, T_\theta)_{\theta \in [0,\bar{\theta}]}$ *with* $K_\theta > K_{\theta'}$ *for some* $\theta > \theta'$. *Then, there exists an interval* $(\underline{\gamma}, \overline{\gamma})$ *such that* $\mathcal{M}$ *is not incentive compatible whenever* $\gamma \in (\underline{\gamma}, \overline{\gamma})$.

The proof of Proposition 4.4 starts by showing that there exists a value $\hat{\gamma}$ such that $V_{\hat{\gamma}}(f_\theta, K) = V_{\hat{\gamma}}(f_{\theta'}, K)$. Further, we show how the necessary condition (4.9) is violated whenever $\gamma = \hat{\gamma}$. Then we show how whenever $K_\theta > K_{\theta'}$ this result translates to $\mathcal{IC}(\gamma, \theta, \theta')$ being violated for any transfers T_θ and $T_{\theta'}$ whenever the initial level of γ is close to $\hat{\gamma}$, hence the existence of the interval $[\underline{\gamma}, \overline{\gamma}]$. These results are illustrated in Figure 24. Namely, Figure 24 (a) plots the two curves $IC_{\theta \to \theta'}(\gamma, K, K)$ and $IC_{\theta' \to \theta}(\gamma, K, K)$ and their difference $\Delta IC(\gamma, K, K) := IC_{\theta \to \theta'}(\gamma, K, K) - IC_{\theta' \to \theta}(\gamma, K, K)$. Hence, as can be seen, condition (4.9) which is equivalent to $\Delta IC(\gamma, K, K) > 0$ always holds except at one point $\hat{\gamma}$ where $\Delta IC(\gamma, K, K) = 0$. In Figure 24 (b) we plot the same curves except now assuming that reporting the higher risk type θ implies a capital requirement of $\hat{K} > K$. In this case, we can see that our necessary condition is violated for all $\gamma \in (\underline{\gamma}, \overline{\gamma})$ whereby $\Delta IC(\gamma, \hat{K}, K) < 0$. The main exercise in proving Proposition 4.4 comes in proving, in Lemma 4.8 in the appendix, that there exists an interval of values around $\hat{\gamma}$ such that $IC_{\theta \to \theta'}(\gamma, K, K)$ decreases by more than $IC_{\theta' \to \theta}(\gamma, K, K)$ when increasing the θ type capital requirement from K to $\hat{K} > K$. Then the fact that the interval is more skewed towards lower values of γ with respect to $\hat{\gamma}$ comes from the fact that this decrease in the incentive functions when increasing K_θ from K to $\hat{K}$ is exactly equal at some value of γ strictly less than $\hat{\gamma}$.



(a) Incentive constraints when $K_\theta = K_{\theta'} = K$

(b) Incentive constraints when $K_\theta = \hat{K} > K$

Figure 24: Incentive constraints when f_θ is normally distributed as $\mathcal{N}(1, \sigma_\theta)$ with $\sigma_{\theta'} = 1$, $\sigma_\theta = 3$, $K = 1$, and $\hat{K} = 1.1$.

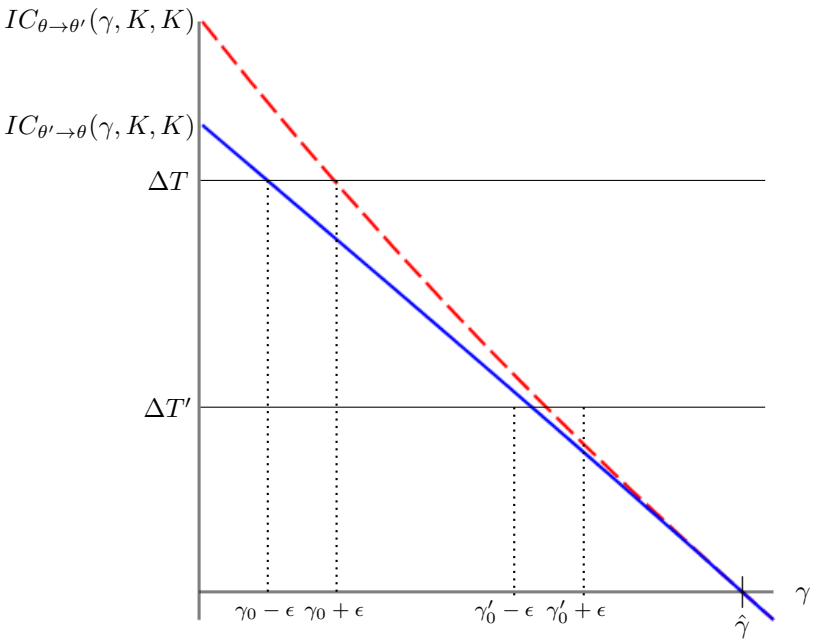


Figure 25: Incentive constraints when $K_\theta = K_{\theta'} = K$

4.4 Robust Mechanisms to Changes in Risk Aversion

In this section we ask the question: what is the form of the optimal mechanism that is *robust* to small variations in risk aversion? Meaning starting with some initially known level of risk aversion γ_0 what is the form of the optimal mechanism $\mathcal{M}$ that remains incentive compatible when the true value of γ is some ϵ-perturbation of γ_0. This leads to the following definition.

Definition 4.5. For any $\epsilon > 0$ and $\gamma_0 \in [0, 1]$ we say that the mechanism $\mathcal{M}$ is (γ_0, ϵ)-robust if is incentive compatible for all values of $\gamma \in [\gamma_0 - \epsilon, \gamma_0 + \epsilon]$.

