Exploring Modifications to the Perception of Economic Agents: Impartial, Farsighted, and Optimistic Perspectives

Inauguraldissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften (Dr. rer. pol.) an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld

Stefan Berens

June 2019

First Supervisor: Prof. Dr. Herbert Dawid

Second Supervisor: J.-Prof. Dr. Jan-Henrik Steg

Abstract. In this thesis, we explore three modifications to the perception of economic agents. Our work introduces impartial, farsighted, and optimistic perspectives to economic frameworks about decision-making under uncertainty, the formation of global trade agreements, and the funding of entrepreneurs respectively.

First, we extend Harsanyi's Impartial Observer Theorem by introducing Knightian Uncertainty in the form of individual belief systems. It features an axiomatic framework of societal decision-making in the presence of individual uncertainty. The model allows the analysis of scenarios where individuals agree on the ranking but not on the likelihood of social outcomes. The preferences of the impartial observer are represented by a weighted sum of utilities - each representing individual preferences with different belief systems. In order to incorporate common criticism of the framework of Harsanyi (1953), our approach is based on the generalized version by Grant et al. (2010). The belief systems are introduced as second-order beliefs following Seo (2009).

Second, we study and compare the stability of trade policy arrangements in two different regulatory scenarios, one with and one without Preferential Trade Agreements (PTAs), i.e. current vs. modified WTO rules. Unlike the existing literature, our work considers an extensive choice set of trade constellations, containing both available PTAs, Customs Unions (CUs) and Free Trade Agreements (FTAs), as well as Multilateral Trade Agreements (MTAs), while assuming unlimited farsightedness of the negotiating parties. With symmetric countries and under both the current and the modified WTO rules, the Global Free Trade (GFT) regime emerges as the unique stable outcome. In the case of asymmetry, the results are driven by the relative size of the countries. If the world is in the vicinity of symmetry and two out of three countries are close to identical while relatively smaller than the other one, the area where the GFT regime is stable increases when prohibiting PTAs. However, when two similar countries are relatively larger, the availability of PTAs is conducive to the stability of the GFT regime. Finally, if the world is further away from symmetry, full trade liberalization is not attainable at all and an area where the Most-Favoured-Nation (MFN) regime is stable appears in the scenario without PTAs. Thus, the direction of the effect of PTAs on trade liberalization depends on the degree of asymmetry among countries.

Third, we consider a project of unknown quality that is available to an entrepreneur and requires both funding and effort to be realized. The entrepreneur is either of optimistic or realistic type and accordingly receives a distorted signal about the quality of the project. The entrepreneur is aware of the potential bias but ultimately believes to be a realist. In terms of funding, the entrepreneur has access to two different sources of funding, namely a bank and a venture capitalist. The banks only contribution is the funding, while the venture capitalist is on top of that able to invest effort into the project. Similarly, the entrepreneur can provide additional effort both in advance and simultaneously with the venture capitalist. Our work ultimately characterizes different constellations of the model parameters under which an increase in the share of overconfidence leads to an increase in social welfare.

Contents

Figures

Acknowledgement

All in all, about a decade of study culminated in this thesis - graciously supported by a host of people and institutions along the way. I am grateful to all of you:

I want to express my deepest gratitude to my advisor, Herbert Dawid, for the extensive support of all of my research projects that now led to this thesis. Especially, I want to thank you for your patience, inspiration, and knowledge that guided me on this journey. I knew I could always count on you and your advice, irrespective of the problem. I also want to extend my gratitude to my co-advisor, Jan-Henrik Steg, for all the constructive discussions and sincere assistance. Whenever I needed to, you would always find time to talk with me about my work.

Apart from my two thesis advisors, I also want to express my sincere gratitude to Gerald Willmann for all the insightful discussions and candid guidance. Additionally, I am grateful to Frank Riedel for encouraging the joint work with Lasha Chochua and providing us with invaluable advice and opportunities.

I am particularly grateful to my co-author and dear friend, Lasha Chochua, who accompanied me on this journey from start to finish. I always cherished our debates. I simply could not have hoped for a better partner to work with.

Furthermore, I am grateful to the following people for supporting this thesis through various fruitful discussions: Thibault Gajdos, Simon Grant, Luca Rigotti, Michael Chwe, and Mauro Napoletano. I also want to extend my gratitude to all the others at numerous conferences who influenced my work via valuable feedback.

I am grateful to the Bielefeld Graduate School of Economics and Management (BiGSEM) of Bielefeld Universit for funding my research and supporting my projects. Moreover, I am grateful to everyone else at Bielefeld University who supported me one way or the other.

Finally, I am incredibly grateful to my family and to all my friends for the words of encouragement on this journey. In particular, it is safe to say that this thesis would not have been possible without the endless and unconditional support of my parents and grandparents: Birgit and Joachim Berens, and Karin and Karl Fuchs. You should know that I am grateful for all the opportunities you provided me with.

Introduction

Every economic model implicitly, sometimes explicitly, contains other models, which describe the world view of the involved economic agents. It is those models within a model that ultimately determine the behavior of the agents. In this thesis, we employ already established economic frameworks and then study modifications to the perception of the corresponding economic agents. In the coming three parts, we move from a societal to an individual level, touch on different areas of research, but always focus on the underlying world view of the different economic agents. Specifically, we introduce an impartial, a farsighted, and an optimistic perspective. Thematically, we examine decision-making under uncertainty, the formation of global trade agreements, and the funding of entrepreneurs respectively. Finally, this thesis falls into the broad category of economic theory.

Context and Overview

In 2005 a team of four U.S. Navy SEALs set out to find a Taliban leader in the Afghan mountains near the Pakistan border. Just as the team set up their base overlooking the area to fulfill their reconnaissance mission, two Afghan goatherds stumbled upon them - a young boy with them. Due to the nature of their mission and other circumstances, the team considered killing or releasing the civilians the only two viable options. Eventually, the team cast a vote, where one soldier abstained and two voted two ways. The commander of the unit then made the decisive call to release them. The civilians later informed the Taliban in a nearby village about the presence of the soldiers. In the subsequent ambush three of the four soldiers died, leaving the commander as the lone survivor.

The story of the 'Afghan Goatherds' (as shown in Sandel (2010)) presents us with a factual not a hypothetical moral dilemma, which stands in contrast to the famous trolley problem (as formulated by Foot (1967)). As such, it raises the question which theoretical model adequately captures the essence of this type of practical problem.

Chapter 1 presents a potential answer to this question by providing a model for normative decision-making in scenarios where a group of individuals faces uncertainty. It introduces an impartial perspective to general decision-making under uncertainty based on the theory of societal judgments developed by Harsanyi (1953, 1955, 1977). In the tradition of various other economists (and philosophers), Harsanyi argues that individuals should make such value judgments about social alternatives from the perspective of an 'impartial observer'.[1] As such, you judge different arrangements using the personal preferences of all members of society simultaneously, essentially imagining yourself independent of your own identity. From this idea then followed

[1] A tradition dating back as far as Adam Smith's 'Theory of Moral Sentiments' (1759), which also includes John Rawls' concept of a 'veil of ignorance' in 'A Theory of Justice' (1971).

Harsanyi's Impartial Observer Theorem, which states that when an individual faces risky prospects over social outcomes it should rank these as a weighted utilitarian, i.e. according to the weighted aggregate of individuals' expected utilities, where the weights follow from a supposed lottery over societal identities. In contrast to this framework, where objective probabilities over the set of social outcomes are known by each member of society, our approach provides an extension to scenarios with subjective belief systems about the likelihood of the social outcomes. Here, the impartial observer necessarily takes these individual belief systems into consideration (via the personal preferences) but without aggregating them in the process.

The presented framework provides an axiomatic foundation for the extension of Harsanyi's Impartial Observer Theorem via Knightian uncertainty. It is based on the generalized version of Harsanyi (1953) by Grant et al. (2010), which addresses the issues of fairness and attitude towards mixing of the original approach through the use of additional functions that transform the individual utility. Using informal notation, Harsanyi's representation $\sum_{Indiv\ i} Prob(i) ExpUtil_i(SocialAlternative)$ thereby became Grant et al.'s $\sum_{Indiv\ i} Prob(i)\phi_i(ExpUtil_i(SocialAlternative))$. Note that both of these models employ von Neumann-Morgenstern expected utility, which necessarily requires a change in order to accommodate Knightian uncertainty. Specifically, the introduction of the individual belief systems to the framework follows Seo (2009), who formulated a model for decision-making under uncertainty using so-called second-order belief systems, i.e. beliefs over probability measures. It allows the introduction of Knightian uncertainty to the framework without any additional modifications or extensive assumptions.[2] Ultimately, the representation of the preferences of the impartial observer under uncertainty takes the form of a weighted aggregate of the individuals' (transformed) second-order subjective expected utilities, i.e. the nature of a weighted utilitarian remains. Alternatively, $\sum_{Indiv\ i} Prob(i)\phi_i(SecOrdSubjExpUtil_i(SocialAlternative))$ when (ab-)using the informal notation from before.

By extending the (generalized) framework to a new type of moral value judgments, it is possible to analyze situations where individuals agree on the ranking but not on the likelihood of social outcomes. The story of the 'Afghan Goatherds' illustrates such a scenario that is purely driven by individual belief systems. In particular, the accompanying exemplary comparison of subjective risk and uncertainty provides further justification for the introduction of uncertainty to the framework. Moreover, an example of an exchange economy (inspired by Eichberger and Pethig (1994)) showcases the framework's suitability for traditional economic problems by comparing two alternative modes of wealth (re-)distribution, namely the Walrasian auctioneer and the Egalitarian rule. It features a scenario where the degree of fairness of the impartial observer, captured via the utility transforming functions, ultimately determines the overall ranking - contributing another example to the discussion on the issue of fairness.

As indicated before, from a certain point of view this framework introduces uncertainty to decision-making with an impartial observer, but alternatively it adds an impartial observer to decision-making under uncertainty. In that regard, the model represents the first modification to the perception of economic agents. Note that at its core the principle of the impartial observer is hypothetical even though

[2]It uses, in contrast to Klibanoff, Marinacci and Mukerji (2005), a domain of preferences similar to that of Anscombe and Aumann (1963), which remains minimalistic by comparison.

the underlying scenarios might be factual (like that of the 'Afghan Goatherds'). Now, in the next part let us instead consider a concrete modification while still staying within the sphere of societal (or alternatively national) decision-making. Specifically, let us introduce farsightedness to the formation of global trade agreements as part of an extension of the toolset for the negotiations.

Since 1995, the World Trade Organization (WTO), as the direct successor to the General Agreement on Tariffs and Trade (GATT) of 1947, provides the rule set for the trade liberalization process of a significant number of countries - accounting for over 90 percent of world trade, GDP, and population as of 2007 (Source: WTO).[3] Its Article I acts as the foundation for all multilateral trade liberalization by formulating the so-called Most-Favoured-Nation (MFN) principle: Any concession granted to one member needs to be extended to all other members of the WTO. Contrary to this core MFN principle, Article XXIV Paragraph 5 explicitly allows countries to form so-called Preferential Trade Agreements (PTAs), i.e. Customs Unions (CUs) and Free Trade Agreements (FTAs), whereby a country does not need to extend the concessions granted within these arrangements to other countries. However, Subparagraphs (a) to (c) require these to be without any (negative) impact on other trade relations.

The overall (direction of the) influence of relaxing the WTO's core principle on the global trade liberalization process is a contested issue. Specifically, it is a controversial topic whether these PTAs ultimately function as 'building blocks' or 'stumbling blocks' on the path towards global free trade (Bhagwati (1993)). Generally, multilateral (non-discriminatory) negotiations, as the Doha Round, stand in stark contrast to bilateral (discriminatory) negotiations, currently leading to an ever-increasing number of PTAs - with forty percent of all countries/territories being a member of more than five PTAs and with about a quarter in more than ten (Source: WTO). All these trade negotiations usually entail potentially significant effects on the countries' economies, which makes for a complicated multi-party process accompanied by elaborate studies about feasibility and future developments. Yet, while a number of different papers try to tackle the aforementioned controversy, these papers typically consider a limited selection of trade agreements or otherwise assume a limited farsightedness of the negotiating countries.

Chapter 2 attempts to rectify this situation by introducing a farsighted perspective to the formation of global trade agreements while simultaneously expanding the set of trade agreements under consideration. As the underlying trade model, the framework utilizes a three-country two-good general equilibrium model of international trade, similar to that of Saggi and Yildiz (2010), which itself is a modification of the one in Bagwell and Staiger (1997). Each country then ranks different trade agreements based on the evaluation according to this trade model. The extensive choice set contains a baseline case, i.e. MFN, feasible PTAs, i.e. CUs and FTAs, as well as potential Multilateral Trade Agreements (MTAs), i.e. all those trade agreements that are consistent with the MFN principle. In order to take the farsightedness of the negotiating parties into consideration, the framework combines the ranking of these various trade agreements with the concept of 'consistent sets' as stable sets of trade agreements - a notion proposed by Chwe (1994). In a nutshell, a collection of trade agreements is considered stable as long as any deviation by any parties from an element in the stable set leads back to an element in the stable set without an

[3]All information obtained via `http://www.wto.org`.

improvement for the deviating party - in other words, the negotiating parties always consider all possible moves following their potential moves.

Ultimately, the framework produces a set of trade agreements which potentially emerge when the negotiating parties exhibit farsightedness, members and non-members of coalitions might freely interact with each other, and trade agreements constitute annullable contracts. A number of these capture important mechanisms present in the world and influence the composition of the stable set significantly when compared to other stability concepts. It is however an approach that is 'not so good at picking out, but ruling out with confidence' (Chwe (1994)). Consequently, when analyzing the effect of a varying relation of the countries' sizes in the form of their endowments, the unstable trade agreements provide effectively as much insight as the stable ones.

Unfortunately, the answer to the question whether PTAs are 'building blocks' or 'stumbling blocks' on the path towards global free trade is not as straightforward as one would like it to be. When comparing the two different regulatory scenarios (with and without PTAs), the difference ultimately depends on the size distribution of the countries. Whenever all countries are similar in size, Global Free Trade (GFT) emerges as the unique stable element under both the existing and the hypothetical institutional arrangement. Leaving the vicinity of symmetry produces mixed results. With two smaller countries PTAs reduce the area where the GFT regime is stable, while the opposite holds for two larger countries. However, once the world moves further away from symmetry, full trade liberalization is generally not attainable at all and PTAs function as a safeguard against the worst possible state from the perspective of overall world welfare, the non-cooperative MFN regime.

After exploring the introduction of a farsightedness perspective to the formation of global trade agreements, let us shift from the societal (or national) to the individual level. While the introduction of farsightedness ultimately expanded the set of information used by the economic agents, the last modification moves in the opposite direction. Specifically, let us introduce optimism to the funding of entrepreneurs in the form of rose-tinted glasses through which part of the entrepreneurs perceive their project.

Throughout contemporary history, a significant share of every generation wants to start new businesses, despite the knowledge about a poor chance of survival and return on investment - a phenomenon called 'a private equity premium puzzle' by Moskowitz and Vissing-Jorgensen (2002). In this context, the two central issues for the aspiring entrepreneurs are the source of funding and the level of commitment. Usually, the two funding choices under consideration are bank and venture capitalist, i.e. debt and equity, while the commitment, i.e. the non-monetary investment, possibly takes many different forms. Simultaneously, every potential financier needs to assess how realistic (or optimistic) the underlying business plan actually is. Naturally, the two choices of the entrepreneur also depend on an assessment of the project's productivity, which possibly deviates from that of the potential financier. In fact, various empirical studies, for example Cooper, Dunkelberg, and Woo (1988), and Camerer and Lovallo (1999), establish a connection between entrepreneurship and the bias of (unrealistic) optimism and overconfidence, thereby pointing towards a possible explanation for the aforementioned puzzle.

While the literature on the source of funding for entrepreneurs is extensive, 'A survey of venture capital research' (2011) by Da Rin, Hellmann, and Puri concludes

that despite the progress there are still open questions with respect to 'the choice between alternative sources of financing'. Specifically, to our knowledge the literature lacks a comprehensive model that features the investment of effort in a start-up under the assumption of the aforementioned perception bias while simultaneously answering the question about the source of financing.

Chapter 3 attempts to fill this perceived gap in the literature by introducing an optimistic perspective to the funding of entrepreneurs. In the framework, the entrepreneur faces a project of unknown quality that requires the investment of both funds and effort. In terms of the source of funding, the entrepreneur can either choose a venture capitalist, who provides both effort and funding at the cost of sharing the profit, or resort to a bank, which only provides funding at the cost of a payment of interest. With respect to the effort, the entrepreneur needs to choose the level of investment both in advance and during the actual project, while the provision of effort by the venture capitalist takes place during the actual project as well - in case the entrepreneur chose the venture capitalist. The introduction of the perception bias takes the form of different types of entrepreneurs, namely optimistic and realistic ones. A realistic entrepreneur receives an accurate signal about the quality of the project, while an optimistic entrepreneur receives a distorted one. However, both types interpret the signal as an undistorted one, which represents another perception bias (on a meta level). Irrespective of the type, the signal might further be affected by non-overconfidence noise, but both types account for this.

In terms of technical tools, the presented framework employs the concept of Perfect Bayesian Equilibrium with a focus on pure strategies. Further, with respect to the underlying signaling game between the entrepreneur and the venture capitalist, the analysis remains limited to separating equilibria in order to permit a scenario where realistic and optimistic entrepreneurs potentially choose different strategies.

Ultimately, nine potential scenarios emerge per each of the two quality signals, namely three scenarios with funding via the bank, three via the venture capitalist and three where the venture capitalist acts as a pseudo-bank - essentially providing the entrepreneur with the terms of funding of a bank. The three bank scenarios correspond to different levels of (re-)payment of the funds, i.e full, partial, and none. Meanwhile, the three venture capitalist scenarios coincide with different constraints, i.e. the participation of the venture capitalist, a technical zero bound, and none. An analysis of different comparative static effects of the model parameters shows that the bargaining process between the entrepreneur and the venture capitalist, when compared to a fixed contract with the bank, introduces strategic interactions that potentially lead to an underinvestment of the entrepreneurial effort, in part due to a hold-up problem.

A comparison with a social planner then provides an intuition for the potential beneficial impact of overconfidence on social welfare. In a nutshell, the distortion potentially alleviates other distortions (like the aforementioned bargaining process) present in the model, changes the probabilities of the two equilibrium scenarios, or alters them entirely. As a consequence, an increase in overconfidence might result in the second-best scenario, compared to the first-best scenario of a social planner. A numerical approach then characterizes the corresponding constellations of the model parameters through the analysis of pairwise comparative static effects with respect to the degree of overconfidence. A significant portion of these features a change in terms of the source of funding, specifically from the venture capitalist as a

pseudo-bank to the bank proper. There, a degree of overconfidence beyond a critical level forces the venture capitalist out of the competition with the bank, which implies better conditions for social welfare. While it is a local not a global statement, it does dispel the notion that an increase in overconfidence is necessarily detrimental for social welfare. In fact, a setting which facilitates a potential beneficial impact of overconfidence ideally features a relatively high entrepreneurial effort productivity with low effort productivity of the venture capitalist, a significant difference between the two project qualities, a relatively low or high fixed effort of the venture capitalist, a relatively high cost of investment, and a relatively low risk-free interest rate as well as interest rate premium. Furthermore, it points towards a low to medium pre-project entrepreneurial effort productivity and probability for a high quality, and a medium to high probability for a noisy signal.

Finally, the remainder of this thesis is organized as follows. First, Chapter 1 presents the Impartial Observer under Uncertainty. Second, Chapter 2 discusses the Farsighted Stability of Global Trade Policy Arrangements. Third, Chapter 3 involves the Funding of Overconfident Entrepreneurs. Each of those parts consists of a self-contained paper.

Contributions

The joint work with Lasha Chochua ultimately spawned two research projects, which found their way into this thesis in the form of Chapter 1 and 2. During these research projects, Lasha Chochua was a doctoral student and part of the BiGSEM at Bielefeld University. It is impossible to attribute any part of the joint work to any particular person. Each of us is equally responsible for the content.

Declaration

I hereby declare that I am aware of the current doctoral regulations of the faculty. The thesis is my and my co-author's original work. Any other contributions are marked as such. All other sources and utilized tools are listed. No third parties benefited financially (in a direct or indirect way) from work related to the content of this thesis. The thesis was never submitted as part of an examination before.

CHAPTER 1

The Impartial Observer under Uncertainty

1. Introduction

Individuals, as members of society, continually face choices among moral rules, institutional arrangements, government policies or patterns of wealth distribution; therefore, they are repeatedly involved in value judgements about which social alternative to choose. The history of economists (and philosophers) arguing that individuals should make such decisions under sympathetic interest in the welfare of each member of society without any bias towards particular participants dates back as far as Adam Smith's 'Theory of Moral Sentiments' (1759).

Harsanyi (1953, 1955, 1977) developed a rational theory of societal judgements. According to this theory, such choices should be made based on individual's 'social' or 'moral' preferences which are derived from the concept of an 'impartial observer'. As such, you imagine a situation where you do not know your actual place in society when comparing different social arrangements. Instead, you judge the desirability of the alternatives under the personal preferences of all members of society.[1] Thus, the original premise remains that this type of theory, unlike the theory of individual rational behavior or game theory, should be independent of selfish considerations.

The main result of this theory, now known as Harsanyi's Impartial Observer Theorem, combines Adam Smith's ideas of a sympathetic and impartial spectator with Kant's universality criterion and the utilitarian tradition of social utility maximization using von Neumann-Morgenstern expected utility theory. In the end, Harsanyi argued that an individual facing risky prospects over social outcomes and a hypothetical lottery over identities in society should rank these according to the weighted aggregate of individuals' expected utilities.

However, Harsanyi's Impartial Observer Theorem with its implicit utilitarianism only considers scenarios where each of the involved individuals faces objective risk. It is a theory analyzing societal judgements when objective probabilities over a set of social outcomes are known by each member of society. Our goal in this paper is to extend Harsanyi's Impartial Observer Theorem to include Knightian uncertainty in the model. By introducing individual belief systems about the likelihood of the social outcomes (which the impartial observer necessarily takes into consideration), our approach allows the application of the original framework to a new area of social value judgements. In particular, it allows the analysis of scenarios where individuals agree on the ranking but not on the likelihood of social outcomes. The impartial observer in our model does not aggregate these belief systems separately though.

[1]The imaginary construct of impartiality is similar to John Rawls' idea of a 'veil of ignorance' in 'A Theory of Justice' (1971), as these two metaphors are attempts at capturing the same stance.

When the impartial observer imagines herself being a particular individual, she adopts not only that individual's preferences but the belief system as well.[2]

The main result of our paper is a generalized utilitarian representation of the preferences of the impartial observer under uncertainty. It is a weighted sum of Second-Order Subjective Expected Utility (SOSEU) functions, each representing the preferences of an individual. Our framework is based on the generalized version of Harsanyi (1953) by Grant et al. (2010), which accommodates common criticism of Harsanyi's approach, specifically the issue of fairness and attitude towards mixing. The introduction of individual belief systems to our framework follows Seo (2009) and as a result the SOSEU functions supersede the Expected Utility (EU) functions of both Harsanyi's and Grant et al.'s approach.

In addition to the framework, our paper includes two illustrative examples. First, the moral dilemma of the 'Afghan Goatherds' (as presented in Sandel (2010)) showcases a scenario of agreement on the ranking but disagreement on the likelihood of social outcomes. Second, an example of a simple exchange economy with endowments and different alternatives of wealth (re-)distribution demonstrates the frameworks suitability for traditional economic problems - it serves as a proof of concept in that regard.

In order to motivate the introduction of uncertainty to moral value judgements, let us preview the story of the Afghan Goatherds. In 2005 a team of four soldiers, all U.S. Navy SEALs, set out to find a Taliban leader in the Afghan mountains near the Pakistan border. Just as the team set up their base overlooking the area to fulfill their reconnaissance mission, two Afghan goatherds stumbled upon them - a young boy with them. Due to the nature of their mission and other circumstances, the team considered killing or releasing the civilians the only two viable options. Eventually, the team cast a vote, where one soldier abstained, two voted two ways and the commander of the unit made the decisive call to release them. The civilians later informed the Taliban in a nearby village about the presence of the soldiers. In the subsequent ambush three of the four soldiers died, leaving the commander as the lone survivor.

In retrospect it is easy to make the correct call for this specific scenario. However, imagine you wanted to create a guideline for commanders about how to make such moral value judgements when in the field. In that case, you would naturally assume the role of an impartial observer and evaluate the situation based on individual preferences. In order to truly assume a person's view however, it is necessary to also take on that persons belief system - which is not possible (or included) in the traditional setting. Our approach therefore serves as an extension of the framework to include such cases where belief systems play an important role. This paper also contains a formal presentation of this specific moral dilemma and thereby connects the model with reality.

Section 2 discusses the related literature and focuses on the two relevant streams of literature, that is social choice theory and decision theory under uncertainty. Section 3 presents a minimal version of a framework based on Grant et al. (2010) and then introduces uncertainty following Seo (2009). Section 4 provides an analysis of the model, including a comparison with Grant et al. (2010). Section 5 contains

[2]In our opinion, this is a natural extension of Harsanyi's concept of impartiality. It differs therefore from the type of group decisions that are presented in Raiffa (1970), which features and discusses aggregation of belief systems.

the aforementioned illustrative examples as it revisits the Afghan Goatherds and presents the economic example. Section 6 summarizes and concludes our paper. The appendix contains additional details for the example of the Afghan Goatherds.

2. Related Literature

The important philosophical tradition of impartiality for moral value judgements about collective life has a long history. Vickrey (1945) and Harsanyi (1953) both independently introduced the idea to the economic literature and as Mongin (2001) formulates it: 'All in all, Harsanyi, if perhaps not Vickrey, should count as a major representative of the ethics of impartiality among 20th century writers.'

The (related) literature on Harsanyi's Impartial Observer Theorem is substantial. It is not our aim to review this vast literature as a whole, but instead to concentrate on the building blocks of our work and other closely related research.

As mentioned in the introduction, Grant et al. (2010) and Seo (2009) are the inspiration for the building blocks of our framework. In the first one, the authors revisit Harsanyi's Impartial Observer Theorem; they consider two major criticisms, concerning fairness and different risk attitudes, and derive a generalized version of the theorem that accommodates these criticisms. Furthermore, in the special case of an impartial observer that is indifferent between identity and outcome lotteries ('accidents of birth' and 'life chances') the generalized version of the theorem boils down to the standard Harsanyi doctrine. In consequence, the setting of the paper actually yields a new axiomatization of Harsanyi's utilitarianism. The resulting generalized utilitarianism serves as inspiration for the foundation of our approach. It gives us the possibility to extend the original framework of Harsanyi from risk to uncertainty while also accommodating common criticism of it.

In the second paper mentioned above, Seo formulates a model for decision making under uncertainty using second-order beliefs, i.e. beliefs over probability measures. Existing models in this stream of literature essentially differ in the choice of the domain of preference. Seo takes the domain of Anscombe and Aumann (1963) and a similar axiomatic foundation. Klibanoff, Marinacci and Mukerji (2005) by contrast require an additonal (sub-)domain with preferences. The domain selection of Seo allows us to introduce uncertainty to the (generalized) framework without any additional modifications. The Second-Order Subjective Expected Utility (SOSEU) representation of the preferences by Seo therefore translates to a corresponding version in the context of an impartial observer.

A variety of papers already deals with Harsanyi's Impartial Observer Theorem under uncertainty. Gajdos and Kandil (2008) provide an extension of the framework where the impartial observer considers sets of identity lotteries. In their model, unlike ours, uncertainty is introduced on a societal (not individual) level. The impartial observer's preference (under additional assumptions) is then characterized by a convex combination of Harsanyi's utilitarian and Rawls' egalitarian criteria.

A work closer to ours is the one by Nascimento (2012) which presents a model of aggregating preference orderings under subjective uncertainty. A fundamental difference is the setting of each paper. Namely, the one of Nascimento is that of a group of individuals that necessarily agrees on the ranking of certain risky objects. In contrast, our setting is one where a group of individuals does not agree on the ranking of any objects (risky or ambiguous). The assumption of Nascimento fits a group of experts or specialists in a field where there is a certain consent. However,

in our opinion this assumption is too restrictive for other cases, like the economic example in this paper (see Section 5 for a formal discussion). Nevertheless, the results are actually closely related, compare specifically Theorem 1 by Nascimento and Theorem 1 in this paper. In a sense, our work arrives at a similar representation but with a different axiomatic foundation and also with applications in mind that are explicitly excluded otherwise. Furthermore, as the analysis of the example of the Afghan Goatherds shows, our framework is actually also able to include scenarios with consent (see Section 5 and Appendix A).

It is worth mentioning (and repeating) that the impartial observer in our model always takes on the individual beliefs as part of the preferences thereby avoiding any aggregation of belief systems. In consequence, our model stays true to Harsanyi's thought experiment and also avoids the impossibility result of Mongin (1995).[3]

3. Model

Let (X, τ) be a topological space. Then, denote by $\mathcal{B}_X$ its Borel σ-algebra and by $\Delta(X)$ the set of all probability measures on $(X, \mathcal{B}_X)$. By x for $x \in X$ refer both to the actual element in X and to the induced one in $\Delta(X)$ - depending on context. Endow $\Delta(X)$ with the weak convergence topology. Also, endow any product of topological spaces with the product topology.

3.1. **General Setting.** Let $\mathcal{I} = \{i_1, \ldots, i_I\}$, $I \geq 2$, be a finite set of individuals facing a societal decision problem in the presence of individual uncertainty. Each social choice is modeled as a three-layered object (of different types of risk).[4] First, the (final) outcome space is given by $\mathcal{X}$ - a compact metrizable space with $|\mathcal{X}| \geq 2$. The outcome lotteries $p \in \Delta(\mathcal{X})$, also called one-stage lotteries, are the first layer (featuring objective risk). Further, let $\mathcal{S} = \{s_1, \ldots, s_S\}$, be the finite set of states of the world, which introduces uncertainty via individual beliefs about its probability distribution. The functions $h \colon \mathcal{S} \to \Delta(\mathcal{X})$, also called acts, are the second layer (featuring subjective risk or simply ambiguity). Denote by $\mathcal{H}$ the set of all acts. The act lotteries $P \in \Delta(\mathcal{H})$, also called two-stage lotteries, are the third layer (featuring objective risk again).

The individuals in this situation imagine themselves as an impartial observer, i.e. treating their (social) identity as an unknown component in the decision problem. As such, they face both identity lotteries $z \in \Delta(\mathcal{I})$ and act lotteries $P \in \Delta(\mathcal{H})$. Thus, the individual preferences $\succeq_i$, $i \in \mathcal{I}$, are each defined on $\Delta(\mathcal{H})$ while that of the impartial observer $\succeq$ is defined on $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$.[5] For all of these preferences, we assume a couple of 'standard' properties:

Assumption 1 (Individual)**.** *For each i in $\mathcal{I}$ the preference $\succeq_i$ on $\Delta(\mathcal{H})$ is complete, transitive and continuous. Its asymmetric part $\succ_i$ is non-empty.*

[3] A number of papers deals with decision making of societies using a mechanism of aggregating different individual beliefs, for example Cres, Gilboa and Vielle (2011), Alon and Gayer (2016), Danan, Gajdos, Hill and Talon (2016), and Qu (2017).

[4] The introduction of uncertainty via these three-layered objects follows Seo (2009) and by extension Anscombe and Aumann (1963).

[5] The impartiality that is presented here is based on the framework of Grant et al. (2010) which generalized the concept of Harsanyi (1953). In Grant et al. (2010) the impartial observer's preferences are defined on $\Delta(\mathcal{I}) \times \Delta(\mathcal{X})$, which naturally extends to $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ - incorporating the framework of Seo (2009). By contrast, the corresponding set in Harsanyi (1953) is $\Delta(\mathcal{I} \times \mathcal{X})$. See Grant et al. (2010) for a detailed discussion on this difference.

Assumption 2 (Impartial Observer)**.** *The preference $\succeq$ on $\Delta(\mathcal{I})\times\Delta(\mathcal{H})$ is complete, transitive and continuous. Its asymmetric part $\succ$ is non-empty*

Note, that by continuous we mean that the weak upper and lower contour sets are closed with respect to the corresponding topologies. In the case of the individual this means with respect to the weak convergence topology and in the case of the impartial observer the product topology of the weak convergence topologies.

Axiom 1 (Acceptance Principle)**.** *For all i in $\mathcal{I}$ and all P, Q in $\Delta(\mathcal{H})$:*

$$P \succeq_i Q \Leftrightarrow (i, P) \succeq (i, Q)$$

The acceptance principle establishes the intuitive link between the preferences of the individuals and that of the impartial observer. The intuition is that when the impartial observer imagines herself to be a particular individual that she in fact takes on the preferences of that individual (including the belief system).

Axiom 2 (Independence over Identity Lotteries)**.** *Suppose elements $(z, P), (z', Q)$ in $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ are such that $(z, P) \sim (z', Q)$. Then, for all $\tilde{z}, \tilde{z}'$ in $\Delta(\mathcal{I})$ and all α in $(0, 1]$:*

$$(\tilde{z}, P) \succeq (\tilde{z}', Q) \Leftrightarrow (\alpha\tilde{z} + (1-\alpha)z, P) \succeq (\alpha\tilde{z}' + (1-\alpha)z', Q)$$

The independence over identity lotteries as well as the acceptance principle are each concerned with the nature of the impartial observer's preferences with respect to identities. As our approach considers uncertainty on the level of outcomes and not identities, these two axioms naturally carry over from the traditional setting.

Assumption 3 (Absence of Unanimity)**.** *For all P, Q in $\Delta(\mathcal{H})$:*

$$\exists i \in \mathcal{I} : P \succ_i Q \Rightarrow \exists j \in \mathcal{I} : Q \succ_j P$$

The absence of unanimity can be interpreted as a required heterogeneity on the social alternatives and (preferences of) individuals. It is also not a new addition, but controversial enough to require an additional comment. First of all, normative decision-making is clearly trivial when all individuals agree on all rankings. Thus, it is possible to exclude this extreme case without losing any explanatory power. However, in our opinion it is too restrictive to completely leave out the opposite where everyone disagrees about everything - like it is done in Nascimento (2012) with the requirement of agreement on risky prospects. In general, our aim is to focus on scenarios that exhibit substantial heterogeneity in terms of (dis-)agreement.[6]

Next, let us state a lemma (which is going to be useful later on) about the representation of the preferences of the impartial observer and the individuals. Now, the structure of the results and the proof itself follow the ideas of Grant et al. (2010), in particular of their Lemma 8:

Lemma 1. *Suppose absence of unanimity applies. Then, the impartial observer satisfies the acceptance principle and independence over identity lotteries if and only if there exists a continuous function $V\colon \Delta(\mathcal{I}) \times \Delta(\mathcal{H}) \to \mathbb{R}$ that represents $\succeq$ and*

[6]It might seem that absence of unanimity is too restrictive as well (just in the other direction). Yet, adding a dummy individual that provides the (technically) required heterogeneity allows us to relax the restriction while still staying in our framework. See Section 5 and Appendix A for the formal presentation of the story of the Afghan Goatherds with a demonstration of a dummy.

for each individual i in $\mathcal{I}$ a function $V_i\colon \Delta(\mathcal{H}) \to \mathbb{R}$ that represents $\succeq_i$ such that for all (z, P) in $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$:

$$V(z, P) = \sum_{i \in \mathcal{I}} z_i V_i(P)$$

Moreover, the functions V and V_i, $i \in \mathcal{I}$, are unique up to common positive affine transformation.

Proof. In order to employ Lemma 8 of Grant et al. (2010) it is necessary to show that the set $\mathcal{H}$ is compact and metrizable.

For any compact topological space (X, τ) the set $\Delta(X)$ is always compact with the weak convergence topology. Therefore, as $\mathcal{X}$ is compact by the initial assumption, $\Delta(\mathcal{X})$ is compact as well. Any finite Cartesian product of compact spaces is compact with the product topology. Thus, with $\Delta(\mathcal{X})$ compact, $\mathcal{H} = \Delta(\mathcal{X})^S$ is compact too.

Using the fact that $\mathcal{X}$ is compact and metrizable (implying separable) it follows that $\Delta(\mathcal{X})$ is metrizable - for example using the Lévy-Prokhorov metric. Furthermore, by combining the metrices of the product for example with a p-norm, $1 \leq p$, $\mathcal{H} = \Delta(\mathcal{X})^S$ is metrizable as desired. □

Note that the proof actually works for a general compact and metrizable set $\mathcal{H}$. It requires no specific structure of $\mathcal{H}$. Thus, a modified Lemma 1 could potentially serve as a foundation for conceptually similar approaches to ours that only differ in terms of additional structure, specifically with respect to individual utilities.

Now, up to this point all assumptions and axioms follow Grant et al. (2010). Specifically, their axiom of independence over outcome lotteries (for individuals) is the only missing axiom. However, in the next part the axioms follow Seo (2009) instead and introduce uncertainty to the framework.

3.2. **Introducing Uncertainty.** In order to formulate the remaining axioms and introduce uncertainty to the model, it is necessary to define additional objects. Namely, let us define what we mean when talking about mixing two acts or lotteries (of acts). In the end, the two different kind of mixtures depend on the timing of the resolution of uncertainty (or of the mixing - depending on how you look at it).

First, consider the case where, when combining two (pure) acts, the uncertainty is resolved first and then the mixing takes place:

Definition 1. *For f, g in $\mathcal{H}$ and α in $[0, 1]$ and for $s \in S$, $B \in \mathcal{B}_{\mathcal{X}}$ we set*

$$(\alpha f \oplus (1 - \alpha)g)(s)(B) = \alpha f(s)(B) + (1 - \alpha)g(s)(B).$$

This operation is called a second-stage mixture.

Now, with this in mind, we introduce a 'standard' independence axiom with respect to second-stage mixtures:

Axiom 3 (Second-Stage Independence)**.** *For all i in $\mathcal{I}$, all α in $(0, 1]$ and lotteries p, q, r in $\Delta(\mathcal{X})$:*

$$\alpha p \oplus (1 - \alpha)r \succeq_i \alpha q \oplus (1 - \alpha)r \Leftrightarrow p \succeq_i q$$

Second, consider the case where, when combining two lotteries of acts, the mixing takes place first and then the uncertainty is resolved:

Definition 2. *For P, Q in $\Delta(\mathcal{H})$ and α in $[0,1]$ and for $B \in \mathcal{B}_{\mathcal{H}}$ we set*

$$(\alpha P + (1-\alpha)Q)(B) = \alpha P(B) + (1-\alpha)Q(B).$$

This operation is called a first-stage mixture.

Again, with this in mind, we introduce a 'standard' independence axiom with respect to first-stage mixtures:

Axiom 4 (First-Stage Independence)**.** *For all i in $\mathcal{I}$, all α in $(0,1]$ and lotteries P, Q, R in $\Delta(\mathcal{H})$:*

$$\alpha P + (1-\alpha)R \succeq_i \alpha Q + (1-\alpha)R \Leftrightarrow P \succeq_i Q$$

Finally, it is necessary to introduce an additional technical object that essentially serves as a tool for scenario analysis (for the individuals):

Definition 3. *Each $f \in \mathcal{H}$ and $\mu \in \Delta(\mathcal{S})$ induce a one-stage lottery*

$$\Psi(f, \mu) := \bigoplus_{s \in \mathcal{S}} \mu(s) f(s),$$

each $P \in \Delta(\mathcal{H})$ and $\mu \in \Delta(\mathcal{S})$ induce a two-stage lottery

$$\Psi(P, \mu)(B) := P(\{f \in \mathcal{H} \colon \Psi(f, \mu) \in B\})$$

for $B \in \mathcal{B}_{\mathcal{H}}$.

In other words, the element $\Psi(P, \mu)$ is the (induced) lottery that corresponds to the lottery P in the scenario where μ is the probability distribution over the states.

Axiom 5 (Dominance)**.** *For all i in $\mathcal{I}$, all P, Q in $\Delta(\mathcal{H})$:*

$$\Psi(P, \mu) \succeq_i \Psi(Q, \mu)\ \forall \mu \in \Delta(\mathcal{S}) \Rightarrow P \succeq_i Q$$

Imagine an individual only knows there exists a 'true' probability distribution but does not know which one it is. Then, the axiom of dominance captures the intuition that if the individual prefers one induced lottery over another one - substituting all possible probability distributions as the true one, then this individual should prefer that one lottery over the other (and vice-versa).

Finally, using the additional axioms with Lemma 1 it is possible to formulate the main result of our paper:

Theorem 1. *Suppose absence of unanimity applies. Then, the impartial observer satisfies the acceptance principle as well as independence over identity lotteries, and each individual satisfies first-stage and second-stage independence, and dominance if and only if the impartial observer's preference admits a representation $\langle \{U_i, \phi_i\}_{i \in \mathcal{I}} \rangle$ of the form*

$$V(z, P) = \sum_{i \in \mathcal{I}} z_i \phi_i(U_i(P))$$

where each

- *$\phi_i \colon \mathbb{R} \to \mathbb{R}$ is an increasing continuous function*
- *$U_i \colon \Delta(\mathcal{H}) \to \mathbb{R}$ is a SOSEU representation of $\succeq_i$*

i.e. the impartial observer is a generalized (weighted) utilitarian under uncertainty.

In addition, the U_i are unique up to uniqueness of the SOSEU representation. Further, the functions V and $\phi_i \circ U_i$ are unique up to a common positive affine transformation.

Proof. Let absence of unanimity apply.

Part 1 ('⇐'):

First, the representation of the impartial observer is affine in identity lotteries and therefore satisfies the acceptance principle and independence over identity lotteries. Note that alternatively this specific step also follows by application of Lemma 1. Second, the representation of each individual is of SOSEU form and thus satisfies its axioms, that is first-stage and second-stage independence, and dominance, using Theorem 4.2 of Seo (2009).

Part 2 ('⇒'):

In this part, the result of Lemma 1 is required. Namely, as the impartial observer satisfies the acceptance principle and independence over identity lotteries, it gives us a continuous function $V\colon \Delta(\mathcal{I}) \times \Delta(\mathcal{H}) \to \mathbb{R}$ representing $\succeq$ and for each $i \in \mathcal{I}$ functions $V_i\colon \Delta(\mathcal{H}) \to \mathbb{R}$ representing $\succeq_i$ such that for all $(z, P) \in \Delta(\mathcal{I}) \times \Delta(\mathcal{H})$:

$$V(z, P) = \sum_{i \in \mathcal{I}} z_i V_i(P)$$

The preferences of each individual satisfy first-stage and second-stage independence, and dominance, and so by Theorem 4.2 of Seo (2009) each V_i is a SOSEU function. However, this only holds up to transformation via an increasing continuous function. Thus, for each $i \in \mathcal{I}$ it follows that $V_i = \phi_i \circ U_i$ where U_i is a SOSEU function and ϕ_i is a transformation.

The uniqueness of the U_i follows directly from Lemma C.1 in Seo (2009), while the uniqueness of the V and $\phi_i \circ U_i$ follows from Lemma 1. □

In the theorem and its proof the actual SOSEU representation remained hidden. The following remark (re-)states the formal definition of a SOSEU representation and its uniqueness properties (see Seo (2009)). It will be helpful in the analysis and for the applications later on.

Remark 1. *A SOSEU representation is generally characterized by a triple (u, v, m) and of the form*

$$U(P) = \int_{\mathcal{H}} \int_{\Delta(\mathcal{S})} v\left(\int_{\mathcal{S}} u(f) d\mu\right) dm(\mu) dP(f)$$

for $P \in \Delta(\mathcal{H})$, where

- *u is a bounded continuous mixture linear function, $u\colon \Delta(\mathcal{X}) \to \mathbb{R}$,*
- *v is a bounded continuous strictly increasing function, $v\colon u(\Delta(\mathcal{X})) \to \mathbb{R}$,*
- *m is a probability measure, $m \in \Delta(\Delta(\mathcal{S}))$.*

The uniqueness of the SOSEU representation implies that for any another triple (u', v', m')

i) u and u' as well as $v \circ u$ and $v' \circ u'$ are each identical up to positive affine transformation,

ii) $\int_{\Delta(\mathcal{S})} \varphi dm = \int_{\Delta(\mathcal{S})} \varphi dm'$ for all continuous functions φ on $\Delta(\mathcal{S})$ for which there exists a Borel signed measure λ on $T := [u(\Delta(\mathcal{X}))]^S$ with bounded variation such that for all $\mu \in \Delta(\mathcal{S})$ it holds that $\varphi(\mu) = \int_T v(\mu \cdot t) d\lambda(t)$.

4. Analysis

Naturally, the starting point of our analysis is the connection between our result and that of Grant et al. (2010). As expected, completely eliminating uncertainty by reducing it to risk produces their result as a special case of ours (see 4.1).

Furthermore, one (significant) indeterminacy in Theorem 1 is the specific form of the ϕ_i, $i \in \mathcal{I}$. In the current setting, the only restriction is that each of them is an increasing continuous function. Let us therefore analyze the specific form of these functions in relation with different (additional) properties of the preferences. In particular, let us compare this to the findings presented in Grant et al. (2010). Fortunately, their results and proofs actually do not rely on the underlying structure of the outcome space and therefore all of their findings translate into our setting without any modifications.

4.1. **The Special Case of Risk.** In the special case of (only) risk, i.e. $\mathcal{S} = \{s\}$, all belief systems are trivial. In consequence, the three-layer objects containing uncertainty in the middle reduce to two-layer objects with only risk present. Further, in order to identify this with the setting of Grant et al. (2010) assume that each v_i, $i \in \mathcal{I}$, is actually a linear function (corresponding to indifference to uncertainty) or equivalently assume reversal of order or reduction of compound lotteries.[7] Thus, the two-layer objects with risk collapse to single-layer objects with risk.

In this (special case of a) setting, first/second-stage independence simply reduces to independence over outcome lotteries, dominance is now an empty statement, and the representation of the impartial observer takes on the form of the generalized (weighted) utilitarian of Grant et al. (2010).

4.2. **Fairness.** One common criticism of Harsanyi's utilitarianism is concerned with fairness (see for example Diamond (1967)). It is one of the two issues that Grant et al. (2010) solved by generalizing the theory. The notion of fairness in this context refers to a preference of the impartial observer for mixing act lotteries over mixing identity lotteries.[8]

Consider from now on those tuples of identity and act lotteries between which the impartial observer is indifferent, that is (z, P') and (z', P) with $(z, P') \sim (z', P)$. Following the arguments in favor of fairness, the impartial observer always prefers mixing these pairs on the level of acts over mixing on the level of identities, i.e. $(z, \alpha P + (1-\alpha)P') \succeq (\alpha z + (1-\alpha)z', P)$ for all $\alpha \in (0,1)$, which is also referred to as a preference for life chances (compared to accidents of birth). In the case of risk, Grant et al. (2010) show that this holds if and only if each ϕ_i, $i \in \mathcal{I}$, is concave, which actually translates one-to-one into our setting.

Conversely, consider a scenario where the impartial observer is indifferent between life chances and accidents of birth, that is $(z, \alpha P + (1-\alpha)P') \sim (\alpha z + (1-\alpha)z', P)$

[7]Take $i \in \mathcal{I}$. Then, reversal of order and reduction of compound lotteries each describe a property of the preferences of individual i with respect to first-stage and second-stage mixtures.

Namely, reversal of order is satisfied if for all $f, g \in \mathcal{H}$ and $\alpha \in [0,1]$

$$\alpha f \oplus (1-\alpha)g \sim_i \alpha f + (1-\alpha)g.$$

Similarly, reduction of compound lotteries is satisfied if for all $p, q \in \Delta(\mathcal{X})$ and $\alpha \in [0,1]$

$$\alpha p \oplus (1-\alpha)q \sim_i \alpha p + (1-\alpha)q.$$

The axiom of dominance implies the equivalence of these two properties (see Seo (2009)).

[8]The paper of Grant et al. (2010) also provides an example justifying this notion of fairness.

for all $\alpha \in (0,1)$. In the case of risk, Grant et al. (2010) show that this holds if and only if each ϕ_i, $i \in \mathcal{I}$, is affine, which again translates one-to-one into our setting.[9]

4.3. **Mixtures.** In the previous part, the focus was on the relation between mixing either identity or act lotteries. Another criticism of Harsanyi's utilitarianism is concerned with different attitudes among the individuals towards mixing.[10]

Fix two individuals, say i and j, and consider from now on those act lotteries which the impartial observer ranks equally from each perspective, that is $P, \tilde{P}, Q, \tilde{Q}$ with $(i,P) \sim (j,Q)$ and $(i,\tilde{P}) \sim (j,\tilde{Q})$. Imagine that the impartial observer prefers facing the (first-stage) mixtures of each of those two pairings of act lotteries as i rather than as j, i.e. $(i, \alpha P + (1-\alpha)\tilde{P}) \succeq (j, \alpha Q + (1-\alpha)\tilde{Q})$ for all $\alpha \in (0,1)$. Now, Grant et al. (2010) show that this holds if and only if the composite function $\phi_i^{-1} \circ \phi_j$ is convex on the domain $\mathcal{U}_{ji} := \{u \in \mathbb{R} \,|\, \exists P, Q \in \Delta(\mathcal{H})\colon (i,P) \sim (j,Q) \wedge U_j(Q) = u\}$ for the case of risk but it actually also applies to our setting under uncertainty.

Alternatively, imagine that the impartial observer is indifferent when comparing to face these mixtures as different individuals: $(i, \alpha\tilde{P} + (1-\alpha)P) \sim (j, \alpha\tilde{Q} + (1-\alpha)Q)$ for all $\alpha \in (0,1)$ and all $i, j \in \mathcal{I}$. Following Grant et al. (2010) this holds if and only if $\phi_i = \phi$, $i \in \mathcal{I}$, both for risk and under uncertainty. Further, let i_1 and i_2 be a pair of individuals such that there exists a sequence of individuals $j_1, \ldots, j_N$ with $j_1 = i_1$ and $j_N = i_2$ where each $\mathcal{U}_{j_n, j_{n-1}}$ has non-empty interior. Then, the functions U_i, $i \in \mathcal{I}$, are unique up to a common positive affine transformation.

5. Applications

In the following, two applications (or examples) for our approach are presented. The first example picks up the story of the Afghan Goatherds from the introduction; the second one is a simple economic example.

As mentioned in the introduction, the moral dilemma of the Afghan Goatherds features a scenario where individuals agree on the ranking of social outcomes, but disagree on the likelihood of these. It showcases the effect of the introduction of uncertainty, as its results are purely driven by the nature of different belief systems.

The economic example on the other hand essentially serves as a proof of concept. It illustrates that our framework is able to accommodate not only pure philosophical but also economic problems. As a bonus, the example allows us to demonstrate the effect of the degree of fairness on the level of the impartial observer.

5.1. **Afghan Goatherds.** First, recall the moral dilemma of the Afghan Goatherds described in the introduction. As mentioned there, our claim is actually not that the commander of the unit necessarily made the decision using anything related to our approach (even though it could very well be the case). However, our approach allows a normative analysis of this (and similar) situations. You could for example be developing a moral guideline for the military in the vein of the United States Army Field Manuals. Naturally, you would then be in the position of a neutral or impartial observer and evaluate the situation using the point of view of each involved individual including their perception of the situation's uncertainty.

[9] Note that in this case the resulting representation is actually equivalent to a framework with a single probability distribution over states for every individual and simultaneously a modified probability distribution over individuals that includes belief systems.