Proposition 4.6. *For any $\epsilon > 0$ there always exists $\underline{\gamma}(\epsilon) < \overline{\gamma}(\epsilon)$ such that no (γ_0, ϵ)-robust mechanism exists whenever $\gamma_0 \in [\underline{\gamma}(\epsilon), \overline{\gamma}(\epsilon)]$.*

This proposition states that it is impossible to provide (γ_0, ϵ)-robust incentives for the revelation of bank risk whenever $\gamma_0 \in [\underline{\gamma}(\epsilon), \overline{\gamma}(\epsilon)]$. Note that this does not necessarily imply that it is impossible to design an incentive compatible mechanism that reveals the bank's information to the market. It simply states that for certain values of γ_0, those incentives will not be robust to small perturbations of γ_0.

The main logic behind this result comes from the fact that the regulator must set capital requirements and transfers independently of the potential perturbations of γ_0. Proposition 4.4 tells us that when more risk implies a larger capital requirement, then incentive compatibility is impossible to obtain for certain values of γ_0. Therefore, in this case robust incentive

compatibility is naturally also impossible. As mentioned above though, in the case where revealing more risk does not affect the bank's the capital requirement it is always possible to obtain incentive compatibility. What Proposition 4.6 implies then is that those incentives (and the incentives provided under a risk sensitive capital requirement) will not be robust to small perturbations in risk aversion for certain initial values of γ_0 as illustrated in Figure 25. Namely, Figure 25 plots the incentive constraints $IC_{\theta\to\theta'}(\gamma, K, K)$ and $IC_{\theta'\to\theta}(\gamma, K, K)$ when both θ and θ' banks face the same capital requirement K. Now, recall that incentive compatibility between θ and θ' (i.e. conditions under which neither type wishes to report they are the other) requires

$$IC_{\theta\to\theta'}(\gamma, K, K) \geq T_\theta - T_{\theta'} \geq IC_{\theta'\to\theta}(\gamma, K, K) \tag{4.10}$$

Figure 25 also plots two hypothetical values of $\Delta T = T_\theta - T_{\theta'}$ and $\Delta T' = T'_\theta - T'_{\theta'}$. Now, looking at ΔT, we can see that whenever $\gamma = \gamma_0$ then the mechanism that sets $K_\theta = K_{\theta'} = K$ and $T_\theta - T_{\theta'} = \Delta T$ satisfies the incentive constraints of the θ and θ' types for the value of ϵ illustrated in the figure. Namely, using transfers whose difference is equal to ΔT, then both conditions of (4.10) are satisfied for all $\gamma \in [\gamma_0 - \epsilon, \gamma_0 + \epsilon]$ as illustrated in Figure 25. On the other hand, when $\gamma = \gamma'_0 > \gamma_0$ then we can see from Figure 25 that the regulator cannot provide robust incentives for the same value of ϵ. Namely, whenever $\gamma = \gamma'_0 - \epsilon$ then it can be seen that $IC_{\theta'\to\theta}(\gamma, K, K) > \Delta T'$ and when $\gamma = \gamma'_0 + \epsilon$ then $IC_{\theta\to\theta'}(\gamma, K, K) < \Delta T'$. Further, no adjustment of $\Delta T'$ can guarantee that both of these conditions are satisfied for all values of $\gamma \in [\gamma_{0'} - \epsilon, \gamma'_0 + \epsilon]$.

4.4.1 Controlled Risk Taking

In this section we will now allow for a bank of type θ to choose any level of risk $\sigma_0 + \tilde{\theta}$ for all $\tilde{\theta} \in [0, \theta]$. We simply note that this case is a straight forward extension of the previous, only that there exists $\hat{\gamma} \in (\underline{\gamma}, \overline{\gamma})$ such that incentive compatibility is guaranteed between θ and θ' as in that case both types θ and θ' will optimally pool on the same level of risk $\tilde{\theta} \leq \theta'$ and therefore will be identical. Thus, the following corollary restates Proposition 4.4 and Proposition 4.6, simply noting that the upper bound of each proposition is lowered to the value at which the bank of type θ prefers to choose a level of risk at or below θ'.

Corollary 4.7. *Suppose that for any θ, the bank of type θ can choose any level of risk $\sigma_0 + \tilde{\theta}$ for all $\tilde{\theta} \in [0, \theta]$.*
(1) Consider any mechanism $\mathcal{M} := (K_\theta, T_\theta)_{\theta\in[0,\bar{\theta}]}$ with $K_\theta > K_{\theta'}$ for some $\theta > \theta'$. Then there exists an interval $(\underline{\gamma}, \hat{\gamma})$ such that $\mathcal{M}$ is not incentive compatible whenever $\gamma \in (\underline{\gamma}, \hat{\gamma})$.
(2) For any $\epsilon > 0$ there always exists $\underline{\gamma}(\epsilon) < \hat{\gamma}(\epsilon)$ such that no (γ_0, ϵ)-robust mechanism

exists whenever $\gamma_0 \in (\underline{\gamma}(\epsilon), \hat{\gamma}(\epsilon)$.

References

[1] Acharya, V. V., Schnabl, P., Suarez, G. (2013): "Securitization without Risk Transfer." *Journal of Financial Economics*, 515-536.

[2] Acharya, V., Mehran, H., Scheurmann, T., and Thakor, A. (2011): "Robust Capital Regulations." *Federal Reserve Bank of New York Staff Reports*, no. 490.

[3] Ashcraft, A., Goldsmith-Pinkham, P., Hull, Peter, and Vickery, J. (2011): "Credit Ratings and Security Prices in the Subprime MBS Market." *American Economic Review: Papers & Proceedings*, 101 (3), 115-119.

[4] Basel Committee on Banking Supervision (2013): "The regulatory framework: balancing risk sensitivity, simplicity and comparability." Discussion paper.