[10] In Grant et al. (2010) this issue is actually referred to as 'different attitude towards risk'.

Let us define the formal structure of the decision problem now. It is deliberately kept simplistic compared to the actual events. Hence, it might not be a perfect fit for every aspect of the original story, but should still serve as a proper demonstration of an application of our theory. First, let $\mathcal{X} = \{0,1\}^2 \backslash \{(0,0)\}$ be the set of outcomes. Each element $x = (x_1, x_2)$ corresponds to the survival of the soldiers, x_1, and that of the afghan goatherds, x_2. Each entry then indicates either 'alive', 1, or 'dead', 0. Furthermore, let $\mathcal{S} = \{t, u\}$ be the states of the world, where t and u correspond to talking to the Taliban about the soldiers and keeping quiet about them, respectively. Finally, the two available moral choices are killing or sparing the civilians, denoted by K and L, respectively. Thus, using the previous notation, they are given by:

$$K = \begin{cases} (1,0) & s = t, u \end{cases}$$

$$L = \begin{cases} (0,1) & s = t \\ (1,1) & s = u \end{cases}$$

Note that these elements only contain subjective risk (with respect to the states). Together with the remaining specifications, this is actually going to guarantee that the example is purely driven by individual belief systems.

In addition, let $I = \{1, \dots, 3\}$ be the set of individuals - corresponding to the team of soldiers.[11] Following the traditional setting, the identity lottery that is part of the choice problem is going to be fixed to $(1/3, 1/3, 1/3)$. Note that in this setting, i.e. with this set of individuals and this (fixed) identity lottery, the specifications of the next part actually conflict with our initial assumption of absence of unanimity. Appendix A analyzes and corrects this problem by introduction of a dummy variable without any changes to the preferences. In the rest of this analysis therefore consider the assumption of absence of unanimity to be fulfilled.

5.1.1. *Specifying the Moral Dilemma.* In order for us to conduct a detailed analysis, we need to specify more about the preferences of the individuals and in particular about the the nature of their belief systems. First of all, to ensure that our results are purely driven by the beliefs of the soldiers, we assume that their preferences are otherwise completely identical, that is $u_i = u$ and $v_i = v$ for each $i \in \mathcal{I}$. Similarly, assume that the impartial observer treats the soldiers identical with respect to facing similar mixtures. Consequently, $\phi_i = \phi$, $i \in \mathcal{I}$, by Section 4.3.

Furthermore, let us specify the ranking of all of the possible survival outcomes. Naturally, you would assume that the soldiers rank the survival of all the highest. Additionally, our assumption is going to be that the soldiers, when confronted with the exclusive survival outcomes, prefer their own survival over that of the civilians. This essentially captures the idea of universal self-preservation instincts. Therefore, after using a positive affine transformation (to specify the lower and upper bound):

$$0 = u(0,1) < u(1,0) < u(1,1) = 1$$

Moreover, the individuals are assumed to be uncertainty-averse, which then implies a concave function v. In particular, it is going to be of the form z^q for $q \in (0,1)$.

[11] In the original story, the team consists of four soldiers including the commander of the unit. Each of the three regular soldiers exhibited different individual preferences, i.e. $K \succ_1 L$, $L \succ_2 K$ and $K \sim_3 L$, which is already enough to construct a simple (yet interesting) example. Therefore, excluding the commander as an individual is of no (significant) consequence to our analysis.

Assume further a preference for life chances (of the impartial observer). Thus, Section 4.2 yields that ϕ is also a concave function. As before, take z^r for $r \in (0,1)$.

In general, the belief systems m_i are probability distributions over $\Delta(\mathcal{S})$, which in this specific scenario is equivalent to probability distributions over $[0,1]$ by simply identifying $\mu \in \Delta(\mathcal{S})$ with $p \in [0,1]$ via $p = \mu(u)$ (alternatively $p = \mu(t)$). Now, the actual belief systems are going to be truncated normal distributions on $[0,1]$.[12] Let (μ_i, σ_i) denote a pair of parameters of the initial normal distributions. Then, assume that $\mu_i = \mu = 0.5$ for $i \in \mathcal{I}$ and furthermore $0 < \sigma_1 < \sigma_2 < \sigma_3 < +\infty$. Hence, the individual belief systems are mean-preserving spreads of each other, which allows us to showcase the effect of introducing uncertainty to the framework. Additionally, the centered mean captures the idea of unbiased individuals. Finally, combining everything together yields the following utility of the impartial observer for the two moral choices:

$$\begin{aligned} V(z,K) &= \sum_{i=1}^{3} \frac{1}{3} \phi(v(u(1,0))) \\ &= (u(1,0)^q)^r \\ V(z,L) &= \sum_{i=1}^{3} \frac{1}{3} \phi\left(\int_0^1 v(pu(1,1) + (1-p)u(0,1)) dm_i(p) \right) \\ &= \sum_{i=1}^{3} \frac{1}{3} \left(\int_0^1 p^q dm(\mu, \sigma_i)(p) \right)^r \end{aligned}$$

5.1.2. *Numerical Analysis.* In addition to calculating the results for specific values, the aim is to demonstrate the effect of uncertainty in our framework. Therefore, consider a modified version of the framework where each individual only uses a single (subjective) probability distribution over states but on a societal level there is an additional probability distribution over these subjective probability distributions.[13] In other words, instead of uncertainty, the model features subjective risk. Denote by $\tilde{V}(z,P)$ the evaluation of (z,P) corresponding to the modified model. Ideally, the comparison between our model and this modified one produces a difference and thus justifies to a certain degree the use of the concept of uncertainty in our model.

Finally, fix $q = 0.75$, $r = 0.25$, $\sigma_1 = 0.01$, $\sigma_2 = 0.1$, $\sigma_3 = 1$, and $u(1,0) = 0.48$. Now, the parameter choices here (and also the previous functional choices) should be understood as part of an example. Our (numerical) analysis uses reasonable but still debatable choices to enable us to showcase interesting phenomena for one of potentially many formal interpretations of the story within our framework. Anyhow, using these parameters produces the following values when rounded to the fourth decimal:

	V_1	V_2	V_3	V	$\tilde{V}$
K	0.5767	0.5767	0.5767	0.8714	0.8714
L	0.5946	0.5923	0.5723	0.8751	0.8657

[12] Alternatively, take beta distributions $Beta(\alpha, \beta)$ with varying $\alpha = \beta$.

[13] Formally, this corresponds to the use of indicator functions as belief systems, i.e. $m_i = \mathbb{1}_{\mu_i}$ for $\mu_i \in \Delta(\mathcal{S})$ and all $i \in \mathcal{I}$, and an extended set of individuals that includes beliefs, i.e. $\mathcal{I} \times \Delta(\mathcal{S})$ where the density function is given by $f_z(i,\mu) = z_i \cdot m_i(\mu)$ for $(i,\mu) \in \mathcal{I} \times \Delta(\mathcal{S})$, instead of $\mathcal{I}$ and the corresponding $z \in \Delta(\mathcal{I})$.

As a consequence, the values produce the following rankings:

$$V_3(L) < V_{1/2/3}(K) < V_2(L) < V_1(L)$$
$$V(z,K) < V(z,L)$$
$$\tilde{V}(z,L) < \tilde{V}(z,K)$$

Therefore, the higher the (perceived) uncertainty on the level of the individuals the lower the evaluation of L compared to that of K, which stays constant. In fact, the dynamic actually produces different individual rankings of the two alternatives. Additionally, the comparison of our model and the modified one actually produces a different ranking on the level of the impartial observer. Thus, the introduction of uncertainty influences (and to a certain extent drives) the final ranking. Further, the impartial observer actually agrees with the result of a simple majority vote in this specific case (which in general is not guaranteed).

5.2. **Exchange Economy.** Consider a simple exchange economy with two goods and two individuals, with each of them receiving endowments. In this setting, compare two alternative re-distributions rules, namely the Walrasian auctioneer and the Egalitarian rule.[14] Uncertainty enters the model via a possible bias in the distribution of the endowments. This specific example serves as a proof-of-concept in the sense that it shows an interpretation of a traditional economic problem within our framework. It is based on an example by Eichberger and Pethig (1994).

Let $\mathcal{I} = \{1, 2\}$ be the set of individuals and let x and y denote the two goods. In order to keep it simple, let us assume that the total endowment for each good is set to 3 and restricted to positive integers. Thus, the possible initial endowments are given by the two by two matrix

$$\begin{pmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{pmatrix} = \begin{pmatrix} ((1,1),(2,2)) & ((1,2),(2,1)) \\ ((2,1),(1,2)) & ((2,2),(1,1)) \end{pmatrix}$$

where $e_{i,j}$ in row i and column j corresponds to individual 1 receiving (i, j) and individual 2 receiving $(3 - i, 3 - j)$ of the pair of goods $(e_{x,k}, e_{y,k})$, $k = 1, 2$.

In the following, our analysis focuses on two possible re-distributions of these initial endowments, namely the Walrasian auctioneer and the Egalitarian rule. Let the utility function of an individual i for the (re-distributed) goods x_i and y_i be given by the Cobb-Douglas form $(x_i y_i)^{\alpha_i}$, where $\alpha_i \in \mathbb{R}_{>0}$. Then, the individuals evaluate the results of the re-distributions as follows (depending on endowments):

In case of the Egalitarian rule, any endowment vector $((e_{x,1}, e_{y,1}), (e_{x,2}, e_{y,2}))$ yields the consumption bundles $((3/2, 3/2), (3/2, 3/2))$. In consequence, the utility for an individual i is always given by $(9/4)^{\alpha_i}$, independent of initial endowments. Now, in case of the Walrasian auctioneer rule, any fixed endowment vector induces a corresponding unique Walrasian equilibrium. The resulting utility of individual i is then given by $((e_{x,i} + e_{y,i})^2/4)^{\alpha_i}$, where $e_{x,i}$ and $e_{y,i}$ are the initial endowments. Using our matrix notation, the following then characterizes the re-distribution rules

[14] In general there is also a combination of the two, namely the Walras rule from Equal Division. However, in this scenario, it coincides with the Egalitarian rule. See Nagahisa and Suh (1995) for a characterization of the Walras rules.

with respect to the individual utilities for all possible initial endowments:

$$\tilde{E}\colon \begin{pmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{pmatrix} \mapsto \begin{pmatrix} ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) & ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) \\ ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) & ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) \end{pmatrix}$$

$$\tilde{W}\colon \begin{pmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{pmatrix} \mapsto \begin{pmatrix} (1, 4^{\alpha_2}) & ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) \\ ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) & (4^{\alpha_1}, 1) \end{pmatrix}$$

The notation with the tilde is deliberate in order to distinguish these descriptions from their counterparts that take the uncertainty into consideration.

As mentioned earlier, the uncertainty is about the probability distribution over the initial endowments. Assume that there are two states, i.e. $\mathcal{S} = \{s_1, s_2\}$ where s_1 corresponds to a bias towards individual 1 and analogously s_2 towards 2. Thus, these are described via the following probability distributions (π_1 for s_1, π_2 for s_2):

$$\pi_1(e_{i,j}) = \begin{cases} \frac{1}{2} & i = j = 2 \\ \frac{1}{4} & i = 1, j = 2 \text{ or } i = 2, j = 1 \\ 0 & i = j = 1 \end{cases}$$

$$\pi_2(e_{i,j}) = \begin{cases} \frac{1}{2} & i = j = 1 \\ \frac{1}{4} & i = 1, j = 2 \text{ or } i = 2, j = 1 \\ 0 & i = j = 2 \end{cases}$$

In a sense, our example exhibits uncertainty about (the state of) the economy instead of a fixed (state of the) economy with inherent uncertainty. Consequently, society chooses the re-distribution rule before knowing the initial distributions.

Finally, to formally state the two rules taking uncertainty into consideration, i.e. to formulate the act lotteries corresponding to them, combine the aforementioned functions and probability distributions as follows:

$$E\colon s_k \mapsto \left(\tilde{E}(e_{i,j}) \text{ with probability } \pi_k(e_{i,j})\right)$$

$$W\colon s_k \mapsto \left(\tilde{W}(e_{i,j}) \text{ with probability } \pi_k(e_{i,j})\right)$$

Furthermore, using the inherent symmetries and other similarities simplifies it to:

$$E\colon s \mapsto ((9/4)^{\alpha_1}, (9/4)^{\alpha_2})$$

$$W\colon s \mapsto \begin{cases} \left(4^{\alpha_1 \mathbb{1}_{\{s_1\}}(s)}, 4^{\alpha_2 \mathbb{1}_{\{s_2\}}(s)}\right) & \text{with probability } 1/2 \\ ((9/4)^{\alpha_1}, (9/4)^{\alpha_2})) & \text{with probability } 1/2 \end{cases}$$

Following the traditional setting, take again the 'fair' uniform identity lottery, that is $z = (1/2, 1/2)$. As in the previous example assume moreover that the impartial observer treats individuals identical with respect to similar mixtures, thus $\phi_i = \phi$, $i \in \mathcal{I}$. In addition, set $\alpha_i = 1/2$ for both i (assuming similar risk-aversion) and fix v_i to be the identity function for both i (assuming uncertainty-neutrality).[15] Also, take $\phi = z^r$ again but with $r \in (0, +\infty)$ this time.

As individual belief systems m_i take truncated normal distributions on $[0, 1]$ again, where $p \in [0, 1]$ corresponds to the probability of state s_i realizing. Further, let (μ_i, σ_i) denote a pair of parameters of the initial normal distributions. Then,

[15]It seems completely counter-intuitive to assume uncertainty-neutrality in our framework as the introduction of uncertainty is our main contribution. However, in this example and specifically this (numerical) analysis our focus is on the effect of different transformations ϕ.

assume that $\mu_1 = 0.75$, $\mu_2 = 0.25$, and $\sigma_i = \sigma = 0.25$, i.e. each of the individuals suspects a bias towards individual 1 and the same level of volatility.

Note that due to the exclusive nature of the game, essentially a zero-sum game, and the additional assumptions, absence of unanimity requires no modifications. Finally, combine everything together for the following evaluations:

$$\begin{aligned} V(z,E) &= \sum_{i=1}^{2} \frac{1}{2}\phi_i \left(v_i \left(\left(\frac{9}{4} \right)^{\alpha_i} \right) \right) \\ &= \left(\frac{3}{2} \right)^{r} \\ V(z,W) &= \sum_{i=1}^{2} \frac{1}{2}\phi_i \left(\int_0^1 v_i \left(\frac{1}{2} \left(\frac{9}{4} \right)^{\alpha_i} + \frac{1}{2} \left(4^{\alpha_i} p + 1\,(1-p) \right) \right) dm_i(p) \right) \\ &= \sum_{i=1}^{2} \frac{1}{2} \left(\frac{5}{4} + \frac{1}{2} \int_0^1 p dm(\mu_i, \sigma)(p) \right)^{r} \end{aligned}$$

Consider different values for $r \in (0, +\infty)$ now, which captures different degrees of fairness on the level of the impartial observer. It results in the following values when rounded to the fourth decimal:

	V_1	V_2	$V_{\|r=1.5}$	$V_{\|r=1}$	$V_{\|r=0.5}$
W	1.5897	1.4103	1.8396	1.5000	1.2242
E	1.5000	1.5000	1.8371	1.5000	1.2247

Evidently, the ranking of the impartial observer depends on the exact value of r and the rankings of the individuals are diametrically opposed:

$$\begin{aligned} V_2(W) < V_2(E) = V_1(E) < V_1(W) \\ V(z,E)_{|r=1.5} < V(z,W)_{|r=1.5} \\ V(z,E)_{|r=1} = V(z,W)_{|r=1} \\ V(z,E)_{|r=0.5} > V(z,W)_{|r=0.5} \end{aligned}$$

In other words, a preference for life chances of the impartial observer actually leads to a preference of the Egalitarian rule over that of the Walrasian auctioneer in a setting with a clear bias towards one specific individual. It essentially provides another example for the discussion on the issue of fairness.

6. Conclusion

The focus of this paper is the normative decision-making of an individual when considering all other individuals in society and under the influence of uncertainty. Based on the works of Grant et al. (2010) and Seo (2009) we provide an axiomatic foundation for an extension of Harsanyi's Impartial Observer Theorem that includes Knightian uncertainty while also accommodating certain common criticism of the traditional result. The main result shows that the impartial observer's preferences admit a representation in form of a weighted average of the individual (transformed) second-order subjective expected utilities. This representation allows for a tractable analysis of the normative choice problems under consideration. Furthermore, the framework re-establishes links between additional properties of the preferences of the impartial observer on one hand (e.g. the issues of fairness and mixtures) and the specific form of the individual transformations on the other.

The main appeal of our model is the extension to normative decision-making in situations where a group of individuals faces subjective instead of objective risk. The story of the Afghan Goatherds is an example for such moral value judgements. It is also purely driven by individual belief systems and thus provides justification for the introduction of uncertainty to the framework. Moreover, the example shows that our model in fact extends beyond the limitations of absence of unanimity via the use of a dummy individual. Finally, the economic example is an application of our theory to a scenario that demonstrates the effect of a preference for life chances - compared to accidents of birth - of the impartial observer. It also serves as a proof-of-concept for the application of our theory to economic problems in general.

Appendix A. Dummy for the Afghan Goatherds

As mentioned in the main part of the paper, in order to actually satisfy the assumption of absence of unanimity, without distorting any preferences, the introduction of a dummy individual d is necessary. The following explains this necessity:

Assuming sufficiently heterogeneous beliefs, it is certainly possible to imagine a ranking of the form $K \succ_1 L$, $L \succ_2 K$ and $K \sim_3 L$ - essentially mirroring reality. However, the assumption of absence of unanimity also applies to all degenerate outcome lotteries, like the one always yielding the outcome $(1, 1)$ irrespective of the state of the world and also the one always yielding $(0, 1)$. Certainly, it is counter-intuitive to assume that in reality one of the soldiers would prefer the second over the first one, i.e. prefering being dead over being alive with everything else fixed. Now, the dummy individual takes care of this problem by prefering $(0, 1)$ over $(1, 1)$ and therefore maintaining absence of unanimity. At the same time, the probability of imagining yourself as the dummy individual is set to zero (for both options) to prevent any distortion on the level of preferences of the impartial observer.

A dummy individual seems artifical, especially one that is necessary because of an assumption that is imposed by us on the model. Yet, it is not an actual restriction or invalidates the assumption. It is merely a technical solution to a technical problem. The two presented (degenerate) lotteries that conflict with absence of unanimity otherwise are (or were) not part of the set of feasible options in reality anyway. A dummy individual in this example is necessary due to the homogeneity of the individuals (and their preferences), which is the result of a simplistic structure. Thus, a dummy individual allows us to apply our theory to examples where the size of the choice set collides with absence of unanimity otherwise. Essentially, this weakens absence of unanimty while still remaining in the framework of our theory.

Formally: Consider $\mathcal{I}' = \mathcal{I} \cup \{d\}$ with the (fixed) identity lottery $(1/3, 1/3, 1/3, 0)$ and set $u_d(x_1, x_2) = u(x_1, x_2)$ and $v_d(y) = 1 - v(y)$ with $m_d = m_1$ (or $m_d = m_{2/3}$). It results in the following (ultimately irrelevant) utilities for the two moral choices: $V_d(K) = 0.4233$ and $V_d(L) = 0.4054$; Consequently: $V_d(L) < V_d(K)$.

CHAPTER 2

The Farsighted Stability of Global Trade Policy Arrangements

1. Introduction

Following the General Agreement on Tariffs and Trade (GATT) of 1947, an increasing number of signatory countries liberalized their trade policies primarily via two channels: bilateral and multilateral negotiations. To the present day, there have been eight rounds of multilateral trade negotiations with the current ninth one, the Doha Round, still ongoing. At the same time, parallel to the arrangements observed on the multilateral level, the world has seen an ever-increasing number of Preferential Trade Agreements (PTAs) mainly in the wake of bilateral negotiations. Currently, about fourty percent of all countries/territories are a member of more than five PTAs while about a quarter participates in more than ten.[1]

The World Trade Organization (WTO), successor of the GATT in 1995, provides the rule set for the trade liberalization process of a significant number of countries.[2] Its Article I acts as the foundation for any multilateral trade liberalization by formulating the so-called Most-Favoured-Nation (MFN) principle: Any concession granted to one member needs to be extended to all other members of the WTO.[3] In this paper, trade policy arrangements that are consistent with the MFN principle are referred to as Multilateral Trade Agreements (MTAs).[4] Contrary to the core MFN principle, Article XXIV Paragraph 5 explicitly allows countries to form PTAs, specifically Customs Unions (CUs) and Free Trade Agreements (FTAs), that do not need to extend the concessions granted within the arrangement to other countries.[5] However, Article XXIV Paragraph 5 Subparagraph (a), (b), and (c) each require that these are without (negative) influence on other trade relations.

The (direction of the) influence of Article XXIV Paragraph 5 on the development of trade policy arrangements is a controversial topic and the focus of many papers.[6] Likewise, the primary purpose of this paper is the analysis of the stability of different trade policy arrangements in two scenarios, that is with PTAs (current WTO rules) and without PTAs (modified WTO rules). In particular, it is our intent to examine

[1] Source: `http://www.wto.org`

[2] All members of the WTO account for 96.4 percent of world trade, 96.7 percent of world GDP, and 90.1 percent of world population as of 2007 (Source: `http://www.wto.org`).

[3] Article I states that 'any [...] favour [...] granted by any contracting party to any product originating in or destined for any other country shall be accorded immediately and unconditionally to the like product originating in or destined for [...] all other contracting parties' (GATT, 1947).

[4] Furthermore, we interchangeably use the terms trade policy arrangements, trade agreements, trade constellations, and trade relations.

[5] Article XXIV Paragraph 5 states that '[...] this agreement shall not prevent [...] the formation of a customs union or of a free-trade area [...]' (GATT, 1947).

[6] The next part of this paper contains further information on the related literature.

whether PTAs act as 'building blocks' or 'stumbling blocks' on the path towards global free trade (Bhagwati (1993)).

The existing literature usually considers a limited selection of trade agreements or assumes limited farsightedness of the negotiating countries. It certainly allows for a cleaner description of the model and interpretation of its results, but ultimately raises the question about whether or not these restrictions significantly influence the analysis and to what degree these frameworks capture reality. In our opinion, certain empirical observations favor an extensive choice set and full farsightedness. During the past rounds of multilateral trade negotiations, many countries were simultaneously involved in other trade liberalization processes.[7] Moreover, such trade negotiations are usually complicated processes with significant effect on the countries' economies and accompanied by elaborate studies about feasibility and future developments.[8] Taking these assessments into account, the contribution of our paper is an answer to the question concerning the influence on the analysis.

First of all, our paper considers an extensive set of trade agreements, containing PTAs, i.e. CUs and FTAs, as well as MTAs. Next, endogenizing the formation of trade agreements, each country ranks them based on a three-country two-good general equilibrium model of international trade.[9] The stability of all trade agreements is then examined using these rankings together with the concept of 'consistent sets' as stable sets - a notion proposed by Chwe (1994). As a result, our paper expands the set of trade agreements under consideration and also extends the farsightedness of the negotiating parties in comparison to the literature. In fact, to the best of our knowledge, no other paper considers a choice set as extensive as ours.

In the end, our analysis shows that the effect of PTAs on trade liberalization depends on the size distribution of the countries. As long as the countries are close to symmetric, Global Free Trade (GFT) emerges as the unique stable outcome under both the existing and the hypothetical institutional arrangement. However, when two countries are considerably smaller, a modified WTO without PTAs would facilitate the formation of GFT. By contrast, if two countries are relatively larger, this modified WTO would actually obstruct the development towards GFT. Once the world is further away from symmetry, full trade liberalization is not attainable at all and abolishing the exception for PTAs might result in the worst possible state from the perspective of overall world welfare, the non-cooperative MFN regime.

The findings of our paper notably deviate from those of the existing literature. Compared to the paper of Saggi, Woodland and Yildiz (2013), the composition of the stable set of trade policy arrangements differs on a substantial part of the parameter space under consideration (while coinciding on the remainder). Beyond that, the comparison with the work of Lake (2017) yields not only a difference in terms of stability but also with respect to the driving force(s).

---

[7] Maggi (2014) showcases the importance of an extensive set of trade constellations.

[8] Aumann and Myerson (1988) provides a (brief) description of the criticism against the use of limited farsightedness in general: 'When a player considers forming a link with another one, he does not simply ask himself whether he may expect to be better off with this link than without it, given the previously existing structure. Rather, he looks ahead and asks himself, "Suppose we form this new link, will other players be motivated to form further new links that were not worthwhile for them before? Where will it all lead? Is the end result good or bad for me?"'

[9] A model similar to that of Saggi and Yildiz (2010), which itself is a modification of the one in Bagwell and Staiger (1997). The modified one is also used in Saggi, Woodland and Yildiz (2013).

The remainder of this paper is organized as follows. Section 2 focuses on the related literature, Section 3 specifies the model, Section 4 analyzes the findings while further details are discussed in Section 5, and Section 6 concludes our paper.

2. Related Literature

An ever increasing body of literature studies the different aspects of international trade agreements. It is not our goal to completely review this stream of literature.[10] The emphasis of this part of our paper is on the methodology of the related papers. Further details, in particular a comparison of the model predictions, can be found in Section 5. In the following, the focus is on the so-called 'rules-to-make-rules' literature (Maggi (2014)) that tries to determine the role of PTAs in the global trade liberalization process.