[5] Begley, T., Purnanandam, A., and Zheng, K. (2017): "The strategic under-reporting of bank risk." *Review of Financial Studies*, 30, 3376-415.

[6] Bergemann, D. and Morris, S. (2005): "Robust mechanism design." *Econometrica*, 73(6), 1771-1813.

[7] Blum, J. M. (1999): "Do Capital Adequacy Requirements Reduce Risks in Banking?" *Journal of Banking and Finance*, 23(5), 755-771.

[8] Carroll, G. (2015) "Robustness and Linear Contracts." *American Economic Review* 105 (2), 536-563.

[9] Carroll, G. (2018): "Robustness in Mechanism Design and Contracting." *Annual Review of Economics*, 11, 139-166.

[10] Chan, Y.-S., Greenbaum, S. I., and Thakor A. V. (1992): "Is Fairly Price Deposit Insurance Possible?" *The Journal of Finance*, 47(1), 227-245.

[11] Cochrane, J. H. (2017): "Macro-Finance." *Review of Finance*, 21(3), 945-985.

[12] Coudert, V. and Gex, M. (2008): "Does risk aversion drive financial crises? Testing the predictive power of empirical indicators." *Journal of Empirical Finance*, 15(2), 167-184.

[13] Gennotte, G. and Pyle, D. (1991): "Capital Controls and Bank Risk." *Journal of Banking and Finance*, 15, 805-824.

[14] Giammarino, R. M., Lewis, T. R., and Sappington D. E. M. (1993): "An Incentive Approach to Banking Regulation" *Journal of Finance*, 48(4), 1523-42.

[15] Haldane, A. and Madouros, V. (2012): "The dog and the frisbee." *Proceedings- Economics Policy Symposium- Jackson Hole*, 109-159.

[16] Hansen, L. P. and Sargent, T. J. (2001):"Robust Control and Model Uncertainty." *The American Economic Review*, 91, 60-66.

[17] Jones, D. (2000): "Emerging Problems with the Basel Capital Accord: Regulatory Capital Arbitrage and Related Issues." *Journal of Banking & Finance*, 35-58.

[18] Kahane, Y. (1977): "Capital Adequacy and the Regulation of Financial Intermediaries." *Journal of Banking and Finance*, 207-218.

[19] Kim, D. and Santomero, A. M. (1988): "Risk in Banking and Capital Regulation." *The Journal of Finance*, 43, 1219-33.

[20] Koehn, M. and Santomero, A. M. (1980): "Regulation of Bank Capital and Portfolio Risk." *Journal of Finance*, 35, 1235-44.

[21] Le Leslé, V. and Aramova, S. (2012): "Revisiting Risk-Weighted Assets." *IMF Working Paper*, 12/90.

[22] Madarász, K. and Prat, A. (2017): "Sellers with misspecified models." *Review of Economic Studies*, 84(2), 790-815.

[23] Mariathasan, M. and Merrouche, O. (2014): "The manipulation of Basel risk-weights: evidence from 2007-2010." *Journal of Financial Intermediation*, 23(3), 300-321.

[24] Merton, R. C. (1977): "An analytic derivation of the cost of deposit insurance and loan guarantees: An application of modern option pricing theory." *Journal of Banking & Finance*, 1 (1). 3-11.

[25] Kopecki, D. and Moore, M. J. (2013): "Whale of Trade Shown at Biggest U.S. Bank with Best Control" *Bloomberg Markets Magazine*, June 4.

[26] Myers, S., and Majluf, N. (1984): "Corporate Financing and Investment Decisions when Firms Have Information that Investors Do Not Have." *Journal of Financial Economics*, **13**, 187-221.

[27] Rajan, R. G. (2005): "Has Financial Development Made the World Riskier?" *NBER Working Paper No. 11728.*

[28] Rochet, J. C. (1992): "Capital Requirements and the Behavior of Commercial Banks. *European Economic Review 36*, 1137-1178.

[29] Vallascas, F. and Hagendorff, J. (2013): "The Risk Sensitivity of Capital Requirements: Evidence from and International Sample of Large Banks." *Review of Finance*, 17, 1947-1988.

[30] Yorulmazer, T. (2013): "Has Financial Innovation Made the World Risker? CDS, Regulatory Arbitrage and Systemic Risk." Working Paper: https://ssrn.com/abstract=2176493

4.5 Appendix

4.5.1 Reformulating the bank's problem

We will now reformulate the problem into a more tractable form. Namely, denoting by

$$r_{\gamma,\theta}(K) := \frac{V_0(f_\theta, K) - V_\gamma(f_\theta, K)}{K}$$

the per dollar of capital risk premium, we can see that the outside equity holder's break even condition can be reformulated as

$$\alpha_\theta \cdot V_0(f_\theta, K) = (1 + \alpha_\theta \cdot r_{\gamma,\theta}(K))K$$

and therefore the bank's problem can be reformulated as

$$\max_{\tau\in[0,\theta]} (1-\alpha_\tau) \cdot V_\gamma(f_\tau, K) = \max_{\tau\in[0,\theta]} V_0(f_\tau, K) - (1 + r_{\gamma,\theta}(K))K \tag{4.11}$$

This formulation will allow us to better understand the bank's incentives. Namely, note that

$$V_0(f_\theta, K) = \int_{-K}^{\infty} (x+K) f_\theta(x)dx = K + \int_{-K}^{\infty} x f_\theta(x)dx := K + \pi(f_\theta, K).$$

Here $\pi(f_\theta, K)$ denotes the return on equity K which has the familiar formulation:

$$\pi(f_\theta, K) = \int_{-\infty}^{\infty} x f_\theta(x)dx - \int_{-\infty}^{-K} x f_\theta(x)dx := \pi(f_\theta) + L(f_\theta, K)$$