A number of relevant papers are the work of Saggi, Yildiz and various co-authors. Saggi and Yildiz (2010) considers a three-country trade model where the degree and nature of trade liberalization, bilateral and multilateral, are endogenously determined. Using Coalition-Proof Nash Equilibria, the authors study the stability of FTAs and MTAs while varying the extent of asymmetry among the countries with respect to their size. In a subsequent paper Saggi, Woodland and Yildiz (2013) study the complementary case by focusing on the combination of CUs and MTAs while leaving everything else fixed (in terms of their framework). By contrast, the paper of Missions, Saggi and Yildiz (2016) analyzes the effect of both forms of PTAs, i.e. CUs and FTAs, on attaining global free trade, but excludes MTAs. In a sense, this completes their '2 out of 3' pattern of trade agreements under consideration.

Another related paper (in terms of farsightedness) is the work of Lake (2017), who uses a dynamic approach to understand whether FTAs facilitate or impede the formation of GFT. The approach uses a three-country dynamic model where a fixed protocol specifies for each period the exact nature (and order) of negotiations. Then, on the basis of Markov Perfect Equilibria in pure strategies, the author analyzes the effect of country asymmetries on global trade liberalization.

Furthermore, a variety of research focuses purely on analyzing the effect of FTAs. Goyal and Joshi (2006) consider several countries with a homogeneous good in their model and study different degrees of asymmetry across countries. They employ the notion of Pairwise Stability by Jackson and Wolinsky (1996) as the solution method. Furusawa and Konishi (2007) use similar methods but introduce heterogeneity with respect to goods. In a separate section, they also briefly discuss a setting with CUs, but overall focus on FTAs. Another related paper to Goyal and Joshi (2006) is that of Zhang et al. (2013) in which the concept of Pairwise Stability is replaced with Pairwise Farsighted Stability by Herings, Mauleon and Vannetelbosch (2009), thereby comparing myopia with farsightedness in an otherwise fixed framework. Also connected to this is the paper of Zhang et al. (2014), which uses the work of Goyal and Joshi (2006) as a benchmark and analyzes the evolutionary effect of the number of countries in a dynamic framework featuring random perturbations. Now, while all of the aforementioned papers employ (different) network-theoretic concepts, there is also Aghion et al. (2007), which features standard cooperative game theory. In the three-country model presented there, a single country takes on

[10]The reader may want to consult the papers of Maggi (2014), Grossmann (2016), and Bagwell and Staiger (2016) for a detailed review of the related literature.

the role of negotiation leader and decides to either engage in sequential bilateral or single multilateral bargaining with the other countries.

The stability concept of our approach is that of Chwe (1994). It is (in parts) a response to the criticism of the von-Neumann-Morgenstern stable set (solution).[11] The approach aims to achieve two goals, namely to include unlimited consideration of the future by the participants while simultaneously avoiding emptiness of the stable set that plagues other (more) restrictive solution concepts.[12] It is also closely related to the stability concept found in Herings, Mauleon and Vannetelbosch (2009) and its extension (HMV (2014)). In fact, as is noted by the authors, their criterion constitutes a stricter version, but in specific cases (like our model) they coincide.

3. Model

3.1. **Setting.** Let $N = \{a, b, c\}$ denote the set of all (three) countries in the world. Furthermore, let X denote the set of all trade agreements between these countries, see Section 3.3 for an explicit list. Then, the welfare function of each country induces a collection of preferences on X denoted by $\{\prec_i\}_{i\in N}$, see Section 3.2 for a description of the employed trade model that determine the welfare functions. Moreover, the non-empty subsets S of N specify the coalitions of countries, i.e. the grand coalition, coalitions of two, and single coalitions. Naturally, the preferences of the individual countries induce those of the coalitions, namely for $x_1, x_2 \in X$ and $S \subseteq N$, $S \neq \emptyset$: $x_1 \prec_S x_2$ if and only if $x_1 \prec_i x_2$ for all $i \in S$. Further, the actual ability of coalitions to change the status quo of trade agreements is captured via the collection $\{\rightarrow_S\}_{S\subseteq N,\, S\neq\emptyset}$ of effectiveness relations defined on X, see Section 3.5 for the resulting overall network structure. In combination, the preferences together with the effectiveness relations will allow us to analyze the (potential) stability of different trade agreements, see Section 3.4 for a formal definition of the employed concept of stability. Finally, to determine the stable and unstable trade agreements an algorithm numerically evaluates a grid of the parameter space, see Section 3.6 for details.

3.2. **Underlying Trade Model.** In order to study the stability of different constellations of trade agreements, our framework utilizes a three-country trade model with competition via exports. It will determine the welfare of each country and thereby induce preferences and rankings over all regimes. The model itself follows the one used by Saggi and Yildiz (2010).

Recall that $N = \{a, b, c\}$ denotes the set of countries. Further, let $G = \{A, B, C\}$ denote three (corresponding) non-numeraire goods. Now, each country i is endowed with zero units of good I (corresponding capital letter) and e_i units of the others. Ultimately, it will end up importing I and exporting J and K with $J, K \neq I$. To guarantee the 'competing exporters'-structure, a general condition needs to be applied to the degree of asymmetry with respect to the endowments of the countries. For i and j in N with $i \neq j$, in order for the exports from i to j to be non-negative

[11] Consult von Neumann and Morgenstern (1944) for a description of this (solution) concept and Harsanyi (1974) for its criticism.

[12] It is also resistant to the criticism of Ray and Vohra (2015) about the sovereignty of coalitions as their main issues concerned with feasibility and distribution do not apply to our framework. Furthermore, their specific criticism about the explanatory power of Chwe's approach only applies to transferable utility games.

the condition $3e_j \leq 5e_i$ needs to be satisfied. Thus, the general condition reads:

$$\frac{3}{5}\max\{e_j, e_k\} \leq e_i \leq \frac{5}{3}\min\{e_j, e_k\} \qquad \forall\, i, j, k \in N$$

The preferences of individuals in each country are furthermore assumed to be identical. The demand for any non-numeraire good $L \in G$ in country $i \in N$ is given by the function $d(p_i^L) = \alpha - p_i^L$ with p_i^L the price of good L in country i and the (universal) reservation price α.[13] Each country also (possibly) imposes tariffs on the goods imported by them. Let t_{ij} denote the tariff imposed by country i on the import from country j. All prices and tariffs of a specific good $I \in G$ are connected via the following no-arbitrage condition

$$p_i^I = p_j^I + t_{ij} = p_k^I + t_{ik} \tag{2.1}$$

where $i, j, k \in N$ are pairwise distinct. In this model, the resulting prices together with the corresponding endowments are the only factors influencing imports and exports. In particular, the level of imports m_i^I of good I to country i is completely determined by the demand function (depending on the price), $m_i^I = d(p_i^I) = \alpha - p_i^I$. The exports x_j^I of good I from country j are the combination of the demand function (or prices) and the corresponding endowment, $x_j^I = e_j - d(p_j^I) = e_j + p_j^I - \alpha$. Now, a market-clearing condition for any good I requires that country i's import is equal to the total export of the countries j and k (again $i, j, k \in N$ pairwise distinct):

$$m_i^I = x_j^I + x_k^I \tag{2.2}$$

Ultimately, the objective function of country i is its welfare[14], denoted W_i, which includes Consumer Surplus (CS), Producer Surplus (PS), and Tariff Revenue (TR):

$$W_i = \sum_{L \in G} CS_i^L + \sum_{L \in G \setminus \{I\}} PS_i^L + TR_i$$

Now, CS is composed of three parts itself, namely one for each good. The consumer surplus CS_i^I with respect to the foreign good I is $CS_i^I = \frac{1}{2}(\alpha - p_i^I)m_i^I$ and $CS_i^L = \frac{1}{2}(\alpha - p_i^J)(e_i - x_i^J)$ for a domestic good L. Also, PS splits into two. The producer surplus PS_i^L for a domestic good L is given by $PS_i^L = x_i^L(p_l^L - t_{li}) + (\alpha - p_i^L)p_i^L$. Finally, the tariff revenue TR_i is given by $TR_i = x_j^I t_{ij} + x_k^I t_{ik}$.

3.2.1. *Equilibrium.* Let us start by using no-arbitrage (2.1) and market-clearing (2.2) to compute the equilibrium prices:

$$p_i^I = \frac{1}{3}\left(3\alpha - \sum_{j \neq i} e_j + \sum_{j \neq i} t_{ij}\right)$$

Using these equilibrium values, it is possible to calculate imports, exports, and also the welfare of each country up to the value of the tariffs (Appendix B.1). Note, that the maximization of welfare with respect to tariffs is going to be restricted depending on the trade agreement under consideration, see Section 3.3. For example in the case of MFN, country i maximizes W_i under the restriction that $t_{ij} = t_{ik}$.

[13]The demand function is derived from a utility function that is additively separable and also quadratic in each non-numeraire good.

[14]In certain cases (depending on the trade agreement) the objective function of a country includes the welfare of other countries as well. See Section 3.3 for the details.

Therefore, country i aims to maximize its welfare W_i over $(t_{ij}, t_{ik}) \in T_i$ given $(t_{ji}, t_{jk}) \in T_j$ and $(t_{ki}, t_{kj}) \in T_j$, where T_l is the set of possible tariff pairs for country l in a fixed trade agreement.

The full equilibrium of this model is computed as follows. Fix a trade agreement and thereby the restrictions on the tariffs. Compute the best-response functions for each country (with respect to the tarifs) and determine the optimal choices. While Section 3.3 contains all information on the trade agreements that is necessary to compute the equilibria, the actual results are presented in Appendix B.2. Finally, an overview of the (resulting) overall welfare can be found in Appendix B.3.

3.3. **Trade Policy Arrangements.** All trade relations in our model are one of four types: MFN, CU, FTA, and MTA. Each type, except for MFN, naturally induces different combinations of insiders and outsiders. Namely, three combinations of two members and one of three (each for CU, FTA, and MTA).[15] Additionally, the case of FTA contains the possibility of a special hub structure with two FTAs at the same time - adding another three combinations. In total, our model allows for 16 different trade constellations.[16] For each of these trade agreements the tariffs are bounded from below and above by zero and the MFN-tariff respectively, which is discussed in more detail in Appendix B.2. The corresponding set of tariffs for country i, i.e. $[0, t_i^{MFN}]$, is denoted by T_i. Any additional restrictions on tariffs, specific to trade agreements, are listed here:

In the baseline case, i.e. MFN, countries do not liberalize their trade relations at all, but the non-discrimination principle still applies. Each country unilaterally chooses its (optimal) tariffs accordingly. Therefore, the optimization problem of country i is $\max_{(t_{ij},t_{ik})\in T_i^{MFN}} W_i$ with $T_i^{MFN} = \{(t_{ij}, t_{ik}) \in \mathbb{R}^2_{\geq 0} \,|\, t_{ij} = t_{ik}\}$. Note, that in this reference scenario each tariff is chosen from $\mathbb{R}_{\geq 0}$ instead of T_i.

In case country i and j form CU(i,j), each of them removes any trade restriction on the other country and then jointly imposes an optimal tariff on country k. Thus, the optimization problem of country i and j is $\max_{(t_{ij},t_{ik})\in T_i^{CU},(t_{ji},t_{jk})\in T_j^{CU}} W_i + W_j$ with $T_i^{CU} = \{(t_{ij}, t_{ik}) \in T_i^2 \,|\, t_{ij} = 0\}$ and T_j^{CU} analogous. Finally, country k simply follows and applies the principle of MFN (as before). However, as soon as all three countries enter a single CU together, the (common) optimization problem is trivial, because the only possible tariff of each country towards any other country is zero, and the scenario is denoted by CUGFT.

In case country i and j form FTA(i,j), each of them removes any trade restriction on the other country and then unilaterally imposes an optimal tariff on country k. Thus, the (representative) optimization problem of country i is $\max_{(t_{ij},t_{ik})\in T_i^{FTA}} W_i$ with $T_i^{FTA} = \{(t_{ij}, t_{ik}) \in T_i^2 \,|\, t_{ij} = 0\} (= T_i^{CU})$. The optimization problem of country k is identical to that of the third country in case of a CU. Further, in case

[15]Note that in our model Global Free Trade is essentially listed in three different variations, via CUs, FTAs, and MTAs. The actual welfare is necessarily equal across all three variations, but not their position in the network (Section 3.5). In particular, for our concept of stability it is important which group of countries can create or destroy specific trade agreements (Appendix C.1). Occasionally, all three variants together are going to be referred to as 'GFT' (when applicable).

[16]The framework does not contain combinations of different classes of trade agreement due to the possibly conflicting restrictions on tariffs that the different classes entail. In order to circumvent potential conflicts one would need to fix an (arbitrary) ordering in terms of priority (or importance) of trade agreements, which would reduce the explanatory power more than the inclusion of other combinations of trade agreements would increase it (in our opinion).

country i forms an FTA both with j and k, that is FTAHub(i), then both tariffs of country i are set to zero by nature of its trade relation with both other countries. Each of the other two countries operates as before: Country j (k analogous) faces $\max_{(t_{ji},t_{jk})\in T_j^{FTA}} W_j$ where $T_j^{FTA} = \{(t_{ji}, t_{jk}) \in T_i^2 \,|\, t_{ji} = 0\}$. Thus, in terms of decision problem, it does not matter for a country whether its partner also forms another trade agreement with the other country. Finally, if all three countries in pairs of two countries form FTAs, then the optimization problem is identical to the case of CUGFT, denoted FTAGFT, but the actual trade agreement is different in terms of structure and network position, see Section 3.5.

In case country i and j form MTA(i,j), then both jointly change their tariffs with respect to each other and also for the third country (at the same time). Thus, the optimization problem of country i and j is $\max_{(t_{ij},t_{ik})\in T_i^{MTA},(t_{ji},t_{jk})\in T_j^{MTA}} W_i + W_j$ with $T_i^{MTA} = \{(t_{ij}, t_{ik}) \in T_i^2 \,|\, t_{ij} = t_{ik}\}$ and T_j analogous. As seen before, the optimization problem of country k is identical to that of the third country in case of a CU. Again, as soon as all three countries enter a single MTA together, the optimization problem is identical to the case of CUGFT, denoted MTAGFT, but also different in terms of network position, see Section 3.5.

3.4. **Stability Concept.** As concept of stability our framework makes use of the approach of Chwe (1994).[17] Consider the tuple $\Gamma = \big(N, X, \{\prec_i\}_{i\in N}, \{\rightarrow_S\}_{S\subseteq N, S\neq\emptyset}\big)$ that correspondingly describes the evolution of the status quo of trade agreements driven by the combination of preferences and effectiveness relations:

Let $x \in X$ be the status quo of trade agreements at the start. Next, each coalition $S \subseteq N$, $S \neq \emptyset$ (including individuals) is able to make $y \in X$ the new status quo as long as $x \rightarrow_S y$. Continue with such y as the new status quo. If a status quo $z \in X$ is reached without any coalition moving away, then the state is actually realized and each country receives their corresponding welfare.[18] In consequence, any coalition only favors following through on their ability to move, $x \rightarrow_S y$, when preferring the final welfare over the current one, $x \prec_S z$. Formally, this comparison of states by (chains of) coalitions is captured in the definition of direct and indirect dominance:

Definition 1 (Dominance)**.** *Let $x_1, x_2 \in X$. Then,*

i) x_1 is directly dominated by x_2, write $x_1 < x_2$, if there exists $S \subseteq N$, $S \neq \emptyset$, such that $x_1 \rightarrow_S x_2$ and $x_1 \prec_S x_2$.

ii) x_1 is indirectly dominated by x_2, write $x_1 \ll x_2$, if there exist sequences $y_0, y_1, \ldots, y_m \in X$ (with $y_0 = x_1$ and $y_m = x_2$) and $S_0, S_1, \ldots, S_{m-1} \subseteq N$, such that $S_i \neq \emptyset$, $y_i \rightarrow_{S_i} y_{i+1}$, and $y_i \prec_{S_i} y_m$ for $i = 0, 1, \ldots, m-1$.

Note, that if $x_1 < x_2$ for some $x_1, x_2 \in X$, then automatically $x_1 \ll x_2$.

Using this definition, the concept of 'consistent set' describes a (sub-)set that exhibits internal stability in the form of a lack of incentive to deviate:

Definition 2 (Consistent Set)**.** *A set $Y \subseteq X$ is consistent if $y \in Y$ if and only if for all $x \in X$ and all $S \subseteq N$, $S \neq \emptyset$, with $y \rightarrow_S x$ there exists $z \in Y$ where $x = z$ or $x \ll z$ such that $y \not\prec_S z$.*

[17] Consult the paper of Chwe (1994) for the proofs of the propositions that are presented here.

[18] Technically, the model is without any true sense of time. Any start (or end) as well as any sequence of actions should be interpreted as a thought-experiment. Furthermore, a path created in this fashion is generally not unique.

In general, a consistent set is not necessarily unique, but the following proposition allows us to talk about the unique 'largest consistent set', i.e. the (consistent) set that contains all consistent sets:

Proposition 1. *There uniquely exists a $Y \subseteq X$ such that Y is consistent and $Y' \subseteq X$ consistent implies $Y' \subseteq Y$. The set Y is called the largest consistent set or simply LCS.*

Or put differently, it is the unique fixed point of the correspondence $f\colon 2^X \to 2^X$ defined by

$$Y \mapsto f(Y) = \{y \in X \,|\, \forall x \in X, \forall S \subseteq N, S \neq \emptyset, \text{ with } y \to_S x\colon \\ \exists z \in Y \text{ s.t. } (x = z \text{ or } x \ll z) \wedge y \not\succ_s z\}.$$

Now, similar to the internal stability captured in the definition of consistent sets, a form of external stability is captured via an incentive to gravitate towards the consistent set:

Definition 3 (External Stability)**.** *Let $Y \subseteq$ be the largest consistent set. Then, it satisfies the external stability condition if for all $x \in X \setminus Y$ there exsists $y \in Y$ such that $x \ll y$.*

The following result characterizes one setting in which this condition is satisfied:

Proposition 2. *Let X be finite and the underlying preferences irreflexive. Then, the LCS is non-empty and satisfies the external stability property.*

Finally, let us state a couple of comments on the application and interpretation of this stability concept with respect to our model:

3.4.1. *Application.* First of all, applying Proposition 1 to our model is trivial, because it is stated without any (additional) requirements on the involved objects. Furthermore, the application of Proposition 2 is straight forward as well: First, the set of outcomes X is clearly finite in our setting as we are only considering a finite number of different trade agreements. Second, any strict preference is automatically irreflexive and our preferences are induced by strict welfare comparisons. Thus, while the definition of the (largest) consistent set in general only guarantees internal stability, our setting actually implies external stability as well:

Corollary 1. *In our setting, the (unique) LCS is non-empty and satisfies the external stability property (in addition to the internal stability).*

Now, the LCS is going to be the focus point of our analysis. Any trade agreement is considered to be '(potentially) stable' if it is in the LCS, 'unstable' otherwise. The nomenclature is a tribute to the fact that the LCS as a stability concept is 'weak: not so good at picking out, but ruling out with confidence', because ultimately it 'does not try to say what will happen but what can possibly happen' (Chwe (1994)).

3.5. **Network Structure.** The complete network structure consists of a collection of transition matrices $\{A_S\}_{S\subseteq N, S\neq\emptyset}$ induced by $\{\to_S\}_{S\subseteq N, S\neq\emptyset}$. Let $S \subseteq N, S \neq \emptyset$ be any coalition, then the entry $(A_S)_{x_i,x_j}$ is 1 if $x_i \to_S x_j$ and 0 otherwise. Thus, the matrix for $\{a, b, c\}$, the full coalition, is simply given by $(A_{\{a,b,c\}})_{x_i,x_j} = 1$ for all $x_i, x_j \in X$. Further, each of the transition matrices induces a directed graph with the trade agreements as vertices and the effectiveness relations as edges. Therefore,

the corresponding directed graph of the full coalition is a complete directed graph with loops.

It is noteworthy to point out that the relation (or transition) $x \to_S x$ holds for all trade agreements x and all coalitions S, but is ultimately irrelevant for the analysis with respect to the stability. The reason for this is the fact that our model contains no sense of time - essentially stalling negotiations serves no purpose.[19] Therefore, these transitions are ignored from now on or, put differently, the framework only considers a form of equivalence classes, namely modulo loops. Furthermore, whenever coalition S is able to destroy one trade agreement, say x_1, and subsequently create another one, say x_2, then it is able to move directly, i.e. $x_1 \to_S x_2$. Finally, for the remaining coalitions (of two and one country) only the transition graphs are presented here. The corresponding transition tables can be found in Appendix B.4.

Let us now consider the transition graph for a single country coalition $i \in N$ with $j, k \in N \setminus \{i\}$, $j \neq k$, denoting the other two countries. In this case, MFN is connected to a number of other different elements, but not to the three variants of Global Free Trade, CUGFT, FTAGFT and MTAGFT. Now, each of those forms a separate group of connected trade agreements. Thus, the overall transition graph, see Figure 1 (modulo loops), consists of four sub-graphs.

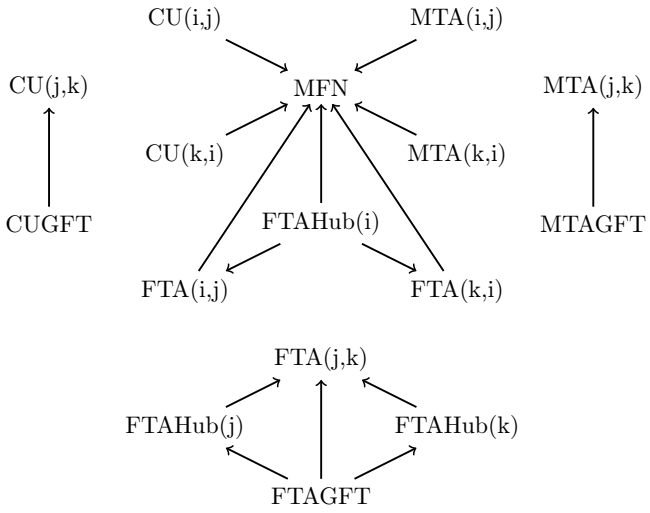


FIGURE 1. The transition graph for coalition $\{i\}$, $i \in N$.

Finally, consider the transition graph for a coalition of two countries $i, j \in N$, $i \neq j$ with $k \in N \setminus \{i, j\}$ denoting the other country. In this case, MFN, CU(i,j), FTA(i,j), and MTA(i,j) are all interconnected. Also, any element connected to one of these is automatically connected to all of them. Thus, in the transition graph, see Figure 2 (again, modulo loops), this group of four corresponds to a complete directed sub-graph pictured as one '(super) node' (dotted box).

[19]While staying in one trade constellation, the overall strategic situation remains the same. Specifically, for each country and each coalition the welfare of each trade agreements only depends on the parameters of the underlying trade model. Similarly, the network structure stays constant. Additionally, the number of (potential) movements in a chain of trade agreements is unlimited.

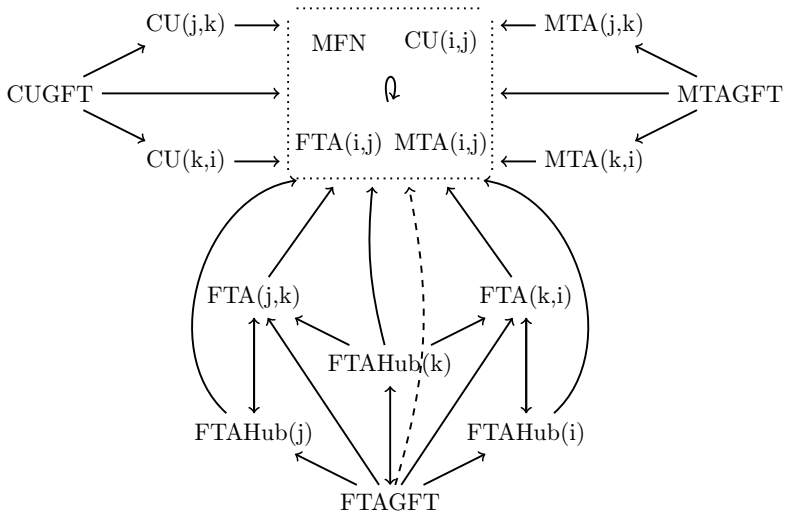


FIGURE 2. The transition graph for coalition $\{i, j\}$, $i, j \in N$, $i \neq j$.

3.6. **Algorithm and Parameters.** The (additional) explanatory power from the introduction of an extensive set of trade agreements and unlimited farsightedness comes at the cost of a complex computational problem. This problem is solved numerically with the help of an algorithm - the pseudocode of which can be found in Appendix A while the corresponding sourcecode is publicly available under the address `http://doi.org/10.4119/unibi/2931412`.[20] The parameter space therefore needs to be specified and discretized:

First, recall that the endowments satisfy $\frac{3}{5}\max\{e_j, e_k\} \leq e_i \leq \frac{5}{3}\min\{e_j, e_k\}$ for all $i, j, k \in N$ in order to guarantee the 'competing exporters'-structure, see Section 3.2. Now, without loss of generality, normalize one endowment to one, namely $e_b = 1$. Consequentially, for $i, j \in N \setminus \{b\}$: $e_{\min} := \frac{3}{5} \leq \frac{3}{5}\max\{1, e_j\} \leq e_i$ and $e_i \leq \frac{5}{3}\min\{1, e_j\} \leq \frac{5}{3} =: e_{\max}$. Furthermore, the resulting parameter space, Figure 3, can be split into six right-angled triangles, which are mirror images of one another (in terms of relative endowments). Thus, again without loss of generality, focus on one of them, namely the marked triangle, and then cover it with a grid for the actual computation.[21]

Additionally, to produce plausible results, e.g positive prices, the factor α needs to be chosen above a minimal value for each tuple of endowments, $\alpha_{\min}(e_a, e_b, e_c)$. Above this minimal value, the results remain unchanged.[22] Thus, by taking the maximum over all these minimal values, $\alpha_{\max\min} = \max_{e_a,e_b,e_c}\{\alpha_{\min}(e_a, e_b, e_c)\}$, adding an epsilon, $\alpha = \alpha_{\max\min} + \epsilon$, and using it for all endowments makes sure that all results are plausible and comparable at the same time.[23]

[20]The authors are grateful to Michael Chwe for the provision of an exemplary algorithm.