Namely, the bank's return on equity is equal to the NPV of the project plus the value of the deposit insurance put (equal to the expected loss to the deposit insurance fund). Now, it is

straightforward to see that

$$\frac{\partial}{\partial K}V_0(f_\theta, K) = 1 - F(-K) > 0$$

so that the value of bank equity is increasing in capital. Yet, the value of inside equity is such that

$$\frac{\partial}{\partial K}[V_0(f_\tau, K) - (1+r_{\gamma,\theta}(K))K] = \frac{\partial}{\partial K}[L(f_\theta, K) - r_{\gamma,\theta}(K)K] = \frac{\partial}{\partial K}L(f_\theta, K) - r_{\gamma,\theta}(K) - K\cdot\frac{\partial}{\partial K}r_{\gamma,\theta}(K)$$

which is negative for small values of γ and positive for large values of γ. Further, it can be observed that for any value of γ, when K is small enough then the value of inside equity is decreasing in newly raised capital K. Further, the relationship between the value of inside equity and K is such that for K small inside equity is decreasing in K and for K large, inside equity is increasing in K. Given that the problem of capital regulation will not be interesting (nor practically relevant) whenever the capital requirement does not bind, we make the following assumption which states that the levels of first best capital requirements and relevant region of risk aversion are such that inside shareholders dislike raising equity (so that capital requirements bind).

4.5.2 Lemma 4.8 and Lemma 4.9

Lemma 4.8. *For any K there exists $\gamma_1 > \gamma(K)$ such that*
1.) $\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'}) < \frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K)$ *whenever* $\gamma < \gamma_1$.
2.) $\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'}) > \frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K)$ *whenever* $\gamma > \gamma_1$.
3.) $|\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'}) - \frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K)|$ *is decreasing in γ whenever $\gamma < \gamma_1$ and increasing in γ whenever $\gamma > \gamma_1$.*

Proof. First note that

$$\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'}) = \frac{\partial}{\partial K}V_\gamma(f_\theta, K) - 1 < 0$$

where the last inequality comes from Assumption 4.3. Next note that

$$\frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K) = \frac{\partial}{\partial K}V_\gamma(f_{\theta'}, K) - \frac{\partial}{\partial K}[K\frac{V_\gamma(f_{\theta'}, K) - T_\theta}{V_\gamma(f_\theta, K) - T_\theta}]$$

and

$$\frac{\partial}{\partial K}[K\frac{V_\gamma(f_{\theta'}, K)}{V_\gamma(f_\theta, K)}] = \frac{V_\gamma(f_{\theta'}, K) - T_\theta}{V_\gamma(f_\theta, K) - T_\theta} + K(\frac{(V_\gamma(f_\theta, K) - T_\theta)\frac{\partial}{\partial K}V_\gamma(f_{\theta'}, K) - (V_\gamma(f_{\theta'}, K) - T_\theta)\frac{\partial}{\partial K}V_\gamma(f_\theta, K)}{(V_\gamma(f_\theta, K) - T_\theta)^2}).$$

Further using this expression and rearranging we obtain

$$\frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K) = (\frac{\partial}{\partial K}V_\gamma(f_{\theta'}, K) - 1)(1 - \frac{K}{V_\gamma(f_\theta, K) - T_\theta})+$$

$$(\frac{\partial}{\partial K}V_\gamma(f_\theta, K)-1)\frac{K}{V_\gamma(f_\theta, K) - T_\theta}\cdot\frac{V_\gamma(f_{\theta'}, K) - T_\theta}{V_\gamma(f_\theta, K) - T_\theta}+(1-\frac{V_\gamma(f_{\theta'}, K) - T_\theta}{V_\gamma(f_\theta, K) - T_\theta})(1-\frac{K}{V_\gamma(f_\theta, K) - T_\theta})$$

Now, we know that when $\gamma = \gamma(K)$, then $V_\gamma(f_\theta, K) = V_\gamma(f_{\theta'}, K)$. Hence, whenever $\gamma = \gamma(K)$ then

$$\frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K) = (\frac{\partial}{\partial K}V_\gamma(f_{\theta'}, K)-1)(1-\frac{K}{V_\gamma(f_\theta, K) - T_\theta})+(\frac{\partial}{\partial K}V_\gamma(f_\theta, K)-1)\frac{K}{V_\gamma(f_\theta, K) - T_\theta}$$

Further, given that $\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'}) = \frac{\partial}{\partial K}V_\gamma(f_\theta, K) - 1 < 0$, $V_\gamma(f_\theta, K) - T_\theta > K$, and $\frac{\partial}{\partial K}V_\gamma(f_{\theta'}, K) > \frac{\partial}{\partial K}V_\gamma(f_\theta, K)$ we can see that

$$\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'}) > \frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K)$$

whenever $\gamma = \gamma(K)$. Now, note that when $\gamma < \gamma(K)$, the term

$$(1 - \frac{V_\gamma(f_{\theta'}, K) - T_\theta}{V_\gamma(f_\theta, K) - T_\theta})(1 - \frac{K}{V_\gamma(f_\theta, K) - T_\theta}) \tag{4.12}$$

is strictly positive and decreasing in γ. Further, $\frac{V_\gamma(f_{\theta'}, K)-T_\theta}{V_\gamma(f_\theta, K)-T_\theta}$ (the additional term that disappears when $\gamma = \gamma(K)$) is increasing in γ and therefore given that this is multiplied by the negative term $\frac{\partial}{\partial K}V_\gamma(f_\theta, K) - 1$ we can see that $\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'}) - \frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K)$ is decreasing in γ up to $\gamma(K)$. Similarly, once $\gamma > \gamma(K)$ the term (4.12) becomes negative and decreasing in γ and the term $\frac{V_\gamma(f_{\theta'}, K)-T_\theta}{V_\gamma(f_\theta, K)-T_\theta}$ becomes greater than 1 and increasing in γ. Hence, there must be some value γ_1 such that $\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'}) = \frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K)$ and $\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'}) < \frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K)$ whenever $\gamma > \gamma_1$.

An immediate corollary of this last point is that $|\frac{\partial}{\partial K}IC_{\theta\to\theta'}(\gamma, K, K_{\theta'})-\frac{\partial}{\partial K}IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K)|$ is decreasing in γ when $\gamma < \gamma_1$ and increasing in γ when $\gamma > \gamma_1$. □

Lemma 4.9. *For any K and any transfers satisfying $\mathcal{IC}(\gamma, \theta, \theta'|\mathcal{M})$ when $K_\theta = K_{\theta'} = K$, there exists $\hat{\gamma}(K)$ such that*

1.) $IC_{\theta\to\theta'}(\hat{\gamma}(K), K, K) = IC_{\theta'\to\theta}(\hat{\gamma}(K), K, K)$

2.) $\frac{\partial}{\partial\gamma}IC_{\theta\to\theta'}(\gamma, K, K) < \frac{\partial}{\partial\gamma}IC_{\theta'\to\theta}(\gamma, K, K)$ whenever $\gamma < \hat{\gamma}(K)$ and $\frac{\partial}{\partial\gamma}IC_{\theta\to\theta'}(\gamma, K, K) > \frac{\partial}{\partial\gamma}IC_{\theta'\to\theta}(\gamma, K, K)$ whenever $\gamma > \hat{\gamma}(K)$.

Proof. 1.) $\hat{\gamma}(K)$ is the value of γ such that $V_\gamma(f_\theta, K) = V_\gamma(f_{\theta'}, K)$ in which case $IC_{\theta\to\theta'}(\hat{\gamma}(K), K, K) = IC_{\theta'\to\theta}(\hat{\gamma}(K), K, K) = 0$.

2.) Note that

$$\frac{\partial}{\partial\gamma}IC_{\theta\to\theta'}(\gamma,K,K)=\frac{\partial}{\partial\gamma}K(\frac{V_\gamma(f_\theta,K)-T_{\theta'}}{V_\gamma(f_{\theta'},K)-T_{\theta'}})=$$

$$\frac{(V_\gamma(f_{\theta'},K)-T_{\theta'})\frac{\partial}{\partial\gamma}V_\gamma(f_\theta,K)-(V_\gamma(f_\theta,K)-T_{\theta'})\frac{\partial}{\partial\gamma}V_\gamma(f_{\theta'},K)}{(V_\gamma(f_\theta,K)-T_{\theta'})^2}$$

while

$$\frac{\partial}{\partial\gamma}IC_{\theta'\to\theta}(\gamma,K,K)=-\frac{\partial}{\partial\gamma}K(\frac{V_\gamma(f_{\theta'},K)-T_{\theta}}{V_\gamma(f_{\theta},K)-T_{\theta}})=$$

$$\frac{(V_\gamma(f_{\theta'},K)-T_{\theta})\frac{\partial}{\partial\gamma}V_\gamma(f_\theta,K)-(V_\gamma(f_\theta,K)-T_{\theta})\frac{\partial}{\partial\gamma}V_\gamma(f_{\theta'},K)}{(V_\gamma(f_{\theta'},K)-T_{\theta})^2}$$

Now note that any mechanism that sets $K_\theta = K_{\theta'} = K$ must be such that $T_\theta = T_{\theta'}$ whenever $\gamma = \gamma(K)$. Further, this implies that our result holds for $T_\theta = T_{\theta'}$ as in that case $\frac{\partial}{\partial\gamma}IC_{\theta\to\theta'}(\gamma,K,K) < \frac{\partial}{\partial\gamma}IC_{\theta'\to\theta}(\gamma,K,K)$ whenever $\gamma < \gamma(K)$ and $\frac{\partial}{\partial\gamma}IC_{\theta\to\theta'}(\gamma,K,K) > \frac{\partial}{\partial\gamma}IC_{\theta'\to\theta}(\gamma,K,K)$ whenever $\gamma > \gamma(K)$. This comes from the fact that the numerators of $\frac{\partial}{\partial\gamma}IC_{\theta\to\theta'}(\gamma,K,K)$ and $\frac{\partial}{\partial\gamma}IC_{\theta'\to\theta}(\gamma,K,K)$ are the same when $T_\theta = T_{\theta'} = T$ and the denominators are $V_\gamma(f_\theta,K)-T$ and $V_\gamma(f_{\theta'},K)-T$ respectively. Hence, given that $\gamma > \gamma(K)$ implies $V_\gamma(f_\theta,K) > V_\gamma(f_{\theta'},K)$ and vice-versa, we obtain our result.

Now, if $\mathcal{M}$ is incentive compatible and $K_\theta = K_{\theta'} = K$, then from the previous step we know that $T_\theta = T_{\theta'}$ when $\gamma = \gamma(K)$. Further, as γ decreases, we know that both $IC_{\theta\to\theta'}(\gamma,K,K)$ and $IC_{\theta'\to\theta}(\gamma,K,K)$ increase as $V_\gamma(f_\theta,K)$ is decreasing in γ faster than $V_\gamma(f_{\theta'},K)$. Therefore, it must be the case that $T_\theta > T_{\theta'}$ for any incentive compatible mechanism with $\gamma < \gamma(K)$. Further, note that if T_θ increases, then the numerator of $\frac{\partial}{\partial\gamma}IC_{\theta'\to\theta}(\gamma,K,K)$ increases as $\frac{\partial}{\partial\gamma}V_\gamma(f_{\theta'},K) - \frac{\partial}{\partial\gamma}V_\gamma(f_\theta,K) > 0$. Further, the numerator decreases when T_θ increases and therefore $\frac{\partial}{\partial\gamma}IC_{\theta'\to\theta}(\gamma,K,K)$ increases. Further, by the same logic we can see that when $T_{\theta'}$ decreases then $IC_{\theta\to\theta'}(\gamma,K,K)$ decreases. Therefore, the results still hold when $\gamma < \gamma(K)$ under any incentive compatible mechanism. Finally, note that when $\gamma > \gamma(K)$ then by the same logic, incentive compatibility requires $T_\theta < T_{\theta'}$ and the converse effects hold as T_θ decreases or $T_{\theta'}$ increases.