[21]The distance is set to 0.0013360053440215 - due to 500 points per dimension of the grid.

[22]The factor α always enters the welfare of country i as $2\alpha e_i$ (see Appendix B.1). Therefore, any changes above the minimal value leave the welfare levels and therefore the rankings unaffected.

[23]In our computation ϵ is simply fixed to 0.01, which yields $\alpha = 1.3988888888888888$.

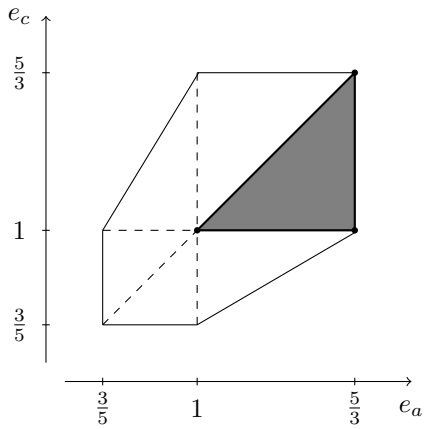


FIGURE 3. The parameter space of the endowments with $e_b = 1$

4. Analysis

Let us now present the resulting structure of stability among trade agreements according to our framework. Figure 4 depicts the parameter space of endowments under consideration for this - it is the (marked) triangle from before. The analysis starts with the three extreme points, then turns to the connecting intervals, and finishes with the entire interior. In each of these cases, two scenarios are examined. The first scenario corresponds to the current WTO institutional arrangement while the second one assumes modified WTO rules without Article XXIV Paragraph 5, which would prevent the formation of PTAs (specifically CUs and FTAs).

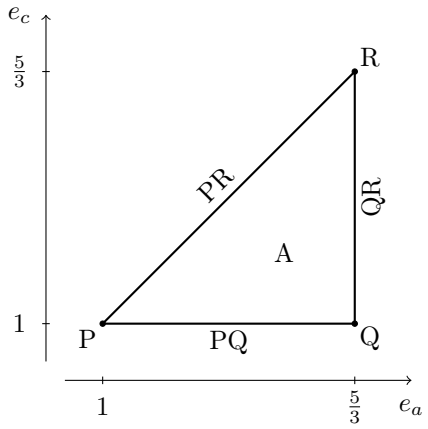


FIGURE 4. Overview of the different points, intervals and areas of interest depending on the (partially normalized) endowment tuple

The remainder of this analysis is structured as follows. First, Section 4.1 considers the symmetric case, see point P in Figure 4, where all countries are identical. Second, Section 4.2 features the two extreme asymmetric cases, points Q and R, with countries that are small, small, and large (Q) or small, large, and large (R). Next, Section 4.3

discusses the three related intervals, sides PQ, QR, and PR, where the countries are small, small, and varying (PQ), small, large, and varying (QR), or small with two varying equally (PR). Finally, Section 4.4 describes the inner area, area A, with three distinct countries.

4.1. **Symmetric Case.** First, let us consider the symmetric case, where symmetry refers to identical endowments for all countries, i.e. $e_a = e_b = e_c = 1 = e_{\min}$, and corresponds to point P in the triangle of Figure 4. As the countries do not differ from one another, the only thing that matters for welfare is whether a country is an insider or an outsider in a specific trade agreement. In the following, we present the ranking of preferences from the perspective of country a, which represents that of all other countries as well, for fixed $i, j \in N \setminus \{a\}$ with $i \neq j$:

$$CU(i,j) \prec_a MFN \prec_a MTA(a,i) \prec_a FTAHub(i) \prec_a FTA(i,j) \prec_a FTA(a,i) \\ \prec_a CU(a,i) \prec_a MTA(i,j) \prec_a GFT \prec_a FTAHub(a)$$

The case where two countries form a CU is the least favorable trade constellation for the third country. Under such circumstances, the outsider faces the second-highest tariffs (with MFN-tariffs the highest), while the insiders cancel the tariffs among themselves. The exports of country a to the other countries, i and j, are the lowest under CU(i,j) compared to all alternative trade agreements. The same applies to the total imports. In other words, the 'trade diversion' effect is the strongest for country a in case of CU(i,j). In general, the MFN regime favors country a when compared to CU(i,j). The tariff revenue remains the same, while the consumer surplus is lower and the producer surplus is higher - the increase offsets the decrease. The MFN regime slackens the 'trade diversion' effect present in the case of CU(i,j) by virtue of increased export values of country a.

Among the group of bilateral trade agreements where the country is an insider, the MTAs result in the lowest welfare (for this country). MTA(a,i) itself generates a higher welfare for country a in comparison with the MFN regime on the grounds of increased consumer and producer surplus. The FTAHub(i) constellation results in even further gains in welfare for country a through higher export values and producer surplus accordingly (the tariff revenue and also the consumer surplus are lower under FTAHub(i) compared to MTA(a,i) though). However, country a does not have an incentive to remain in this constellation. The unilateral deviation from FTAHub(i) to FTA(i,j) comes with a decrease of consumer and producer surplus but enough increase in tariff revenue to ultimately ensure higher welfare under the latter constellation. Nonetheless, among FTAs being an outsider is less desirable than being an insider for any country. The drop in tariff revenue is offset by an expansion of the consumer and producer surplus, resulting in higher welfare for country a in case of FTA(a,i) compared to FTA(i,j). As an insider, country a prefers CU(a,i) over FTA(a,i) though. More precisely, in spite of the decline in the consumer surplus, the actual welfare goes up through an expansion of tariff revenue and producer surplus.

The formation of MTA(i,j) guarantees the highest welfare for country a compared to any other bilateral trade agreement. The driving factor is the MFN-principle, which implies that in case of MTA(i,j) the insiders need to apply the same tariff to both each other and the outsider - a form of free-rider problem. At the same time, country a attains the highest possible tariff revenue.

Each country obtains the second-highest welfare level when the world reaches global free trade. Under full trade liberalization, the producer surplus is also the

second-highest among all trade agreements (effectively driving the ranking). It is only surpassed by that of FTAHub(a). The latter constellation brings about the highest possible welfare for country a. But note that such a trade agreement disproportionally favors the hub country over the other countries.

Countries' strong preference rankings are the crucial ingredient for computing the LCS. In fact, for each country all three variants of global free trade are ranked as second-best option while each first-best option, a hub structure, is ranked considerably lower for the other countries. Intuitively, global free trade seems like a stable compromise. The following proposition and its proof reinforce this:

Proposition 3. *Under symmetry and with the current institutional arrangement of the WTO, the LCS contains three elements: CUGFT, FTAGFT, and MTAGFT. In other words, (the trinity of) global free trade is the unique stable outcome.*

Proof. Based on the definition of indirect dominance and the transition graphs, see Section 3.4 and 3.5, the preference rankings from earlier allow us to derive the indirect dominance matrix. If the entry in the matrix is equal to one (resp. zero), then the trade arrangement corresponding to the row of the entry is (resp. isn't) indirectly dominated by the one corresponding to the column of the entry. For example, $FTAHub(a)$ is indirectly dominated by $CUGFT$ as there exists a (finite) sequence of outcomes and coalitions such that all coalitions in the sequence prefer the final outcome over the current one:

$$FTAHub(a) \rightarrow_{\{b,c\}} CU(b,c) \rightarrow_{\{a,b,c\}} CUGFT$$

Checking for all possible sequences yields the following indirect dominance matrix:[24]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 MFN	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2 $CU(a,b)$	0	0	0	0	1	0	0	0	1	1	0	1	0	1	1	1
3 $CU(b,c)$	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	1
4 $CU(c,a)$	0	0	0	0	1	0	0	0	1	0	1	1	1	1	0	1
5 $CUGFT$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 $FTA(a,b)$	0	1	1	1	1	0	0	0	1	1	0	1	0	1	1	1
7 $FTA(b,c)$	0	1	1	1	1	0	0	0	0	1	1	1	1	0	1	1
8 $FTA(c,a)$	0	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1
9 $FTAHub(a)$	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	1
10 $FTAHub(b)$	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	1
11 $FTAHub(c)$	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	1
12 $FTAGFT$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13 $MTA(a,b)$	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
14 $MTA(b,c)$	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
15 $MTA(c,a)$	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
16 $MTAGFT$	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0

Note that intuitively any outcome is stable if all deviations from it are deterred. Also, a deviation from the outcome is hindered if there is a stable outcome which might be reached and some member of the deviating coalition does not prefer it over the initial outcome. In the following procedure, start with the full set and then keep removing elements that are unstable until the remaining ones are stable

[24] Appendix A contains the pseudocode for this procedure.

Take $x \in \{MFN, FTA(i,j), CU(i,j), MTA(i,j)\}$, where $i,j \in N$ with $i \neq j$, and then consider the joint deviation $x \rightarrow_{\{a,b,c\}} FTAGFT$. The $FTAGFT$ regime is not indirectly dominated by any other outcome (see the matrix above) and also $x \prec_{\{a,b,c\}} FTAGFT$ for each of those x. Thus the deviation $x \rightarrow_{\{a,b,c\}} FTAGFT$ cannot be deterred and therefore no such x can be part of the stable set.

Consider $FTAHub(i)$, $i \in N$, and the deviation $FTAHub(i) \rightarrow_{\{j,k\}} FTAGFT$, $j,k \in N \setminus \{i\}$ with $j \neq k$. Using $FTAHub(i) \prec_{\{j,k\}} FTAGFT$ together with the logic from before eliminates $FTAHub(i)$ for each $i \in N$.

Focus on the set of remaining elements $Y = \{CUGFT, FTAGFT, MTAGFT\}$. Start with any element y in Y. If there is a deviation to any element $x \in X \setminus Y$, then there always exists an indirect dominance path (see indirect dominance matrix) $x \ll y'$ coming back to an element $y' \in Y$. In addition, for any $y_1, y_2 \in Y$, $y_1 \neq y_2$, there does not exist a coalition $S \subseteq N$, $S \neq \emptyset$, for which $y_1 \prec_S y_2$. Thus, the set Y satisfies the (internal) stability condition while being maximal, i.e. $Y = LCS$. □

In the symmetric case, under the current institutional arrangement of the WTO, the global free trade variations appear as the only stable constellation according to our framework. But what would happen without Article XXIV Paragraph 5? In this case, countries would not have the option to liberalize trade through the formation of CUs or FTAs - leaving MTAs as the only possibility. The representative preference ranking of country a would look as follows:

$$MFN \prec_a MTA(a,i) \prec_a MTA(i,j) \prec_a MTAGFT$$

Each country achieves the peak welfare under MTAGFT. Thus, it is reasonable to conjecture stability of MTAGFT. The following proposition proves this intuition:

Proposition 4. *Under symmetry and with the modified institutional arrangement of the WTO (no PTAs), the LCS contains one element: MTAGFT. In other words, global free trade is the unique stable outcome.*

Proof. The indirect dominance matrix is derived as before:

$$\begin{array}{l} \\ 1\,MFN \\ 2\,MTA(a,b) \\ 3\,MTA(b,c) \\ 4\,MTA(c,a) \\ 5\,MTAGFT \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Let us start with the full set of trade agreements again (limited to the setting). If the grand coalition moves from MFN to $MTAGFT$, then the only possibility is to stay there, as $MTAGFT$ is not indirectly dominated by any other outcome. Moreover, $MFN \prec_{a,b,c} MTAGFT$. Thus, MFN cannot be stable. Furthermore, if the grand coalition moves from any bilateral MTA regime to $GFTMTA$, by the same argument, it is clear that no bilateral MTA can be stable. Finally, any deviation from $MTAGFT$ will come back to itself due to the indirect dominance. Consequentially, the set $Y = \{MTAGFT\}$ is consistent and also the largest one. □

If symmetry among all countries holds, then Article XXIV Paragraph 5 does not change anything in terms of stability and corresponding welfare, both individual and overall. The only stable trade constellation is (the trinity of) global free trade.

4.2. **Asymmetric Case - Vertices of the Triangle.** It is natural to start the analysis of the asymmetric case by considering its two extreme scenarios, which correspond to the points Q and R in the triangle of Figure 4. In the following, Section 4.2.1 discusses the case of countries that are small, small, and large (Q) while Section 4.2.2 focuses on countries that are small, large, and large (R).

4.2.1. ***The case of two small and one large country.*** In this scenario, fix $e_a = e_{max}$ and $e_b = e_c = e_{min}$ (point Q). Let us start with the ranking of preferences for country a and another country $i \in N \setminus \{a\}$ - representing also $j \in N \setminus \{a, i\}$:

$$\begin{aligned} CU(i,j), FTA(i,j) \prec_a FTAHub(i) \prec_a MFN, MTA(i,j) \prec_a FTA(a,i) \\ \prec_a MTA(a,i) \prec_a CU(a,i) \prec_a GFT, FTAHub(a) \end{aligned}$$

$$\begin{aligned} MTA(a,i) \prec_i GFT, FTAHub(a) \prec_i CU(a,i) \prec_i FTA(a,i) \prec_i CU(a,j) \\ \prec_i MFN, MTA(i,j) \prec_i FTAHub(i) \prec_i FTA(a,j) \prec_i MTA(a,j) \\ \prec_i FTAHub(j) \prec_i CU(i,j), FTA(i,j) \end{aligned}$$

One immediately notices that small and large countries have different rankings. A large country profoundly dislikes the scenarios where it is an outsider; while the small countries, by contrast, dislike any trade arrangements with the large country. Note that in certain cases countries actually do not differentiate between different trade constellations.[25] For example, CU(i,j) and FTA(i,j) result in same welfare for all countries. In this case, under the given pattern of endowments, the optimal tariffs of the small countries for CU and FTA are above the MFN-tariff. However, the Sub-paragraphs of Article XXIV Article 5 rule this out and therefore the tariffs are capped at the MFN-level. A similar argument applies to the case of FTAHub(a). Here, the optimal tariffs of the small countries would be negative. By restricting tariffs from below by zero implies that FTAHub(a) corresponds to GFT, or rather a Pseudo-GFT. Finally, the MTA between the small countries actually coincides with the MFN regime because of identical optimal tariffs for both cases.

Next, let us analyze the preferences of the large country a. As mentioned above, being the outsider produces the least favorable constellations for a large country. The worst scenarios are those where the small countries form a PTA. In such cases, the export, and hence the producer surplus, is the lowest in the large country. Now, compared to these PTAs among the small countries, both the tariff revenue and the consumer surplus are lower under FTAHub(i), but the comparably strong growth of the producer surplus produces an increase in welfare. Further increases in producer surplus are possible via the MFN regime or MTA(a,i). As soon as the large country forms an FTA, its welfare increases due to trade relations that benefit its exports within the constellation. MTA(a,i) leads to an even higher welfare, but there the driving factor are the tariff revenues. Among all bilateral trade agreements, where the large country is an insider, the optimal outcome is CU(a,i). The highest welfare for the large country occurs when trade is fully liberalized though. In that scenario, it is able to completely reap the trade benefits - the producer surplus peaks when compared to the other trade agreements.

Now, let us consider the preferences of a small country i, the countries i and j are indistinguishable from each other in this respect. Contrary to the preferences of a large country, it is in its best interest for a small country to avoid forming

[25] Additional details on this can be found in Appendix B.2 and Appendix B.5.1.

any trade arrangement with the large country. The smallest welfare of a small country occurs in the case of MTA(a,i), as the constellation generates one of the least desirable combinations of tariff revenue and producer surplus. Under FTAHub(a) or GFT, the small country achieves higher welfare through an increase in producer surplus and despite a decrease in tariff revenue and consumer surplus. CU(a,i) and FTA(a,i) each lead to further welfare improvements for the small country. In both cases, the driving factor is a higher consumer surplus. Next, regardless of the lower consumer and producer surplus under CU(a,j) (compared to FTA(a,i)), its higher tariff revenue actually results in a higher welfare. The tariff revenue stays at its peak under MTA(i,j) or the MFN regimes as well, but with higher welfare. Specifically, an increase in consumer surplus offsets the decrease in producer surplus. By comparison, FTAHub(i) actually increases the producer surplus and thereby also the total welfare. FTA(a,j) then generates its peak tariff revenue from before and through this a higher welfare for the small country as an outsider. Under MTA(a,j) the tariff revenue stays the same and the lower producer surplus gets (more than) compensated by the higher consumer surplus. Compared to this, FTAHub(j) decreases the tariff revenue but increases both the consumer and the producer surplus enough to increase the welfare. Finally, a small country attains the best result by forming a PTA with the other small country, either through FTA(i,j) or CU(i,j). While keeping relatively high tariff revenues, the small countries manage to have a high producer surplus as well.

Using the preference rankings to derive the LCS yields the following proposition:

Proposition 5. *With the endowments given by $e_a = e_{max}$ and $e_b = e_c = e_{min}$, and under the current institutional arrangement of the WTO, the stable constellations are the PTAs between the two small countries, that is CU(b,c) and FTA(b,c).*

Proof. Let us start by giving the indirect dominance matrix:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 MFN	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
2 $CU(a,b)$	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0
3 $CU(b,c)$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 $CU(c,a)$	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0
5 $CUGFT$	1	1	1	1	0	0	1	0	0	0	0	0	0	1	0	0
6 $FTA(a,b)$	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
7 $FTA(b,c)$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8 $FTA(c,a)$	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
9 $FTAHub(a)$	1	0	1	0	0	1	1	1	0	1	1	0	0	1	0	0
10 $FTAHub(b)$	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
11 $FTAHub(c)$	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
12 $FTAGFT$	1	0	1	0	0	1	1	1	0	1	1	0	0	1	0	0
13 $MTA(a,b)$	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
14 $MTA(b,c)$	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
15 $MTA(c,a)$	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
16 $MTAGFT$	1	1	1	1	0	0	1	0	0	0	0	0	1	1	1	0

Recall that X denotes the full set and let $Y = \{CU(b,c), FTA(b,c)\}$ be the candidate for the LCS. Take any element x from the set $X \setminus Y$ and consider the deviation $x \to_{\{b,c\}} CU(b,c)$. Note that $CU(b,c)$ is not indirectly dominated by any

other element from X and furthermore $x \prec_{\{b,c\}} CU(b,c)$ for all $x \in X \setminus Y$. Thus, the deviation $x \rightarrow_{\{b,c\}} CU(b,c)$ can not be deterred for all $x \in X \setminus Y$. Therefore, no such x can be part of the stable set.[26]

As each outcome in $X \setminus Y$ is indirectly dominated by $y \in Y$ (see the matrix), for any coalition and any deviation away from $y \in Y$ there always exists a path of indirect dominance back to Y. Moreover, no coalition is actually better off when coming back to Y, as $x \not\succ_S y$ for all $x, y \in Y$, $x \neq y$, and $S \subseteq N$, $S \neq \emptyset$. Therefore, the set Y satisfies the (internal) stability condition while being maximal, i.e. $Y = LCS$. □

Even though global free trade is the most desirable regime for the large country, the two small countries do not have any incentive to form such a constellation and the large country can not enforce it. As a consequence, country a ends up with the worst trade agreement (from its perspective). Thus, in this scenario the size advantage of the large country does not translate into a favorable stable regime. Moreover, this specific case showcases the relevance of the restrictions on PTAs (remember that insiders are not allowed to raise tariffs on outsiders). The constraint makes the small countries be indifferent between the two forms of PTAs.

Now we turn to the hypothetical scenario without Article XXIV Paragraph 5. Here, the ranking of preferences for the countries, with country a the large one and country b and c small (represented by i and j), are as follows:

$$MTA(i,j), MFN \prec_a MTA(a,i) \prec_a GFT$$

$$MTA(a,i) \prec_i MTAGFT \prec_i MFN, MTA(i,j) \prec_i MTA(a,j)$$

As a result, the best outcome for a small country i is the $MTA(i,j)$ regime, as the PTAs are not available anymore. The next proposition presents the new LCS as a consequence of these changes:

Proposition 6. *With the endowments given by $e_a = e_{max}$ and $e_b = e_c = e_{min}$, and under a modified institutional arrangement of the WTO, the stable constellations are MFN and MTA(b,c).*

Proof. The indirect dominance matrix is given as follows:

$$\begin{array}{l} \\ 1\,MFN \\ 2\,MTA(a,b) \\ 3\,MTA(b,c) \\ 4\,MTA(c,a) \\ 5\,MTAGFT \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

Start with the full set again. If we consider the deviations $MTA(c,a) \rightarrow_c MFN$ and $MTA(a,b) \rightarrow_b MFN$, then no further deviations are expected as MFN is not indirectly dominated by any other outcome. In addition, $MTA(c,a) \prec_c MFN$ and $MTA(a,b) \prec_b MFN$, so $MTA(c,a)$ and $MTA(a,b)$ cannot be part of the stable set. The same argument works in the case of $MTAGFT$ and the deviation

[26] It might appear that this proof deviates from the general approach of eliminating element by element from the full set until the remainder forms the stable set. However, in this proof it is purely a coincidence that in one step all elements but the stable ones can be eliminated with one argument (or rather deviation).

$MTAGFT \rightarrow_{b,c} MFN$, as $MTAGFT \prec_{b,c} MFN$. So, the global free trade regime cannot be stable as well.

Let $Y = \{MFN, MTA(b,c)\}$. Following any deviation from the elements in Y, there is always an indirect dominance path coming back to Y (MFN in this case). In addition, for any $x, y \in Y$ with $x \neq y$ there does not exist coalition S for which $x \prec_S y$. Thus, the set Y is consistent and the largest one as well. □

In summary, when there are two small and one large country, the GFT regime is unstable under the current and hypothetical institutional set-up of the WTO. At best, world trade can be partially liberalized. Additionally, the small countries profit when they can form a PTA instead of an MTA, as the limiting MFN principle can be avoided that way.

4.2.2. ***The case of one small and two large countries***. In this scenario, fix $e_b = e_{min}$ and $e_a = e_c = e_{max}$ (point R). Let us start with the ranking of preferences for country b and another country $i \in N \setminus \{b\}$ - representing also $j \in N \setminus \{b, i\}$:

$$GFT \prec_b CU(i,b) \prec_b FTAHub(i) \prec_b FTA(i,b) \prec_b FTAHub(b) \prec_b MTA(i,b) \prec_b MFN \prec_b CU(i,j) \prec_b MTA(i,j) \prec_b FTA(i,j)$$

$$CU(j,b) \prec_i FTA(j,b) \prec_i MFN \prec_i MTA(i,b), MTA(j,b) \prec_i FTA(i,j) \prec_i MTA(i,j) \prec_i CU(i,j) \prec_i FTAHub(b) \prec_i FTAHub(j) \prec_i GFT \prec_i FTAHub(i) \prec_i FTA(i,b) \prec_i CU(i,b)$$

Under the given pattern of endowments, the preference rankings of the countries are considerably different from the previous cases. For the small country, the MFN regime generates higher welfare than any other trade agreement where it is part of. As for a large country, being an outsider is on the lower end of the ranking, while being an insider in a PTA with a small country is on the other end.

Let us take a closer look at the preference ranking of the small country. First, GFT actually generates the lowest total welfare - driven by no tariff revenue and not enough compensation via consumer and producer surplus. As mentioned before, any trade arrangement involving the small country results in lower welfare compared with other constellations (but higher welfare than GFT). The lowest among those are the CU with any of the large countries, which through increased tariff revenue (and despite a decrease in consumer surplus) yield higher welfare in comparison with the GFT regime. Even though FTAHub(i) reduces those gains in tariff revenue again, by virtue of a growing consumer surplus it still raises the total welfare. Further improvement in the welfare of the small country is possible if the world moves from FTAHub(i) to FTA(i,b); the sole reason is a higher consumer surplus. Under FTAHub(b), the export volumes to the large countries are at its peak and it generates substantially higher producer surplus. As a consequence, it results in the small country preferring to form a hub structure (as the hub node) over an FTA with one of the large countries. Replacing the FTA with an MTA with similar structure is the most desirable configuration for the small country among the constellations where it participates. Under MTA(i,b) the producer surplus is actually the smallest compared to all other alternatives, but high tariff revenue and consumer surplus determine its position in the ranking. The MFN regime surpasses all configurations mentioned above. When there are two large countries, the tariff revenue becomes an important factor in the welfare of the small country. Any further improvements with

respect to the welfare of the small country depend on the large countries liberalizing trade among themselves - the small country essentially free-rides in these cases (exhausting its tariff revenue to the fullest). The driving factor among these three is the export volume. Consequentially, CU(i,j) is the worst option, followed by MTA(i,j), and FTA(i,j) is the (overall) best outcome.

The following discusses the preferences of the two large countries. The least favorable scenario occurs when the other large country forms a CU together with the small country. Its position in the ranking is driven by the lowest export volumes and producer surplus. Now, FTA(j,b) produces higher welfare compared to the previous constellation due to growth in producer surplus (based on rising exports to the small country) which makes up for the drop in consumer surplus. A similar development makes the MFN regime an even better constellation (here the exports to the large country increase). All tariffs (and thus prices) are identical under both MTA(i,b) and MTA(j,b), as a consequence they generate the same welfare. On the grounds of increased exports, the welfare tops that of the MFN regime. Among the class of bilateral trade agreements between the large countries, the ranking goes as follows: FTA(i,j) followed by MTA(i,j) only surpassed by CU(i,j). In comparison with MTA(i,b) and MTA(j,b), the greater consumer and producer surplus of FTA(i,j) guarantees an increase of total welfare. An MTA between the two large countries produces more tariff revenues and actually results in a more desirable outcome. Moving from MTA(i,j) to CU(i,j) decreases tariff revenue and also consumer surplus but the gain in producer surplus through increased exports to the other large country makes more than up for this. FTAHub(b) and even more so FTAHUb(j) further improve the welfare via growth of the tariff revenue and consumer surplus (the case of FTAHub(b)), and increased exports to the other large country (for FTAHUb(j)). Now, the GFT regime allows the large country to raise the exports to the small country while retaining the same level of exports to the other large country. As a consequence, the welfare of GFT surpasses that of the previous mentioned constellations. However, when the large country is part of a hub structure as the hub node itself, then its exports to the small country increase such that the welfare exceeds that of full trade liberalization. Furthermore, the FTA with the small country constitutes the second-best outcome for the large country on the grounds of high tariff revenue accompanied by similar consumer surplus. Finally, CU(i,b) is the most desirable constellation driven by the high exports to the small country.