□

4.5.3 Proof of Proposition 4.4

Proof. In order to prove this result, first note that a necessary condition for incentive compatibility is that when $K_\theta = K_{\theta'}$, the difference in the payoff of the θ type when it reports truthfully rather than imitating the θ' type must be strictly greater than difference in the payoff of the θ' type when it imitates the θ type as opposed to reporting truthfully. In order to see why this is the case, note that the incentive compatibility constraint (4.5) after

rearranging becomes:

$$V_\gamma(f_\theta, K_\theta) - V_\gamma(f_\theta, K_{\theta'}) + K_{\theta'} \cdot \frac{V_\gamma(f_\theta, K_{\theta'}) - T_{\theta'}}{V_\gamma(f_{\theta'}, K_{\theta'}) - T_{\theta'}} - K_\theta \geq T_\theta - T_{\theta'} \tag{4.13}$$

and similarly, writing down the incentive compatibility constraint of the θ' type to report truthfully as opposed to reporting that it is the θ type and rearranging, we obtain:

$$T_\theta - T_{\theta'} \geq V_\gamma(f_{\theta'}, K_\theta) - V_\gamma(f_{\theta'}, K_{\theta'}) + K_{\theta'} - K_\theta \frac{V_\gamma(f_{\theta'}, K_\theta) - T_\theta}{V_\gamma(f_\theta, K_\theta) - T_\theta} \tag{4.14}$$

Incentive compatibility therefore requires that the LHS of (4.13) is greater than the RHS of (4.14). Further, if this equation holds when $K_\theta > K_{\theta'}$ then it must hold with strict inequality when $K_\theta = K_{\theta'} = K$. Therefore, substituting $K_\theta = K_{\theta'} = K$ and rearranging, we obtain a necessary condition for incentive compatibility for any mechanism when $K_\theta > K_{\theta'}$ which is

$$\frac{V_\gamma(f_\theta, K) - V_\gamma(f_{\theta'}, K)}{V_\gamma(f_{\theta'}, K) - T_{\theta'}} > \frac{V_\gamma(f_\theta, K) - V_\gamma(f_{\theta'}, K)}{V_\gamma(f_\theta, K) - T_\theta}$$

Now, noting that $V_0(f_\theta, K) - V_0(f_{\theta'}, K) = L(f_\theta, K) - L(f_{\theta'}, K) > 0$ we can see that this condition can always be satisfied for the right transfers (e.g. $T_\theta = T_{\theta'}$) whenever $\gamma > 0$. But, as γ increases, $V_\gamma(f_\theta, K) - V_\gamma(f_{\theta'}, K)$ is strictly decreasing in γ. Hence, there must exist $\bar{\gamma}$ such that $V_{\bar{\gamma}}(f_\theta, K) = V_{\bar{\gamma}}(f_{\theta'}, K)$, violating the strict inequality for all $\gamma \geq \bar{\gamma}$.

Now, consider the case where $K_\theta > K_{\theta'} = K$ what we would like to prove is that there exists $\underline{\gamma} < \overline{\gamma}$ such that the incentive compatibility conditions are violated whenever $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. In order to do this, note that it still must be the case that $IC_{\theta \to \theta'}(\gamma, K_\theta, K_{\theta'}) \geq IC_{\theta \to \theta'}(\gamma, K_\theta, K_{\theta'})$ and after rearranging this expression, we obtain

$$(V_\gamma(f_\theta, K_\theta) - V_\gamma(f_{\theta'}, K_\theta))(1 - \frac{K_\theta}{V_\gamma(f_\theta, K_\theta) - T_{\theta'}}) \geq (V_\gamma(f_\theta, K_{\theta'}) - V_\gamma(f_{\theta'}, K_{\theta'}))(1 - \frac{K_{\theta'}}{V_\gamma(f_{\theta'}, K_{\theta'}) - T_\theta}) \tag{4.15}$$

Further, by Assumption 4.3 we know that for all relevant values of γ we have $V_\gamma(f_\theta, K_\theta) - V_\gamma(f_{\theta'}, K_\theta) < V_\gamma(f_\theta, K_{\theta'}) - V_\gamma(f_{\theta'}, K_{\theta'})$. Therefore, denoting by $\gamma(K)$ the value of γ such that $V_\gamma(f_\theta, K) - V_\gamma(f_{\theta'}, K) = 0$ we can see that it must be the case that $\gamma(K_{\theta'}) > \gamma(K_\theta)$. Therefore, when $\gamma = \gamma(K_\theta)$ it must be the case that the left hand side of (4.15) is equal to zero while the right hand side of (4.15) is strictly greater than 0, violating the inequality. Therefore, by continuity there must exist $\underline{\gamma}$ and $\overline{\gamma}$ such that $\underline{\gamma} < \gamma(K_\theta) < \gamma(K_{\theta'}) < \overline{\gamma}$ and $\mathcal{M}$ is not incentive compatible whenever $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. □