Let us compute the LCS under these preference rankings in the next proposition:

Proposition 7. *With the endowments given by $e_b = e_{min}$ and $e_a = e_c = e_{max}$, and under the current institutional arrangement of the WTO, the stable constellation is the CU between the two large countries, that is CU(c,a).*

Proof. The indirect dominance matrix is given as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 MFN	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0
2 $CU(a,b)$	1	0	0	1	0	1	1	1	0	1	0	0	1	1	1	0
3 $CU(b,c)$	1	0	0	1	0	1	1	1	0	1	0	0	1	1	1	0
4 $CU(c,a)$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 $CUGFT$	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0
6 $FTA(a,b)$	1	0	0	1	0	0	0	1	0	1	0	0	1	1	1	0
7 $FTA(b,c)$	1	0	0	1	0	0	0	1	0	1	0	0	1	1	1	0
8 $FTA(c,a)$	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
9 $FTAHub(a)$	0	0	0	1	0	1	1	1	0	0	0	0	0	0	1	0
10 $FTAHub(b)$	1	0	0	1	0	0	0	1	1	0	1	1	1	1	1	0
11 $FTAHub(c)$	0	0	0	1	0	1	1	1	0	0	0	0	0	0	1	0
12 $FTAGFT$	0	1	1	1	0	1	1	1	1	0	1	0	0	0	1	0
13 $MTA(a,b)$	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0
14 $MTA(b,c)$	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0
15 $MTA(c,a)$	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
16 $MTAGFT$	0	1	1	1	0	1	1	0	0	0	0	0	0	0	1	0

First, take $x \in \{CU(i,b), FTA(i,b), MTA(i,b), FTAHub(i), FTAHub(b)\}$, with $i \in \{a,c\}$. Country b can destroy such trade agreements and, depending on the initial constellation, either $FTA(c,a)$ or the MFN regime remains. Then, further deviations are possible, namely $MTA(c,a)$ and $CU(c,a)$. However, each of the aforementioned trade agreements is indirectly dominated by $CU(c,a)$ and simultaneously country b is better off compared to the initial situation. Consequently, such deviations can not be avoided and no such x can be part of the stable set.

Now, consider $x \in \{MFN, FTA(c,a), MTA(c,a)\}$ for which $x \rightarrow_{\{a,c\}} CU(a,c)$ presents a deviation that can not be deterred. As in the previous paragraph, $CU(c,a)$ is not indirectly dominated any element and also $x \prec_{\{a,c\}} CU(a,c)$. Thus, no such x can be the part of the stable set as well.

At last, let $x \in \{CUGFT, FTAGFT, MTAGFT\}$ and consider the deviations where country b leaves the agreements. $CU(c,a)$, $FTA(c,a)$, or $MTA(c,a)$ can be the result. We have shown that the last two outcomes can not be stable. As for $CU(a,c)$, we have that for all x considered $x \prec_{\{b\}} CU(a,c)$. As a result, we conclude that no such x can be in the consistent set.

$CU(a,c)$ indirectly dominates each outcome, all deviations from it are deterred. So, the set containing $CU(a,c)$ is consistent and the largest one as well. □

The small country manages to block many desirable outcomes for large countries. Country b can unilaterally deviate from any trade agreement with higher welfare than CU(i,j) for the large countries. Thus, the majority of countries cannot impose their will on the other country. What the large countries can achieve is the best trade agreement that they can reach without the participation of the small country, in this case one among themselves.

A similar story unfolds in the scenario without Article XXIV Paragraph 5. There, the countries' preference rankings are as follows, with country b the small one and country a and c large (represented by i and j):

$$MTAGFT \prec_b MTA(i,b) \prec_b MFN \prec_b MTA(i,j)$$

$$MFN \prec_i MTA(i,b), MTA(j,b) \prec_i MTA(i,j) \prec_i MTAGFT$$

As the logic of the corresponding preference rankings of the countries is similar to before, let us directly present the proposition:

Proposition 8. *With the endowments given by $e_b = e_{min}$ and $e_a = e_c = e_{max}$, and under a modified institutional arrangement of the WTO, the stable constellation is the MTA between the two large countries, that is MTA(c,a).*

Proof. In this case, the indirect dominance matrix has the following form:

$$\begin{array}{l} \\ 1\,MFN \\ 2\,MTA(a,b) \\ 3\,MTA(b,c) \\ 4\,MTA(c,a) \\ 5\,MTAGFT \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

Assume, $x \in \{MTA(a,b), MTA(b,c), MTAGFT\}$ and consider the deviations, where country b dismantles any above mentioned constellation. Two possibilities: Either MFN or $MTA(c,a)$ remain. From MFN either no coalition moves away or, as it is indirectly dominated by $MTA(c,a)$ (see the indirect dominance matrix), the latter might be approached. In either case, b is better off. Thus, no such x can be part of the stable set.

Now, analyze the case of the MFN regime. Take the following deviation: $MFN \rightarrow_{\{a,c\}} MTA(c,a)$. As $MTA(c,a)$ is not indirectly dominated by any other trade agreement and $MFN \prec_{\{a,c\}} MTA(c,a)$, the MFN regime can not be stable as well.

As $MTA(c,a)$ indirectly dominates each trade agreement, all deviations from it are deterred. So, the set consisting of $MTA(c,a)$ is consistent and the largest one as well. □

Thus, similar to the other asymmetric case, one small and two large countries allow for partial but not full liberalization of world trade irrespective of the actual scenario (current vs. modified WTO rules). In terms of overall welfare, the world is better off in the hypothetical scenario without Article XXIV Paragraph 5 though. Individually, the small country is in a better position in case of MTA(i,j) compared to CU(i,j), as it exploits the MFN obligation of the large countries. By contrast, the large countries are better off in the other case. Therefore, while none of the two institutional arrangement facilitate global free trade, they influence the welfare for the stable set (both overall and individual).

4.3. **Asymmetric Case - Edges of the Triangle.** Let us now turn to the cases where the endowments of countries vary along one dimension - corresponding to the sides PQ, QR, and PR in the triangle of Figure 4. Specifically, Section 4.3.1 presents the scenario where the countries are small, small, and varying (PQ), Section 4.3.2 discusses the setting where the countries are small, large, and varying (QR), and Section 4.3.3 describes the case of a small country with two varying equally (PR).

While in the previous cases it was still possible to solve the problems analytically, the following require the use of a numerical approach. The analysis presented here consists of graphics picturing the composition of the stable sets and accompanying descriptions that explore the underlying mechanics. The exact numerical values for these (sub-)intervals can be found in Appendix C.2.

4.3.1. ***The case of one small, one large, and one varying country.*** First, let us consider the case $e_b = e_{min}$, $e_a = e_{max}$, and $e_c \in (e_{min}, e_{max})$ (side QR). Under the given pattern of the endowments, a number of trade agreements can be completely ruled out (with respect to the LCS). The MFN and GFT regimes for example are never part of the stable set. Additionally, none of the PTAs between the small and the large country appear as a stable outcome. The same holds for the hub structures where either the small or the large country is the hub node. As for the actual composition of the LCS, see Figure 5 for a graphical representation.

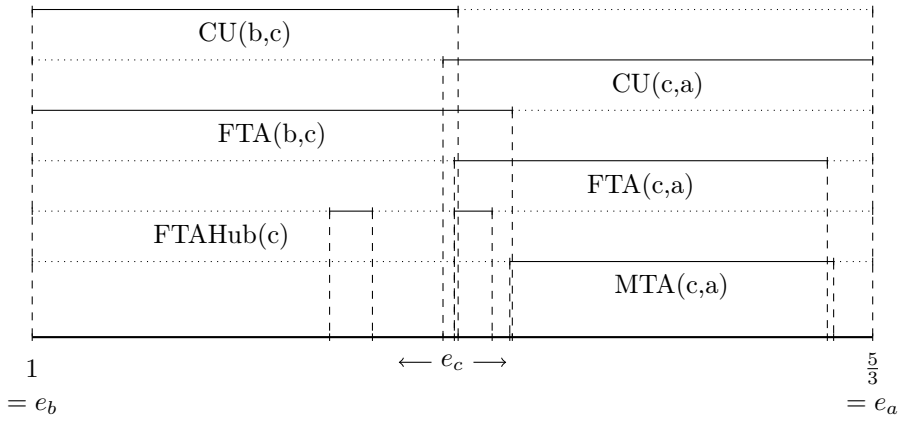


FIGURE 5. Characterization of the case of small, varying, and large country

The general observation is that when the varying country is close in size to the small country, then the PTAs between these smaller countries appear as elements in the stable set. When the country becomes larger the trade constellation between the larger countries replaces these. Additionally, there are two small, separated, regions in the middle of the interval where FTAHub(c) is stable.

In order to get an intuitive understanding of the results, let us identify specific trade agreements that go from stable to unstable (or the other way around) for certain endowment tuples. Then, explore the underlying mechanics to understand why the changes happen.

Start with the PTAs between country b and c, the small and the varying one. Interestingly, the only factor driving their stability are the preferences of country b (with fixed minimal endowments). Once the MFN regime becomes more desirable than CU(b,c) for country b, the constellation CU(b,c) drops from the stable set. Now, an identical story holds for the case of FTA(b,c). Thus, for both constellations it only requires a single change in the preference ranking of country b to influence the stable set.

The PTAs and MTAs between country a and c start to appear in the LCS when country c is becoming relatively large and closer to country a in size. At first both countries actually prefer to form a CU with country b, that is when country c is relatively small (and CU(b,c) actually is an element in the stable set). However, once it is preferable for country b to be the outsider instead of the insider in a CU, CU(c,a) emerges as a stable outcome (even though CU(b,c) still remains stable). Moreover, as soon as country c prefers FTA(c,a) respectively MTA(c,a) over the MFN regime, each of them becomes part of the LCS as well. For the interval where all PTAs and MTAs between country a and c are stable, both countries have fixed preference relations over these outcomes:

$$FTA(c,a) \prec_a CU(c,a) \prec_a MTA(c,a)$$
$$MTA(c,a) \prec_c FTA(c,a) \prec_c CU(c,a)$$

However, as soon as country c also prefers MTA(c,a) over FTA(c,a), the joint FTA drops out of the LCS. Similarly, as soon as country a prefers CU(c,a) over MTA(c,a), this also applies to the joint MTA - leaving CU(c,a) as the only stable outcome.

FTAHub(c) is stable in the two small, separated, regions in (or near) the middle of the interval. In the first region, the stability is driven by the fact that country b starts to value FTAHub(c) more than FTA(b,c) and gets in unison with country a in this respect. Once the preferences of country b over these outcomes get reversed, FTAHub(c) drops out of LCS again. In the second region, the stability of the same hub structure is largely determined by the change in the preferences of country c. Now, as soon as it starts to value FTA(c,a) over the MFN regime, which also puts FTA(c,a) in the LCS, both FTAs with c as a partner are stable and consequentially the corresponding hub structure is stable as well. As soon as the free-riding incentives of country b increase (valuing the MFN regime more than FTAHub(c)), this hub structure is not part of the stable set anymore.

The hypothetical institutional arrangement without Article XXIV Paragraph 5 does not promote the appearance of GFT as part of the stable set. GFTMTA, but also MTA(a,b) and MTA(b,c) never emerge as stable outcomes. Varying the size of country c generates either the MFN regime or MTA(c,a) as the stable element. Figure 6 presents these findings.[27]

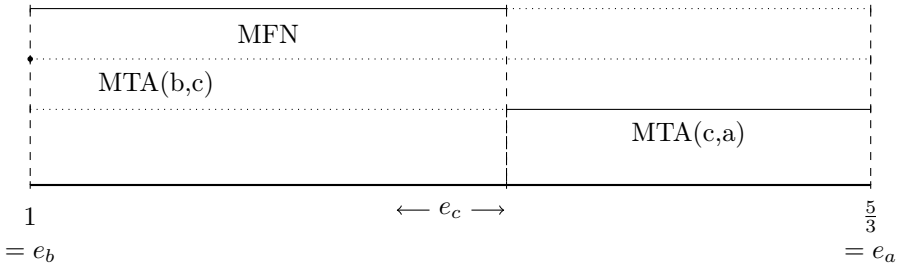


FIGURE 6. Characterization of the case of small, varying, and large country

[27]In addition to the aforementioned elements, it also pictures MTA(b,c) as a single point, see the dot, but this appears only for completion sake because that point corresponds to one of the extreme cases (point Q) discussed earlier.

Over the whole interval, country b does not have any incentive to form an MTA with any of the other countries. This is one reason why the MFN regime is stable over the specific range of the interval. The other reason is that country c prefers to not have a trade agreement with country a as long as its own size is not too large. Once country c gets sufficiently large though, MTA(c,a) presents a better option than the MFN regime. As a consequence, MTA(c,a) replaces the MFN regime as the stable set.

As a sidenote, while in the first scenario (with PTAs), the LCS near and at each respective extreme point corresponded to each other (continuity), the situation is different in the second scenario (without PTAs). When country c and b are equal in size, MTA(b,c) appears in the LCS even though it is not there before. Here, both the MFN regime and MTA(b,c) generate the same welfares for all countries (see also the discussion on point Q in Section 4.2.1).

Finally, under this given pattern of endowments, the GFT regime does not appear as part of the stable set independent of the scenario (with and without PTAs). However, the choice of rules does determine whether partial trade liberalization takes place or not. The possibility of forming PTAs reduces the incentive of the small(est) country to free ride. Otherwise, the MFN regime is the unique stable outcome when there is one small, one large, and one comparably small country.

4.3.2. ***The case of two small, and one varying country.*** Second, let us showcase the scenario with $e_b = e_c = e_{min}$ and $e_a \in (e_{min}, e_{max})$ (side PQ). In contrast to the previous case, it is not possible to rule out many of the trade agreements. Only MFN, FTAHub(a), and MTA(b,c) never appear in the LCS. The stable set is then presented in Figure 7.[28]

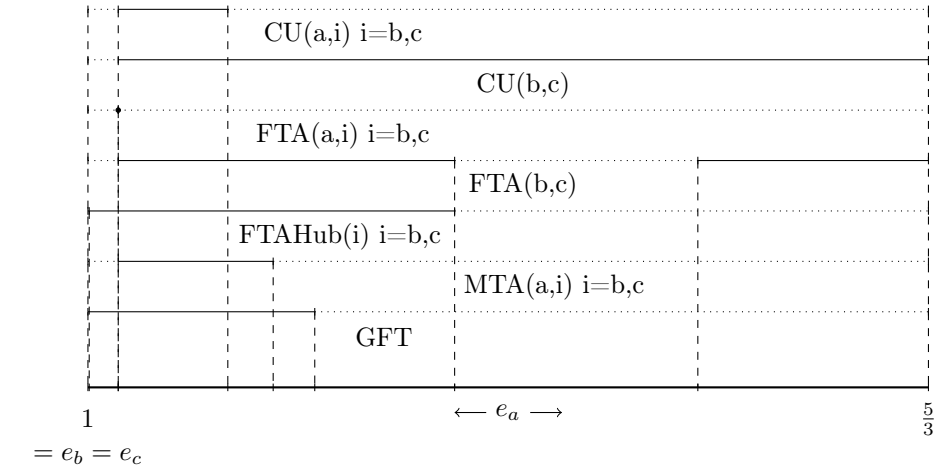


FIGURE 7. Characterization of the case of small, small, and varying country

In the immediate vicinity around symmetry, the GFT regime is the only element of the LCS (or rather the group of the three variants forms the stable set), but both FTAHub(b) and FTAHub(c) emerge as stable outcomes when moving away from

[28]The dot marks a single point again.

the extreme point. On the whole interval a number of different PTAs and MTAs, mostly between a small and the larger country, appear. Near the other point, only PTAs among the small countries are still stable.

First, the spike in the number of stable constellations close to symmetry actually follows a change in the preferences of the varying country with respect to CU(b,c) and the GFT regimes - it starts preferring the first over the latter. Furthermore, FTA(a,b) and FTA(c,a) become unstable because the small countries start to like the MFN regime more than the GFT variants (or rather these are only stable for that instance where it is not the case). When country a gets sufficiently large, country b prefers FTAHub(c) over CU(a,b) and c prefers FTAHub(b) over CU(c,a). As a consequence, both of these CUs drop out from the LCS. Similarly, when country b and c start preferring FTAHub(c) and FTAHub(b) over GFT, the latter stops being stable. A similar argument also applies to the MTAs. When the size of country a increases even more, both country b and c favor CU(b,c) over their respective hub structure, which results in FTA(b,c), FTAHub(b), and FTAHub(c) becoming unstable. When the endowment of country a gets close to maximum, the small countries are constrained by the MFN-tariffs and do not differentiate between CU(b,c) and FTA(b,c) anymore, which makes FTA(b,c) stable again.

In the scenario without Article XXIV Paragraph 5, the interval of GFT increases significantly. Moreover, over two-thirds of this interval the GFT regime is the unique element in the LCS. Additionally, all possible combinations of MTA appear at some point (mostly close to symmetry). Figure 8 demonstrates the results.[29]

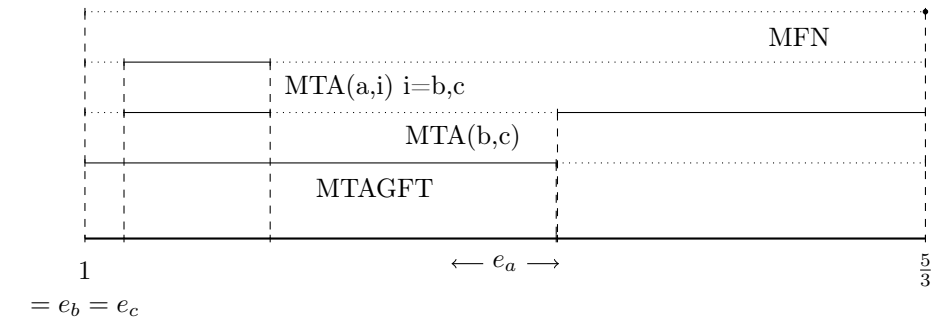


FIGURE 8. Characterization of the case of small, small, and varying country

Around symmetry, GFTMTA is the only element in the stable set. As soon as the small countries start to prefer MTAs with country a over GFTMTA, all three MTAs appear in the LCS. When the size of country a increases, the MTAs drop out from the LCS, because the small countries rank the one with the large country as the worst trade agreements (switching last place with the MFN regime), which actually also influences the stability of the MTA among themselves. Furthermore, the GFT regime becomes unstable when the small countries start to prefer their joint MTA over GFTMTA.

Similar to the previous case, the LCS changes at one extreme point. Namely, when the endowment of country a reaches the maximum, the MFN regime appears

[29]As before, in addition to the mentioned trade agreements, the graphic also contains MFN as a single point, see the dot, at an extreme point (again point Q).

in the LCS, as MFN and MTA(b,c) generate identical welfare for all countries (again, see also the discussion on point Q in Section 4.2.1).

Under this pattern of endowments, the first scenario does not allow for a sharp prediction via the LCS (unlike the previous case). Especially around symmetry, where almost all trade agreements are part of the stable set. In the second scenario, the effect of the PTAs on the stability of the GFT regime is significant though - essentially the abolishment of Article XXIV Paragraph 5 would facilitate the formation of GFT as long as there are two small countries and the third country is not substantially larger.

4.3.3. ***The case of one small, and two varying countries***. Finally, let us turn to the case where $e_b = e_{min}$ and $e_a = e_c \in (e_{min}, e_{max})$ (side PR). In this scenario, depending on the size of the larger countries, any trade agreement can be part of the stable set. The exact composition of the LCS is the basis for Figure 9.

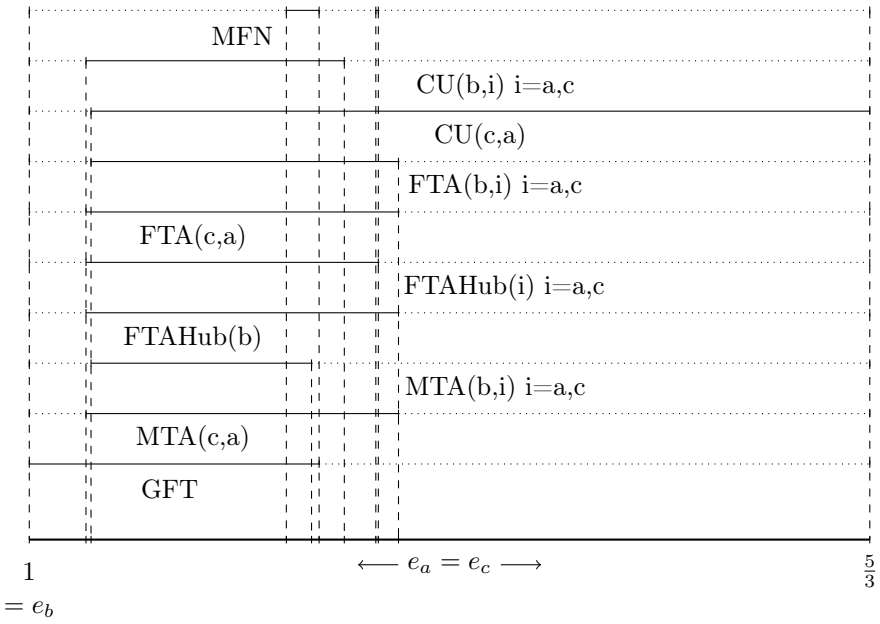


FIGURE 9. Characterization of the case of small, varying, and varying country

In the interval around symmetry, the GFT variants are stable and stay the unique elements of the stable set for longer (compared to the previous case). Also, a collection of different trade agreements is stable relatively close to symmetry. However, near the other extreme point, the CUs between the varying countries is the unique stable outcome. Also, MFN is stable in two small, separated, regions.

Again, the peak in stability near symmetry comes from a shift in the preferences of the varying countries with respect to CUs. At that point, both of these start to prefer a CU with the small country over the different forms of GFT. The occurrence of the MFN regime actually follows a preference of the small country of MFN over GFT (the first region) and then FTAHub(a) and FTAHub(c) (the second region). As countries a and c are getting bigger, first MTA(a,b) and MTA(b,c) drop out

from the LCS when they rank the lowest according to the preferences of country b. The three variants of GFT become unstable once the small country prefers CU(c,a). Next, CU(a,b) and CU(b,c) follow as the small country starts to prefer to be in the MFN regime over a CU with any of the larger countries. As soon as CU(c,a) becomes the more desirable trade agreement for country b when comparing it to any FTA where b participates or any hub structure with a large country as hub, all aforementioned constellations drop out from the LCS.

Contrary to the previous case, switching off Article XXIV Paragraph 5, actually decreases the interval where the GFT regime is part of the stable set. However, this effect is considerably smaller. A similar observation holds for the range where the GFT regime is the unique stable outcome. The exact composition can be seen in Figure 10.

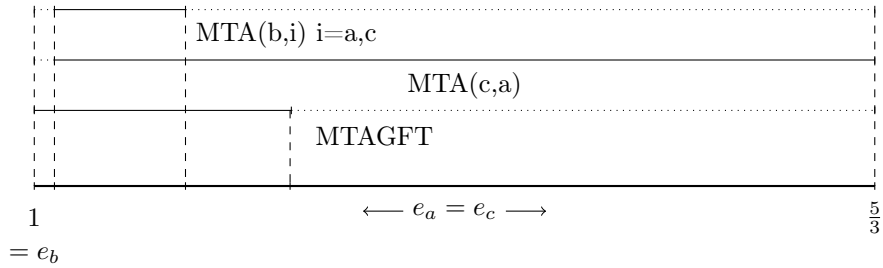


FIGURE 10. Characterization of the case of small, varying, and varying country

The main driving force behind the stability are alterations in the preferences of the small country over the interval. More precisely, it is important where exactly the small country places the MFN regime in its ranking of preferences compared to the other trade agreements. As soon as country b prefers MFN over another constellation, the latter drops out from the LCS. After a certain point, the MTA between the large countries remains the only stable outcome.

Similar to the previous pattern of endowments, this case makes a clear analysis in the first scenario difficult, especially around symmetry where, as before, almost all constellations are stable. The effect of Article XXIV Paragraph 5 actually works in the other direction on the stability of the GFT regime when compared to the previous case though.

4.4. **Asymmetric Case - Interior of the Triangle.** In the following (and final) part of the analysis, the focus lies on the interior of the triangle of Figure 4. Here, unlike in the previous discussions, both CU and FTA appear together under the label of PTA. However, a variation of the graphics of this analysis that actually distinguishes between the two can be found in Appendix C.1. For the purpose of a general overview, this level of abstraction suffices though - in fact, the members of a specific trade agreement are suppressed for clarity as well, i.e. who is insider and outsider.

First, we consider the existing institutional set-up, where PTAs are available to the countries. Figure 11 shows the (simplified) stable sets. In a small region close to symmetry, region one, the trinity of GFT regimes is the unique stable element. In both a neighboring and another distant area, region two, PTAs become

stable as well. The connecting area, region three, adds MTAs as another stable element. In a tiny area near the diagonal, region four, no form of trade agreement can actually be excluded from the stable set. Further along the diagonal and in the asymmetric corners, region five, PTAs are the only stable trade constellation. In between, region six, MTAs are also stable. In another tiny area, also close to the diagonal, region seven, MFN enters the stable set as well. In general, with a certain degree of asymmetry among countries at most partial trade liberalization can be expected.

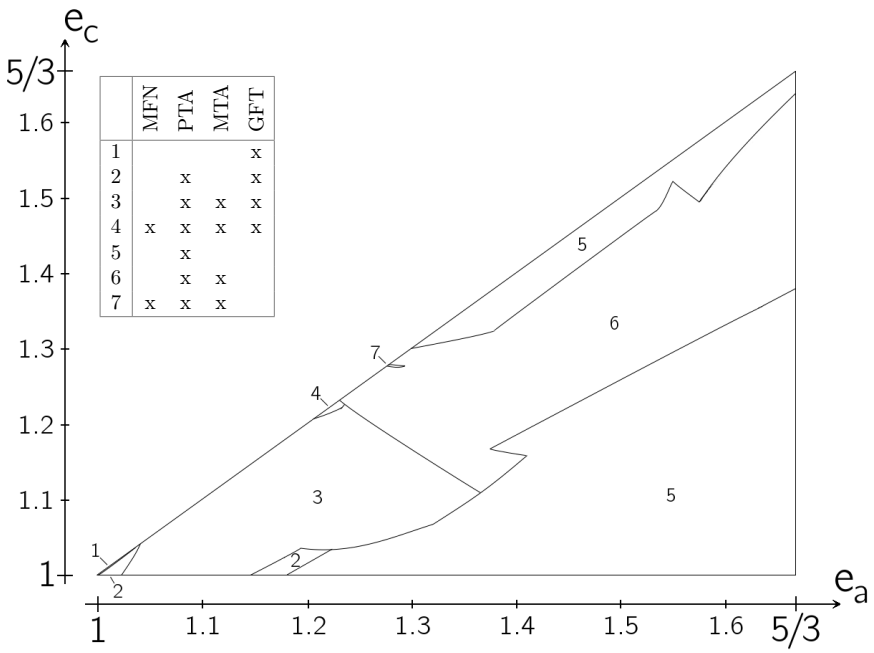


FIGURE 11. Simplified Overall Stability with PTAs

Next, we consider a modified institutional set-up, where PTAs are not available. Figure 12 depicts the corresponding stable sets. In an area near symmetry as well as in a sizeable area away from it, region one, GFT is again the unique stable element. Connected to these are two areas, region two, where MTAs become stable as well. Moving towards the asymmetric corners, region three, yields MTAs as the only stable element. In between, region four, only MFN remains in the stable set.

The comparison of the graphics allows us to deduce two compelling statements. The first noteworthy result is the extent of MFN in each scenario. In the modified institutional arrangement without PTAs the area where MFN is (uniquely) stable increases substantially (note that this effect is present away from symmetry). Under (significant) asymmetry, it seems that PTAs allow countries to move towards their international efficiency frontier (cf. Bagwell et al. (2016)).

The second interesting result is the difference in the extent of stability of GFT in the two regulatory scenarios. First, recall that once the degree of asymmetry surpasses a certain threshold, none of the GFT regimes remains in the stable set, independent of the institutional set-up. Around symmetry the opposite holds in

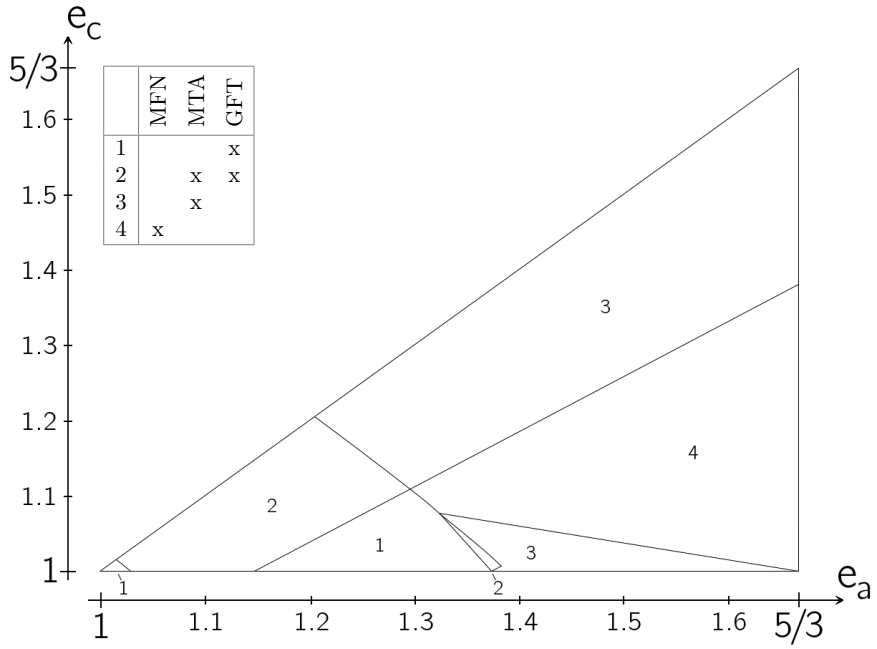


FIGURE 12. Simplified Overall Stability without PTAs

that the GFT regimes are always stable there (in both scenarios). In between, the effect of PTAs on the stability of GFT depends on the structure of asymmetry. See Figure 13 for the different areas of stability depending on the regulatory scenario. Note that region one corresponds to the aforementioned stability around symmetry. In the case of two relatively larger countries (but not too large), the abolishment of PTAs results in a reduction of the area where GFT is stable, see region two. In this instance, PTAs act as 'building blocks' on the road to GFT. But when two countries are relatively smaller (but not too small), the same regulatory action yields the exact opposite effect, see region three. Here, PTAs are 'stumbling blocks'. Thus, whether PTAs are 'building blocks' or 'stumbling blocks' in the vicinity of symmetry depends on the relative size of the majority of the countries.

In a nutshell: If the world is in the vicinity of symmetry and two out of three countries are close to identical while relatively smaller than the other one, the area where the GFT regime is stable increases when prohibiting PTAs. However, when two similar countries are relatively larger, the availability of PTAs is conducive to the stability of the GFT regime. Finally, if the world is further away from symmetry, full trade liberalization is not attainable at all and an area where the MFN regime is stable appears in the scenario without PTAs.

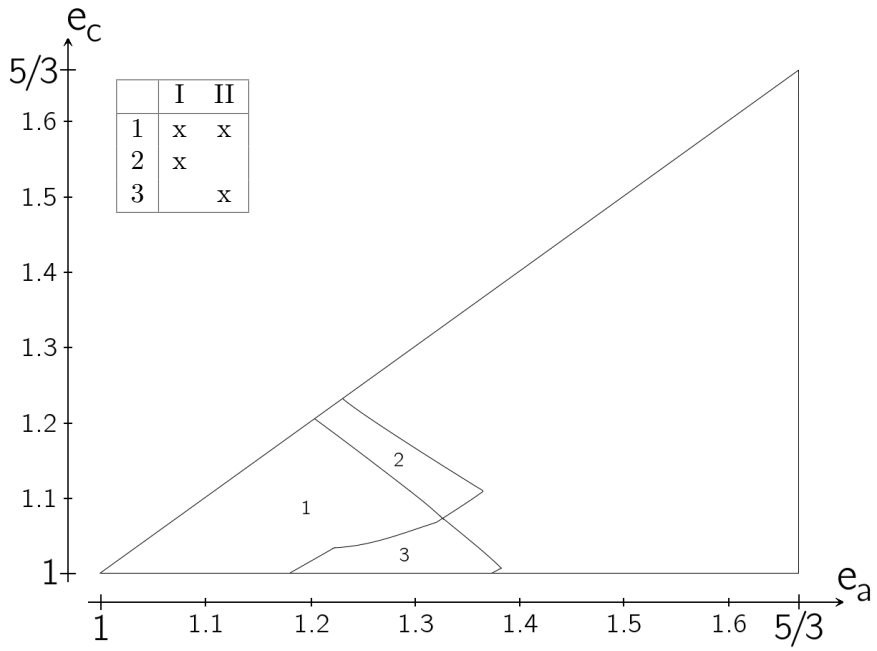


FIGURE 13. The different areas of stability of the GFT regime in the scenario with (I) and without (II) PTAs

5. Discussion

In this section, let us first compare the findings of our paper with those of several similar studies and underline the differences in the modeling strategies, especially with respect to the explanatory power of each approach. Second, this section links our predictions to different empirical observations, thereby validating our approach.

Let us start with the paper of Saggi, Woodland and Yildiz (2013).[30] First note that the underlying trade model in our paper is similar to theirs, which allows a direct comparison of the findings in certain scenarios (found in the next paragraph). The first distinction is the set of trade agreements under consideration. While in our model countries can be involved in multilateral trade liberalization via MTAs or they may choose to carry out their favored form of preferential trade liberalization through CUs or FTAs, Saggi, Woodland and Yildiz (2013) focuses on two out of these three possibilities, namely CUs and MTAs. In our opinion, the expanded set of trade arrangements in our model allows us to fully capture the trade-offs among the alternatives and make the model realistic. The second significant difference is the concept of stability. While our framework uses the notion of LCS, the paper of Saggi, Woodland and Yildiz (2013) utilizes Coalition-Proof Nash Equilibria. As Bernheim, Peleg and Whinston (1987) note, their notion of self-enforceability, which is critical for Coalition-Proof Nash Equilibria, is too restrictive in one crucial aspect, mainly:

[30]As mentioned in Section 2, this paper analyzes the case of CUs and MTA while their other papers (Saggi and Yildiz (2010) resp. Missions, Saggi and Yildiz (2016)) focus on different combinations of trade agreements (FTAs and MTAs resp. CUs and FTAs) but use a similar framework. As a consequence, the comparison of methodology applies to these papers as well.

'When a deviation occurs, only members of the deviating coalition may contemplate deviations from the deviation. This rules out the possibility that some member of the deviating coalition might form a pact to deviate further with someone not included in this coalition.' Importantly, this limitation does not affect the concept of LCS. It is not a pure academic difference. The historic development of the two (disjoint) trade constellations in Europe in the 1960s, the European Economic Community (EEC) and the European Free Trade Association (EFTA), can not be captured by a model using Coalition-Proof Nash Equilibria, because it excludes those strategies that the UK actually followed during that time.[31] While being a member of EFTA the UK applied for EEC membership in 1961 and thereby undermined the stability of the EFTA. Furthermore, 'the more ambitious Kennedy Round between 1964 and 1967 coincided with negotiations to expand the EEC to include Britain, Ireland, Denmark, Greece, and Norway - and was motivated in part by US concerns about being excluded from an ever-broader and more unified European market.' (World Trade Report 2011). Thus, unlike the Coalition-Proof Nash Equilibrium, the LCS allows interactions among members and non-members of coalitions simultaneously, thereby accommodating these historic developments. Additionally, the (conjectured) motivation of the US reinforces the importance of the interaction among different modes and forms of trade liberalization.

In specific cases it is actually possible to directly compare the composition of the stable sets of our paper with those of Saggi, Woodland and Yildiz (2013). In fact, their 'multilateralism game' fits our scenario of a modified institutional arrangement without PTAs. Compare Figures 2 and 5 of Saggi, Woodland, and Yildiz (2013) with Figures 10 and 8 in our paper correspondingly. In the case with one small country and the other two varying, both approaches predict the same stable sets near the endpoints of the interval. In our paper GFT stays part of the stable set even when MTAs, either between the large countries or a small and a large country, become stable as well - which is in contrast to Saggi, Woodland, and Yildiz (2013). A similar observation follows for the case of two small and one varying country, i.e. near the endpoints of the interval results coincide while the appearance of MTAs does not prevent GFT from staying in the stable set. Furthermore, in our model there exists an interval where GFT becomes the unique stable element once more. It seems that one effect of the unlimited farsightedness is the proliferation of GFT.

Another relevant paper is that of Lake (2017). Apart from the stability concept, the approach of Lake also differs with respect to the choice set. There, the focus lies on FTAs. In this respect, a direct comparison of the findings is difficult. Moreover, compared to the previous paper, the 'multilateralism game' is further simplified in Lake (2017), as the only possible regime there is the three country constellation that results in GFT. Furthermore, as the underlying trade model, the paper employs the political economy oligopolistic model. However, according to the paper, the findings are robust with respect to various underlying trade models, including the competition via exports model. Additionally, Lake himself compares his results to those of Saggi and Yildiz (2010). Due to the similarity of the 'multilateralism game' in Saggi and Yildiz (2010) and Saggi, Woodland, and Yildiz (2013), it is only logical to compare our results with those of Lake (2017) as with the previous paper.

According to Lake (2017), specifically Figure 3, the exact role of FTAs under asymmetry depends on the nature of asymmetry (similar to our findings). However,

[31] See Baldwin and Gylfason (1995).

the direction of the effect of PTAs on trade liberalization is the opposite. There, in case of two larger and one small country, FTAs act as '(strong) stumbling blocs', and with two smaller and one larger country, as '(strong) building blocs'. Furthermore, there it seems that the determining factor are the preferences of the larger countries, while in our case the findings are driven by the preferences of the smaller countries. Thus, the aforementioned differences in the choice set and stability concept appear to shift the power to influence the negotiations among the countries, which then produces a different outcome. Specifically, in the case of the 'multilateralism game', which corresponds to our scenario without PTA, there are essentially two areas, that is one where GFT emerges as unique equilibrium and one with MFN instead. In the parameter space triangle the first makes up the upper left part of the triangle, while the second makes up the opposing lower right.[32] Therefore, for two larger countries the GFT regime remains the unique equilibrium for the whole interval, whereas in our model it only stays stable in the vicinity of symmetry (only partially unique) and then the MTA between the larger countries is the unique stable outcome, as seen in Figure 10. Furthermore, for two smaller countries first GFT then MFN is the unique equilibrium, while in the beginning GFT is also stable in our case (although only partially unique) the MTA between the small countries takes it place as the unique stable element near the end, depicted in Figure 8. Finally, in the case of three different countries, it starts with MFN and ends with GFT as the unique equilibrium, which corresponds to our findings for the first part but then in the second part the MTA between the medium and large country is the unique stable element, visible in Figure 6.

Another aspect of Lake (2017) necessitates a remark, namely the assumption that a once created trade agreement remains binding from then on. Lake argues that 'the binding nature of trade agreements is pervasive in the literature and realistic'. However, the latest developments in the world cast doubt on this plausibility. The USA, for example, pulled out of the negotiations for the Trans-Pacific Partnership at the final stage and currently negotiates with South Korea to amend the so-called KORUS FTA. The developments around 'Brexit' are another argument for modeling non-binding trade agreements. Using the LCS as stability concept allowed us to accommodate such deviations.

A final remark on the relation of our research with empirical observations. As the analysis has shown, a growing degree of asymmetry among countries produced a significant area of stability for the MFN regime when PTAs are prohibited, see region four in Figure 12. If one would interpret the expansion of the WTO rule set to an increasing number of countries as an amplification of asymmetry among its member states, then the potential of PTAs to prevent the MFN regime might be one of the driving factors of the prevalence of PTAs in recent history. However, the World Trade Report (2011) casts doubt on this motive. According to the report: 'Approximately 66 per cent of tariff lines with MFN rates above 15 percentage points have not been reduced in PTAs.' Note that a reduction of the tariff peaks for 34 percent of the tariff lines is still a significant effect considering the fact that the majority of tariff peaks occurs in agricultural and labor-intensive manufacturing sectors, which are politically sensitive and countries usually try to exclude them from the trade liberalization via PTAs. Furthermore, according to the same report,

[32]The first corresponds to the regions denoted WBB and SSB in Figure 3 of Lake (2017), while the second matches the regions SBB and WSB.

over time more and more PTAs have included provisions regarding technical barriers to trade - a category of the Non-Tariff Barriers (NTBs). The paper of Kee, Nicita, and Olarreaga (2009) estimates that restrictiveness measures that include NTBs are on average about 87 percent higher than the measures based on tariffs alone. Thus, in order to evaluate the aforementioned motive properly, the effects of NTBs should be included in the analysis as well. Moreover, contrary to the conclusion of the report that 'preference margins are small and market access is unlikely in many cases to be an important reason for creating new PTAs', Keck and Lendle (2012) show that the preferential utilization rates are often high even in the case of small preference margins and they increase both with the preference margin and the export volume. All in all, it is our opinion that there is (partial) evidence for the motive of avoidance of MFN via PTAs which coincides with the predictions of our model about the trade-off between trade liberalization and the MFN regime in the asymmetric part of the parameter space.

As this overview showed, a number of core attributes of the LCS, specifically the farsightedness, the non-binding nature of agreements, and the possibility of interactions between members and non-members of coalitions, capture important mechanisms present in the world and influence the composition of the stable set significantly (when compared to other stability concepts).

6. Conclusion

Under the rules of the WTO (previously GATT), a group of countries can engage in both multilateral and preferential trade liberalization. The formation of global trade agreements is a complex game and the rules of the game influence the nature of the exact outcomes. WTO's Article I aims at creating the global free trade system, while Article XXIV Paragraph 5 allows countries to seemingly circumvent the liberalization process. In this paper, our focus lies on the stability of trade policy arrangements under two different regulatory scenarios (with and without PTAs) assuming unlimited farsightedness of the participants in the trade negotiations and considering an extensive set of trade agreements - moving our model closer to reality.

Unfortunately, the answer to the question whether PTAs are 'building blocks' or 'stumbling blocks' on the path towards global free trade is not as straightforward as one would like it to be. In the end, the results presented here are mixed and depend on the size distribution of the countries. Under symmetry, GFT is the unique stable trade constellation in both regulatory scenarios. But as soon as one moves away from symmetry, GFT might not be reached at all. In between, the effect of switching off Article XXIV Paragraph 5 depends on the exact asymmetry. In case two countries are relatively smaller, prohibiting PTAs increases the area of stability of the GFT regimes. When two countries are relatively larger, it reduces the area. Once the world is further away from symmetry, abolishing the exception for PTAs might result in the worst possible state from the perspective of overall world welfare, the non-cooperative MFN regime. Therefore, under such circumstances, PTAs act as a mechanism that prevents the MFN regime.

Our research also raises a couple of questions in need of further investigation. First, it would be interesting to study the robustness of the findings with respect to the underlying trade model. While the model of competition via exports remains popular in the related literature, economists also extensively use both oligopoly and competition via imports model. Fortunately, the framework presented here

does allow for a different underlying trade model such as the ones mentioned above. Another potential area of inquiry might be an extension of the framework to increase the number of countries. Nowadays, in addition to bilateral negotiations, so-called plurilateral negotiations play an important role in the development of preferential trade liberalization. Recent examples are the Trans-Pacific Partnership (TTP) and the Regional Comprehensive Economic Partnership (RCEP). Including more than three countries in a model would allow us to investigate the strategic interactions among countries whilst taking these negotiations into account. The introduction of political economy considerations to the underlying trade model is another area of interest[33], as it might allow us to understand the nature of tariff peaks occurring after PTAs come into effect. It is our opinion that modifications or extensions of our framework (as mentioned here) are directions worthy of further research.

As a final remark, it is perhaps important, going forward, to move the debate of 'building blocks' vs. 'stumbling blocks' to a level of detail that goes beyond this binary choice.

[33]See for example Facchini et al. (2013).

Appendix A. Pseudocode

Note, that a couple of functions and variables are directly baked into the program without any further explanation in the pseudocode below - for example the matrix that determines the general network structure (for each player and all coalitions). The origin and characterization of these can be found in their respective parts in the main paper. The network structure A and the preference relations B both enter as a collection of $|X| \times |X|$-matrices, $\{A_S\}_{S\subseteq N, S\neq\emptyset}$ and $\{B_S\}_{S\subseteq N, S\neq\emptyset}$ resp., where $(A_S)_{i,j} = \mathbb{1}_{\{i\to_S j\}}(i,j)$ and $(B_S)_{i,j} = \mathbb{1}_{\{i\prec_S j\}}(i,j)$ for $(i,j) \in X \times X$.

Algorithm Largest Consistent Set

Input: Countries N, Outcomes X, Network Structure A, Preference Relations B
Output: Largest Consistent Set $\{Y\}$

procedure ParameterSpaceLCS(N, X, A, B)
  $E = eMaxArea$ ▷ See Section 3.6
  $\alpha = \alpha MinValue(E)$ ▷ See Section 3.6
  for $e \in E$ **do**
    $Y = GeneralLCS(N, X, A, B)$
  return $\{Y\}$

function GeneralLCS(N, X, A, B)
  for $S \subseteq N$ **do**
    $C_S = \min\{A_S, B_S\}$
  $D^0 = \max_{S\subseteq N}\{C_S\}$ ▷ : Direct Dominance
  $n = 0$
  repeat
    $n = n + 1$
    for $S \subseteq N$ **do**
      $A_S^n = (\mathbb{1}_{\{(A_S \cdot D^{n-1})_{i,j}\neq 0\}}(i,j))_{(i,j)\in X\times X}$
      $D_S^n = \min\{A_S^n, B_S\}$
    $D^n = \max_{S\subseteq N}\{D_S^n\}$ ▷ : Indirect Dominance
  until $D^n = D^{n-1}$
  $D = \mathbb{1}_X + D^n$
  $Y^0 = (1)_{x\in X}$
  $m = 0$
  repeat
    $m = m + 1$
    for $x \in X$ **do**
      if $Y_x^{m-1} = 0$ **then**
        $Y_x^m = 0$
      else
        $$y = \max_{k\in X, S\subseteq N}\left\{(A_S)_{x,k}\left(1 - \max_{z\in X}\left\{Y_z^{m-1}(D)_{k,z}\left(1 - (B_S)_{x,z}\right)\right\}\right)\right\}$$
        $Y_x^m = Y_x^{m-1} - y$
  until $Y^m = Y^{m-1}$
  $Y = Y^m$
  return Y

Appendix B. Model

B.1. **Individual Welfare.** The following table lists the individual welfare for each (representative) trade agreement, depending on endowments and tariffs, multiplied with the factor 18. Note that for MFN, CUGFT, FTAGFT, and MTAGFT the welfare W_i resembles W_j and W_k. In case of CU(i,j), FTA(i,j), and MTA(i,j) the welfare W_i is similar to W_j. For FTAHub(i) the welfare W_j resembles W_k.

Trade Agreement	Individual Welfare
MFN	
↳ W_i	$-10e_i^2 + 2e_j^2 + 2e_k^2 - 8t_i^2 + t_j^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k - t_j - t_k)$ $+2e_j(e_k + t_i + t_k) + 2e_k(t_i + t_j)$
CU(i,j)	
↳ W_i	$-10e_i^2 + 2e_j^2 + 2e_k^2 - 11t_{ik}^2 + t_{jk}^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k + t_{jk} - t_k)$ $+2e_j(e_k - 4t_{ik} + t_k) + 2e_k(5t_{ik} - t_{jk})$
↳ W_k	$2e_i^2 + 2e_j^2 - 10e_k^2 + 4t_{ik}^2 + 4t_{jk}^2 - 8t_k^2 + 2e_i(e_j - 2e_k + 2t_{jk} + t_k)$ $+2e_j(-2e_k + 2t_{ik} + t_k) + 4e_k(9\alpha - 2t_{ik} - 2t_{jk})$
CUGFT	
↳ W_i	$-10e_i^2 + 2e_j^2 + 2e_k^2 + 4e_i(9\alpha - e_j - e_k) + 2e_je_k$
FTA(i,j)	
↳ W_i	$-10e_i^2 + 2e_j^2 + 2e_k^2 - 11t_{ik}^2 + t_{jk}^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k + t_{jk} - t_k)$ $+2e_j(e_k - 4t_{ik} + t_k) + 2e_k(5t_{ik} - t_{jk})$
↳ W_k	$2e_i^2 + 2e_j^2 - 10e_k^2 + 4t_{ik}^2 + 4t_{jk}^2 - 8t_k 2 + 2e_i(e_j - 2e_k + 2t_{jk} + t_k)$ $+2e_j(-2e_k + 2t_{ik} + t_k) + 4e_k(9\alpha - 2t_{ik} - 2t_{jk})$
FTAHub(i)	
↳ W_i	$-10e_i^2 + 2e_j^2 + 2e_k^2 + t_{jk}^2 + t_{kj}^2 + 4e_i(9\alpha - e_j - e_k + t_{jk} + t_{kj})$ $+2e_j(e_k - t_{kj}) - 2e_kt_{jk}$
↳ W_j	$2e_i^2 + 2e_j^2 - 10e_k^2 + 4t_{jk}^2 - 11t_{kj}^2 + 2e_i(e_j - 2e_k + 2t_{jk} - 4t_{kj})$ $+e_j(-4e_k + 10t_{kj}) + 4e_k(9\alpha - 2t_{jk})$
FTAGFT	
↳ W_i	$-10e_i^2 + 2e_j^2 + 2e_k^2 + 4e_i(9\alpha - e_j - e_k) + 2e_je_k$
MTA(i,j)	
↳ W_i	$-10e_i^2 + 2e_j^2 + 2e_k^2 - 8t_i^2 + t_j^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k - t_j - t_k)$ $+2e_j(e_k + t_i + t_k) + 2e_k(t_i + t_j)$
↳ W_k	$2e_i^2 + 2e_j^2 - 10e_k^2 + t_i^2 + t_j^2 - 8t_k^2 + 2e_i(e_j - 2e_k + t_j + t_k)$ $+2e_j(-2e_k + t_i + t_k) + 4e_k(9\alpha - t_i - t_j)$
MTAGFT	
↳ W_i	$-10e_i^2 + 2e_j^2 + 2e_k^2 + 4e_i(9\alpha - e_j - e_k) + 2e_je_k$

Table 14. The individual welfare for each trade agreement depending on endowments and tariffs

B.2. **Tariffs.** The following describes the tariffs that the countries choose for each trade agreement. In addition to the specific restrictions mentioned in Section 3.3, all tariffs are bounded both from below and above by zero and the MFN-tariff respectively. As per WTO rule, the formation of any PTA does not allow additional tariffs towards others - which results in the upper bound of the MFN-tariff. Also, any form of subsidies is excluded here - which results in the lower bound of zero. Now, the following determines and describes the optimal tariffs for each scenario and the cases where capping occurs:

B.2.1. *MFN.* In this case, the optimal tariff of country i, given by $t_i^* = \frac{1}{8}(e_j + e_k)$, is always greater than zero as the endowments themselves are greater than zero. Additionally, t_i^* is going to play the role of the maximal tariff for country i for all the other agreements, then denoted t_i^{MFN}.