4.6 Proof of Proposition 4.6

Proof. We will first prove that for any $\theta > \theta'$ and any (θ', θ)-revealing mechanism $\mathcal{M}$ with $K_\theta = K_{\theta'} = K$, there exists $\underline{\gamma}$ and $\overline{\gamma}$ such that $\mathcal{M}$ is not (γ_0, ϵ)-robust whenever $\gamma_0 \in [\underline{\gamma}, \overline{\gamma}]$. In order to prove this, note that

$$IC_{\theta \to \theta'}(\gamma(K), K, K) = IC_{\theta' \to \theta}(\gamma(K), K, K) = 0$$

where $\gamma(K)$ is the value such that $V_{\gamma(K)}(f_\theta, K) = V_{\gamma(K)}(f_{\theta'}, K)$. Therefore, if $\mathcal{M}$ is incentive compatible when $\gamma = \gamma(K)$ then it must be the case that $T_\theta = T_{\theta'}$. Now, from Lemma 4.9 we know that as γ decreases below $\gamma(K)$, then $IC_{\theta \to \theta'}(\gamma, K, K)$ increases by more than $IC_{\theta' \to \theta}(\gamma, K, K)$ whenever transfers are set to ensure incentive compatibility. Therefore, $IC_{\theta \to \theta'}(\gamma, K, K) > IC_{\theta' \to \theta}(\gamma, K, K)$ for all $\gamma < \gamma(K)$. Further, the fact that both functions are increasing implies both are greater than zero when $\gamma < \gamma(K)$ and hence $T_\theta - T_{\theta'} > 0$ is necessary for incentive compatibility. Similarly, whenever $\gamma > \gamma(K)$ both functions decrease below zero and therefore it must be the case that $T_\theta - T_{\theta'} < 0$. Hence, for any $\epsilon > 0$, any $(\gamma(K), \epsilon)$-robust mechanism requires $T_\theta - T_{\theta'} = 0$, $T_\theta - T_{\theta'} < 0$, and $T_\theta - T_{\theta'} > 0$, a contradiction. What this states is that no $(\gamma(K), \epsilon)$-robust mechanism exists for all $\epsilon > 0$. Further, by the same logic we can conclude that for any $\epsilon > 0$, no mechanism can be (γ_0, ϵ)-robust for all $\gamma_0 \in [\gamma(K) - \epsilon, \gamma(K) + \epsilon]$ as $\gamma_0 = \gamma(K) - \epsilon$ requires $T_\theta - T_{\theta'} > 0$, $\gamma_0 = \gamma(K)$ requires $T_\theta - T_{\theta'} = 0$, and $\gamma_0 = \gamma(K) + \epsilon$ requires $T_\theta - T_{\theta'} < 0$.

Now suppose instead that $\mathcal{M}$ is (θ, θ')-risk sensitive. Now, we know that $IC_{\theta \to \theta'}(\gamma(K_{\theta'}), K_{\theta'}, K_{\theta'}) = IC_{\theta' \to \theta}(\gamma(K_{\theta'}), K_{\theta'}, K_{\theta'})$. Further, Lemma 4.8 tells us that there exists some $\overline{\gamma} > \gamma(K_{\theta'})$ such that whenever $\gamma < \overline{\gamma}$ then $\frac{\partial}{\partial K} IC_{\theta \to \theta'}(\gamma, K, K_{\theta'}) < \frac{\partial}{\partial K} IC_{\theta \to \theta'}(\gamma, K, K_{\theta'})$. What this states is that the decrease in $IC_{\theta \to \theta'}(\gamma, K, K_{\theta'})$ is greater than the decrease in $IC_{\theta' \to \theta}(\gamma, K_{\theta'}, K)$ whenever $\gamma < \overline{\gamma}$. Importantly, this implies that for any $K_\theta > K_{\theta'}$ it must be the case that $IC_{\theta \to \theta'}(\gamma(K_{\theta'}), K_\theta, K_{\theta'}) < IC_{\theta' \to \theta}(\gamma(K_{\theta'}), K_{\theta'}, K_\theta)$ and therefore the mechanism is no longer incentive compatible whenever $\gamma = \gamma(K_{\theta'})$. Now, if the capital requirements K_θ and $K_{\theta'}$ are incentive compatible for some value of $\gamma < \gamma(K_{\theta'})$ then we know that there must exist a value $\underline{\gamma} < \gamma(K_{\theta'})$ such that $IC_{\theta \to \theta'}(\underline{\gamma}, K_\theta, K_{\theta'}) = IC_{\theta' \to \theta}(\underline{\gamma}, K_{\theta'}, K_\theta)$ and $IC_{\theta \to \theta'}(\gamma, K_\theta, K_{\theta'}) > IC_{\theta' \to \theta}(\gamma, K_{\theta'}, K_\theta)$ whenever $\gamma < \underline{\gamma}$. Further, that Lemma 4.8 also tells us that whenever $\gamma = \overline{\gamma}$, then $IC_{\theta \to \theta'}(\overline{\gamma}, K_\theta, K_{\theta'}) = IC_{\theta' \to \theta}(\overline{\gamma}, K_{\theta'}, K_\theta)$ and whenever $\gamma > \overline{\gamma}$ then $IC_{\theta \to \theta'}(\gamma, K_\theta, K_{\theta'}) > IC_{\theta' \to \theta}(\gamma, K_{\theta'}, K_\theta)$. In summary, for any $K_\theta > K_{\theta'}$ there

exists $\underline{\gamma} < \overline{\gamma}$ such that

$$IC_{\theta\to\theta'}(\gamma, K_\theta, K_{\theta'}) > IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K_\theta) \text{ whenever } \gamma < \underline{\gamma} \tag{4.16}$$

$$IC_{\theta\to\theta'}(\gamma, K_\theta, K_{\theta'}) = IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K_\theta) \text{ whenever } \gamma = \underline{\gamma} \tag{4.17}$$

$$IC_{\theta\to\theta'}(\gamma, K_\theta, K_{\theta'}) < IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K_\theta) \text{ whenever } \gamma \in (\underline{\gamma}, \overline{\gamma}) \tag{4.18}$$

$$IC_{\theta\to\theta'}(\gamma, K_\theta, K_{\theta'}) = IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K_\theta) \text{ whenever } \gamma = \overline{\gamma} \tag{4.19}$$

$$IC_{\theta\to\theta'}(\gamma, K_\theta, K_{\theta'}) > IC_{\theta'\to\theta}(\gamma, K_{\theta'}, K_\theta) \text{ whenever } \gamma > \overline{\gamma} \tag{4.20}$$

Therefore, we know that no mechanism with capital requirements $K_\theta > K_{\theta'}$ is incentive compatible whenever $\gamma \in (\underline{\gamma}, \overline{\gamma})$. Further, as in the case when $K_\theta = K_{\theta'}$ if there exists a (γ_0, ϵ)-robust mechanism with capital requirements $K_\theta > K_{\theta'}$ then it requires that $\gamma_0 < \underline{\gamma} - \epsilon$ or $\gamma_0 > \overline{\gamma} + \epsilon$ as robustness requires $T_\theta - T_{\theta'} = 0$ when $\gamma = \underline{\gamma}$ or $\gamma = \overline{\gamma}$ but $T_\theta - T_{\theta'} > 0$ whenever $\gamma = \underline{\gamma} - \epsilon$ and $T_\theta - T_{\theta'} < 0$ whenever $\gamma = \overline{\gamma} + \epsilon$.

□

4.7 Proof of Corollary 4.7

Proof. Note that, denoting by $\gamma(K)$ the value of γ introduced in the proofs of Proposition 4.4 and 4.6, then whenever $\gamma > \gamma(K_{\theta'})$, choosing a value $\tilde{\theta} > \theta'$ is suboptimal for both types θ and θ'. Hence, any mechanism that sets $K_\theta > K_{\theta'}$ is guaranteed to ensure that both types θ and θ' choose some level of risk $\tilde{\theta} \leq \theta'$ and therefore $\mathcal{IC}(\gamma, \theta, \theta')$ are trivially satisfied as there are effectively no θ types. Therefore, we simply replace $\overline{\gamma}$ from Proposition 4.4 with $\hat{\gamma} := \gamma(K_{\theta'})$ and $\overline{\gamma}(\epsilon)$ with $\hat{\gamma}(\epsilon) := \gamma(K_{\theta'}) + \epsilon$. □

INSTITUT POLYTECHNIQUE DE PARIS

Titre : Sujet sur la théorie des jeux et conception de méchanism appliquée à la Finance et à la réglementation bancaire.

Mots clés : Théorie des jeux, conception de méchanism, réglementation bancaire

Résumé : Cette thèse développe de nouveaux outils de théorie des jeux et mechanism design pour de multiples application en économie/finance. Le premier chapitre étudie la possibilité d'implémentation d'équilibres de communication dans le cadre de jeux stratégiques lorsque tous les joueurs de réseau peuvent communiquer par l'intermédiaire d'un médiateur impartial. Je dérive les conditions nécessaires et suffisantes sur la structure du réseau de joueurs telles que, pour tout jeu, tout équilibre de communication puisse être implémenté. Le deuxième chapitre propose un modèle d'encombrement de la chaine de production dans lequel les contraintes de capacité produisent de multiples équilibres de Nash Pareto-inefficients. Ce chapitre montre comment l'utilisation d'équilibres corrélés peut résoudre de manière substantielle ces inefficiences. Les deux dernier chapitres traitent de questions liées à la conception des exigences de fonds propres de banques. Dans le chapitre 3, on caractérise les exigences optimales de fonds propres des banques lorsque celles-ci disposent d'informations privées sur la valeur de leurs actifs existants. On montre comment l'implémentation des exigences de fonds propres peut éliminer le coût de l'augmentation de capital pour la banque en révélant ses informations au marché, et les conditions dans lesquelles ce transfert d'informations est optimal. Dans le chapitre 4, on fait l'hypothèse que les banques possèdent de l'information privée sur le risque de leurs actifs plutôt que sur leur valeur. Dans c ecas, si les investisseurs sont suffisamment averses au risque, on montre que n'importe quelle exigence de fonds subordonée au risque des banques incitent ces derniéres à mentir sur leur niveau de risque effectif. Ce résultat met em lumière d'importants problèmes de robustesse.

Title : Essays on Game Theory, Mechanism Design, and Financial Economics

Keywords : Game Theory, Mechanism Design, Bank Regulation

Abstract : This thesis develops and utilizes tools in game theory and mechanism design to study multiple applications in economics and finance. The first chapter studies the problem of implementing communication equilibria of strategic games when players communicate with an impartial mediator through a network. I characterize necessary and sufficient conditions on the network structure such that any communication equilibrium of any game can be implemented on that network. The next chapter studies a model of supply chain congestion whereby capacity constraints lead to very inefficient Nash equilibria and I show how the use of correlsted equilibria can substantially resolve those inefficiencies. The final two chapters study related issues in the design of bank capital requirements. In Chapter 3, I characterize optimal bank capital requirements when banks have private information about the value of their existing assets. I show how the implementation of capital requirements can eliminate the bank's cost of raising capital by revealing their information to the market and conditions under which doing soi s optimal. In Chapter 4, I show how when the bank's private information is about the riskiness of its assets instead, then any risk sensitive capital requirement will lead banks to optimally misreport their risk whenever investors are sufficiently risk averse, highlighting important robustness concerns.