B.2.2. *CU.* Consider the scenario CU(i,j), then the optimal tariff of country i towards country k, given by $t_{ik}^* = \frac{1}{5}(2e_k - e_j)$, is always greater than zero but not always less than the MFN-tariff (and the one towards country j, t_{ij}^*, is always zero):
i) Lower Bound. By assumption on the endowments $e_k \geq \frac{3}{5}e_j$ and thus $e_k > \frac{1}{2}e_j$, which guarantees $t_{ik}^* > 0$.
ii) Upper Bound. By assumption on the endowments $e_k \leq \frac{5}{3}e_j$ however $t_{ik}^* \leq t_i^{MFN}$ requires $e_k \leq \frac{13}{11}e_j$, which leaves the interval $\frac{13}{11}e_j < e_k \leq \frac{5}{3}e_j$ to require capping. For this interval, the (maximal) MFN-tariff is optimal as the derivative of the joint welfare with respect to t_{ik} is always greater than zero on the interval $[0, t_i^{MFN}]$:

$$\frac{\partial(W_i + W_j)}{t_{ik}} = \frac{1}{9}\left(-10t_{ik} - 2e_j + 4e_k\right) \geq \frac{1}{36}\left(-13e_j + 11e_k\right) > 0$$

B.2.3. *FTA.* Consider the scenario FTA(i,j), then the optimal tariff of country i towards country k, given by $t_{ik}^* = \frac{1}{11}(5e_k - 4e_j)$, is neither always greater than zero nor always less than the MFN-tariff (but the one towards country j, t_{ij}^*, is zero):
i) Lower Bound. By assumption on the endowments $e_k \geq \frac{3}{5}e_j$ however $t_{ik}^* \geq 0$ requires $e_k \geq \frac{4}{5}e_j$, which leaves the interval $\frac{3}{5}e_j \leq e_k < \frac{4}{5}e_j$ to require capping. For this interval, the (minmal) zero-tariff is optimal as the derivative of the welfare with respect to t_{ik} is always lesser than zero on the interval $[0, t_i^{MFN}]$:

$$\frac{\partial W_i}{\partial t_{ik}} = \frac{1}{9}\left(-11t_{ik} - 4e_j + 5e_k\right) \leq \frac{1}{9}\left(5e_k - 4e_j\right) < 0$$

ii) Upper Bound. By assumption on the endowments $e_k \leq \frac{5}{3}e_j$ however $t_{ik}^* \leq t_i^{MFN}$ requires $e_k \leq \frac{43}{29}e_j$, which leaves the interval $\frac{43}{29}e_j < e_k \leq \frac{5}{3}e_j$ to require capping. For this interval, the (maximal) MFN-tariff is optimal as the derivative of the welfare with respect to t_{ik} is always greater than zero on the interval $[0, t_i^{MFN}]$:

$$\frac{\partial W_i}{\partial t_{ik}} = \frac{1}{9}\left(-11t_{ik} - 4e_j + 5e_k\right) \geq \frac{1}{72}\left(-43e_j + 29e_k\right) > 0$$

B.2.4. *MTA.* Consider the scenario MTA(i,j), then the optimal tariff of country i, given by $t_i^* = \frac{1}{7}(2e_k - e_j)$, is greater than zero and less or equal to the MFN-tariff as per assumption on the endowments $\frac{3}{5}e_j \leq e_k \leq \frac{5}{3}e_j$.

B.2.5. *Notes.* The analysis considered country i and an agreement with country j, but it naturally extends to all other combinations. Also, the perspective of the third country needs no further analysis as it always chooses the MFN-tariff. Furthermore, the case of FTAHub(i) is simply a combination of FTA(i,j) and FTA(i,k). Finally, the three variants of GFT require no additional analysis as every country always chooses the zero-tariff. Information on another form of GFT, Pseudo-GFT, that technically exists but turns out to be negligible, can be found in Appendix B.5.1.

B.3. **Overall Welfare.** The following table lists the overall welfare for each (representative) trade agreement, depending purely on endowments, computed modulo $2\alpha\left(\sum_{n\in N} e_n\right)$, which is the common term associated with the factor α. Also, the notation l_c and l^c is used to indicate that country l is capped in terms of tariffs from below or above respectively. Note that one specific comparison of trade agreements is presented in more detail in Appendix B.5.2.

Trade Agreement	Overall Welfare
MFN	
↳ no cap	$\frac{11}{32}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$
CU(i,j)	
↳ no cap	$\frac{1}{1600}(-563e_i^2 - 550e_ie_j - 448e_ie_k - 563e_j^2 - 448e_je_k - 704e_k^2)$
↳ i^c	$\frac{1}{1600}(-563e_i^2 - 550e_ie_j - 448e_ie_k - 550e_j^2 - 550e_je_k - 627e_k^2)$
↳ i^c, j^c	$\frac{11}{32}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$
CUGFT	
↳ no cap	$\frac{1}{3}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$
FTA(i,j)	
↳ no cap	$\frac{1}{7744}(-2963e_i^2 - 2662e_ie_j - 1728e_ie_k - 2963e_j^2 - 1728e_je_k - 3648e_k^2)$
↳ i_c	$\frac{1}{23232}(-8889e_i^2 - 7986e_ie_j - 5184e_ie_k - 7865e_j^2 - 7744e_je_k - 9344e_k^2)$
↳ i_c, i_c	$\frac{1}{192}(-65e_i^2 - 66e_ie_j - 64e_ie_k - 65e_j^2 - 64e_je_k - 64e_k^2)$
↳ i^c	$\frac{1}{7744}(-2963e_i^2 - 2662e_ie_j - 1728e_ie_k - 2662e_j^2 - 2662e_je_k - 3155e_k^2)$
↳ i^c, j^c	$\frac{11}{32}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$
FTAHub(i)	
↳ no cap	$\frac{1}{363}(-153e_i^2 - 81e_ie_j - 81e_ie_k - 146e_j^2 - 121e_je_k - 146e_k^2)$
↳ j_c	$\frac{1}{363}(-137e_i^2 - 81e_ie_j - 121e_ie_k - 146e_j^2 - 121e_je_k - 121e_k^2)$
↳ j_c, k_c	$\frac{1}{3}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$
↳ j^c	$\frac{1}{23232}(-8889e_i^2 - 5184e_ie_j - 7986e_ie_k - 9344e_j^2 - 7744e_je_k - 7865e_k^2)$
↳ j^c, k^c	$\frac{1}{192}(-66e_i^2 - 66e_ie_j - 66e_ie_k - 65e_j^2 - 64e_je_k - 65e_k^2)$
FTAGFT	
↳ no cap	$\frac{1}{3}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$
MTA(i,j)	
↳ no cap	$\frac{1}{3136}(-1083e_i^2 - 1078e_ie_j - 960e_ie_k - 1083e_j^2 - 960e_je_k - 1216e_k^2)$
MTAGFT	
↳ no cap	$\frac{1}{3}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$

TABLE 15. The overall welfare for each trade agreement depending on endowments

B.4. **Transition Tables.** The following lists the network structure of Section 3.5, specifically Figure 1 and 2, in the form of transition tables:

$x_1 \in X$	$x_2 \in X \setminus \{x_1\}$ with $x_1 \to_{\{i\}} x_2$
MFN	-
CU(i,j)	MFN
CU(j,k)	-
CU(k,i)	MFN
CUGFT	CU(j,k)
FTA(i,j)	MFN
FTA(j,k)	-
FTA(k,i)	MFN
FTAHub(i)	MFN, FTA(i,j), FTA(k,i)
FTAHub(j)	FTA(j,k)
FTAHub(k)	FTA(j,k)
FTAGFT	FTA(j,k), FTAHub(j), FTAHub(k)
MTA(i,j)	MFN
MTA(j,k)	-
MTA(k,i)	MFN
MTAGFT	MTA(j,k)

(A) The transition table for coalition $\{i\}$, $i \in N$.

$x_1 \in X$	$x_2 \in X \setminus \{x_1\}$ with $x_1 \to_{\{i,j\}} x_2$
MFN	CU(i,j), FTA(i,j), MTA(i,j)
CU(i,j)	MFN, FTA(i,j), MTA(i,j)
CU(j,k)	MFN, CU(i,j), FTA(i,j), MTA(i,j)
CU(k,i)	MFN, CU(i,j), FTA(i,j), MTA(i,j)
CUGFT	MFN, CU(i,j), CU(j,k), CU(k,i), FTA(i,j), MTA(i,j)
FTA(i,j)	MFN, CU(i,j), MTA(i,j)
FTA(j,k)	MFN, CU(i,j), FTA(i,j), FTAHub(j), MTA(i,j)
FTA(k,i)	MFN, CU(i,j), FTA(i,j), FTAHub(i), MTA(i,j)
FTAHub(i)	MFN, CU(i,j), FTA(i,j), FTA(k,i), MTA(i,j)
FTAHub(j)	MFN, CU(i,j), FTA(i,j), FTA(j,k), MTA(i,j)
FTAHub(k)	MFN, CU(i,j), FTA(i,j), FTA(j,k), FTA(k,i), FTAGFT, MTA(i,j)
FTAGFT	MFN, CU(i,j), FTA(i,j), FTA(j,k), FTA(k,i), FTAHub(i), FTAHub(j), FTAHub(k), MTA(i,j)
MTA(i,j)	MFN, CU(i,j), FTA(i,j)
MTA(j,k)	MFN, CU(i,j), FTA(i,j), MTA(i,j)
MTA(k,i)	MFN, CU(i,j), FTA(i,j), MTA(i,j)
MTAGFT	MFN, CU(i,j), FTA(i,j), MTA(i,j), MTA(j,k), MTA(k,i)

(B) The transition table for coalition $\{i, j\}$, $i, j \in N$, $i \neq j$.

TABLE 16. The network structure as transition tables

B.5. **Additional Remarks.**

B.5.1. *Pseudo-GFT.* In Appendix B.2 a special case of 'Pseudo-GFT' is a possibility. Namely, in the case of a hub structure with both non-hub nodes capping at zero the trade agreement amounts to the same tariff structure (and welfare) of a GFT. If it were ever part of the stable set, then it would necessarily need to be considered de facto GFT even though it is not de jure GFT. However, in our analysis this case never occured and it is therefore a negligible oddity.

B.5.2. *A Special Case.* As can be seen in Table 15 the overall welfare is equal in case of MFN, CU(i^c, j^c), and FTA(i^c, j^c) even though the tariff structure is different. The following explores this equivalence in order to provide an insight into the underlying mechanics. In terms of tariff structure both CU(i^c, j^c) and FTA(i^c, j^c) are the same and therefore it is sufficient to compare MFN with CU(i^c, j^c) when only interested in (effects on) welfare. Now, Table 17 shows us the differences in the welfare (components) both on the individual as well as on the joint/overall level, which are computed from the expressions in Table 18 and 19.

$\Delta(MFN, CU(i^c, j^c))$	
TR_i	$1/24(e_j + e_k)(2e_j - e_k)$
CS_i	$(32e_i^2 + 64e_ie_k - 13e_j^2 - 26e_je_k + 19e_k^2)/1152$
PS_i	$1/12e_i(-e_i - e_k)$
W_i	$(-64e_i^2 - 32e_ie_k + 83e_j^2 + 22e_je_k - 29e_k^2)/1152$
TR_j	$1/24(e_i + e_k)(2e_i - e_k)$
CS_j	$(-13e_i^2 - 26e_ie_k + 32e_j^2 + 64e_je_k + 19e_k^2)/1152$
PS_j	$1/12e_j(-e_j - e_k)$
W_j	$(83e_i^2 + 22e_ie_k - 64e_j^2 - 32e_je_k - 29e_k^2)/1152$
TR_k	0
CS_k	$19(-e_i^2 - 2e_ie_k - e_j^2 - 2e_je_k - 2e_k^2)/1152$
PS_k	$1/24e_k(e_i + e_j + 2e_k)$
W_k	$(-19e_i^2 + 10e_ie_k - 19e_j^2 + 10e_je_k + 58e_k^2)/1152$

(A) The difference in the individual welfare (components) depending on endowments

$\Delta(MFN, CU(i^c, j^c))$	
$TR_i + TR_j$	$1/24(2e_i^2 + e_ie_k + 2e_j^2 + e_je_k - 2e_k^2)$
$CS_i + CS_j$	$19(e_i^2 + 2e_ie_k + e_j^2 + 2e_je_k + 2e_k^2)/1152$
$PS_i + PS_j$	$1/12(-e_i^2 - e_ie_k - e_j^2 - e_je_k)$
$W_i + W_j$	$(19e_i^2 - 10e_ie_k + 19e_j^2 - 10e_je_k - 58e_k^2)/1152$
$\sum_{n\in N} TR_n$	$1/24(2e_i^2 + e_ie_k + 2e_j^2 + e_je_k - 2e_k^2)$
$\sum_{n\in N} CS_n$	0
$\sum_{n\in N} PS_n$	$1/24(-2e_i^2 - e_ie_k - 2e_j^2 - e_je_k + 2e_k^2)$
$\sum_{n\in N} W_n$	0

(B) The difference in the joint/overall welfare (components) depending on endowments

TABLE 17. The difference in the welfare (components) depending on endowments

Welfare Components	
MFN	
TR_i	$(e_j + e_k)^2/32$
CS_i	$(18e_i^2 + 13e_j^2 + 13e_k^2 + 8e_je_k + 18e_i(e_j + e_k))/128$
PS_i	$-e_i(-16\alpha + 6e_i + 3e_j + 3e_k)/8$
TR_j	$(e_i + e_k)^2/32$
CS_j	$(13e_i^2 + 18e_j^2 + 13e_k^2 + 8e_ie_k + 18e_j(e_i + e_k))/128$
PS_j	$-e_j(-16\alpha + 3e_i + 6e_j + 3e_k)/8$
TR_k	$(e_i + e_j)^2/32$
CS_k	$(13e_i^2 + 13e_j^2 + 18e_k^2 + 8e_ie_j + 18e_k(e_i + e_j))/128$
PS_k	$-e_k(-16\alpha + 3e_i + 3e_j + 6e_k)/8$
CU(i^c, j^c)	
TR_i	$-(5e_j - 7e_k)(e_j + e_k)/96$
CS_i	$(65e_i^2 + 65e_j^2 + 49e_k^2 + 49e_je_k + e_i(81e_j + 49e_k))/576$
PS_i	$-e_i(-48\alpha + 16e_i + 9e_j + 7e_k)/24$
TR_j	$-(5e_i - 7e_k)(e_i + e_k)/96$
CS_j	$(65e_i^2 + 65e_j^2 + 49e_k^2 + 49e_ie_k + e_j(81e_i + 49e_k))/576$
PS_j	$-e_j(-48\alpha + 9e_i + 16e_j + 7e_k)/24$
TR_k	$(e_i + e_j)^2/32$
CS_k	$(17e_i^2 + 17e_j^2 + 25e_k^2 + 25e_ie_k + 25e_je_k + 9e_ie_j)/144$
PS_k	$-e_k(-24\alpha + 5e_i + 5e_j + 10e_k)/12$

(A) The individual welfare (components) depending on endowments

Welfare Components	
MFN	
$TR_i + TR_j$	$1/32(e_i^2 + 2e_ie_k + e_j^2 + 2e_je_k + 2e_k^2)$
$CS_i + CS_j$	$1/128(31e_i^2 + 36e_ie_j + 26e_ie_k + 31e_j^2 + 26e_je_k + 26e_k^2)$
$PS_i + PS_j$	$1/8(-6e_i^2 - 6e_ie_j - 3e_ie_k - 6e_j^2 - 3e_je_k) + 2\alpha(e_i + e_j)$
$\sum_{n\in N} TR_n$	$1/16(e_i^2 + e_ie_j + e_ie_k + e_j^2 + e_je_k + e_k^2)$
$\sum_{n\in N} CS_n$	$11/32(e_i^2 + e_ie_j + e_ie_k + e_j^2 + e_je_k + e_k^2)$
$\sum_{n\in N} PS_n$	$1/4(-3e_i^2 - 3e_ie_j - 3e_ie_k - 3e_j^2 - 3e_je_k - 3e_k^2) + 2\alpha\left(\sum_{n\in N} e_n\right)$
CU(i^c, j^c)	
$TR_i + TR_j$	$1/96(-5e_i^2 + 2e_ie_k - 5e_j^2 + 2e_je_k + 14e_k^2)$
$CS_i + CS_j$	$1/288(65e_i^2 + 81e_ie_j + 49e_ie_k + 65e_j^2 + 49e_je_k + 49e_k^2)$
$PS_i + PS_j$	$1/24(-16e_i^2 - 18e_ie_j - 7e_ie_k - 16e_j^2 - 7e_je_k) + 2\alpha(e_i + e_j)$
$\sum_{n\in N} TR_n$	$1/48(-e_i^2 + 3e_ie_j + e_ie_k - e_j^2 + e_je_k + 7e_k^2)$
$\sum_{n\in N} CS_n$	$11/32(e_i^2 + e_ie_j + e_ie_k + e_j^2 + e_je_k + e_k^2)$
$\sum_{n\in N} PS_n$	$1/24(-16e_i^2 - 18e_ie_j - 17e_ie_k - 16e_j^2 - 17e_je_k - 20e_k^2) + 2\alpha\left(\sum_{n\in N} e_n\right)$

(B) The joint/overall welfare (components) depending on endowments

TABLE 18. The welfare (components) depending on endowments

Trade Agreement	Individual/Joint/Overall Welfare
MFN	
↳ W_i	$1/128(-78e_i^2 + 256\alpha e_i - 30e_ie_j - 30e_ie_k + 17e_j^2 + 16e_je_k + 17e_k^2)$
↳ W_j	$1/128(17e_i^2 - 30e_ie_j + 16e_ie_k - 78e_j^2 + 256\alpha e_j - 30e_je_k + 17e_k^2)$
↳ W_k	$1/128(17e_i^2 + 16e_ie_j - 30e_ie_k + 17e_j^2 - 30e_je_k - 78e_k^2 + 256\alpha e_k)$
↳ $W_i + W_j$	$1/128(-61e_i^2 - 60e_ie_j - 14e_ie_k - 61e_j^2 - 14e_je_k + 34e_k^2) + 2\alpha(e_i + e_j)$
↳ $\sum_{n\in N} W_n$	$1/32(-11e_i^2 - 11e_ie_j - 11e_ie_k - 11e_j^2 - 11e_je_k - 11e_k^2) + 2\alpha\left(\sum_{n\in N} e_n\right)$
CU(i^c, j^c)	
↳ W_i	$1/576(-319e_i^2 + 1152\alpha e_i - 135e_ie_j - 119e_ie_k + 35e_j^2 + 61e_je_k + 91e_k^2)$
↳ W_j	$1/576(35e_i^2 - 135e_ie_j + 61e_ie_k - 319e_j^2 + 1152\alpha e_j - 119e_je_k + 91e_k^2)$
↳ W_k	$1/288(43e_i^2 + 36e_ie_j - 70e_ie_k + 43e_j^2 - 70e_je_k - 190e_k^2 + 576\alpha e_k)$
↳ $W_i + W_j$	$1/288(-142e_i^2 - 135e_ie_j - 29e_ie_k - 142e_j^2 - 29e_je_k + 91e_k^2) + 2\alpha(e_i + e_j)$
↳ $\sum_{n\in N} W_n$	$1/32(-11e_i^2 - 11e_ie_j - 11e_ie_k - 11e_j^2 - 11e_je_k - 11e_k^2) + 2\alpha\left(\sum_{n\in N} e_n\right)$

TABLE 19. The welfare depending on endowments

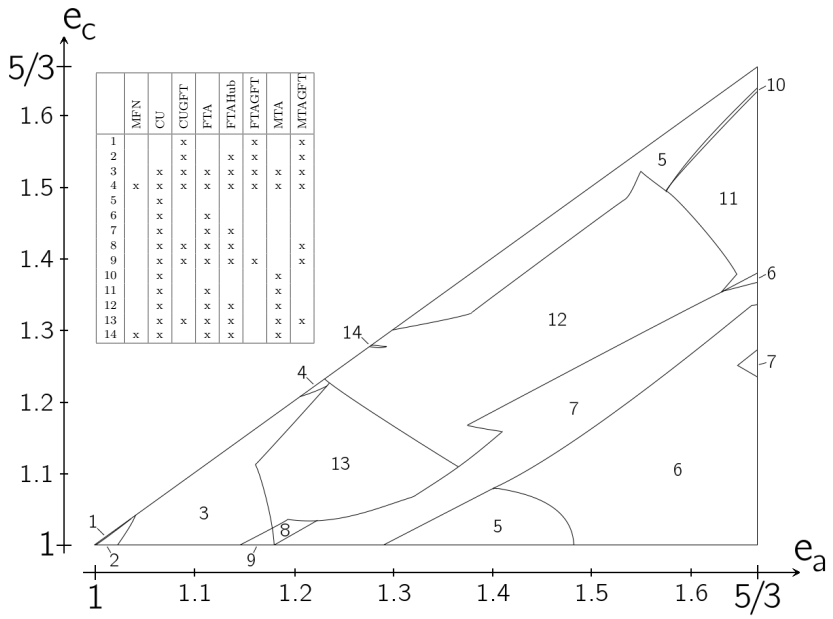
Now, recall that capping at the MFN-tariff for both members of a customs union, in this case CU(i^c, j^c), occurs when the endowment of the non-member is above a minimal value determined by the endowments of the members, $\max\{e_i, e_j\} < \frac{11}{13}e_k$ (Appendix B.2). Using this together with the general assumptions on the relation of endowments, the following effects on welfare (components) take place, where each expression of the type '$+c$' for some c is positive and each '$-c$' negative:

$\Delta(MFN, CU(i^c, j^c))$		
Country i	Country j	Country k
TR		
$+\tau_i$	$+\tau_j$	0
$+\tau_{ij}$		0
$+\tau$		
CS		
$+\gamma_i$	$+\gamma_j$	$-\gamma_k$
$+\gamma_{ij}$		$-\gamma_k$
0		
PS		
$-\rho_i$	$-\rho_j$	$+\rho_k$
$-\rho_{ij}$		$+\rho_k$
$-\rho$		
W		
$\pm\omega_i$	$\pm\omega_j$	$+\omega_k$
$-\omega_{ij}$		$+\omega_k$
0		

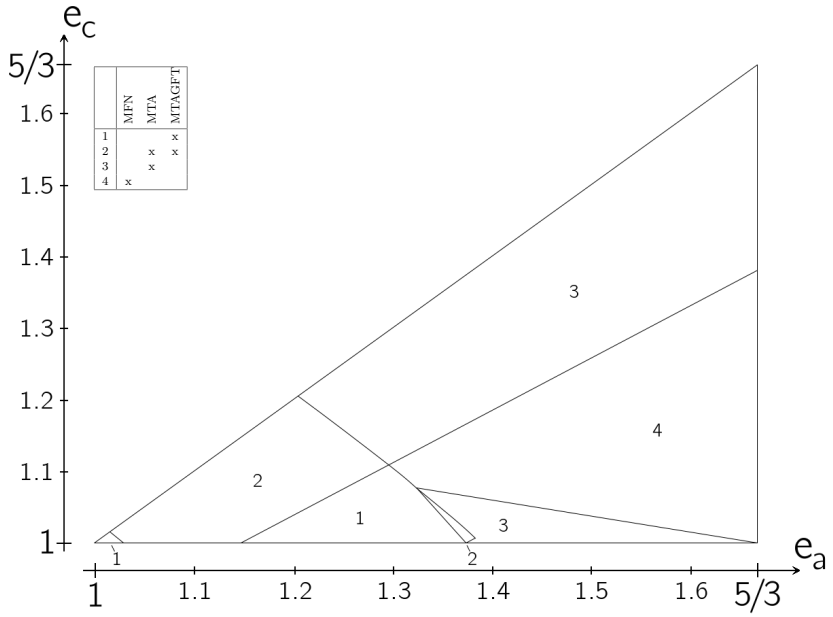
TABLE 20. The effect on the welfare (components)

Appendix C. Analysis

C.1. **Additional Graphics.** The following provides detailed figures:

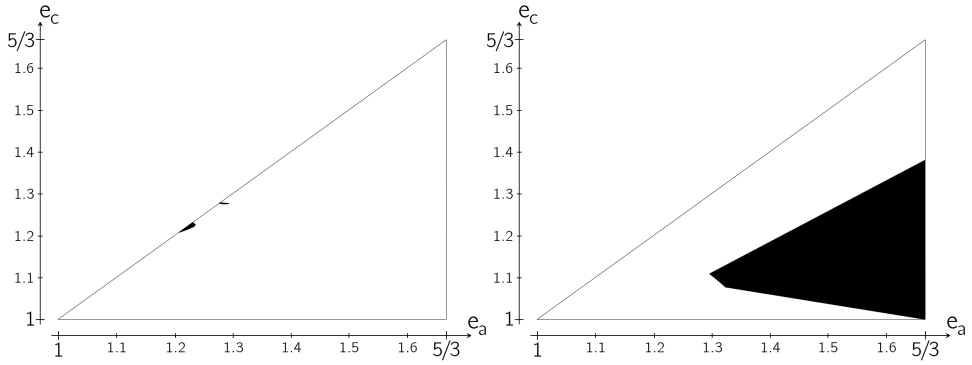


(A) Overall Stability with PTAs



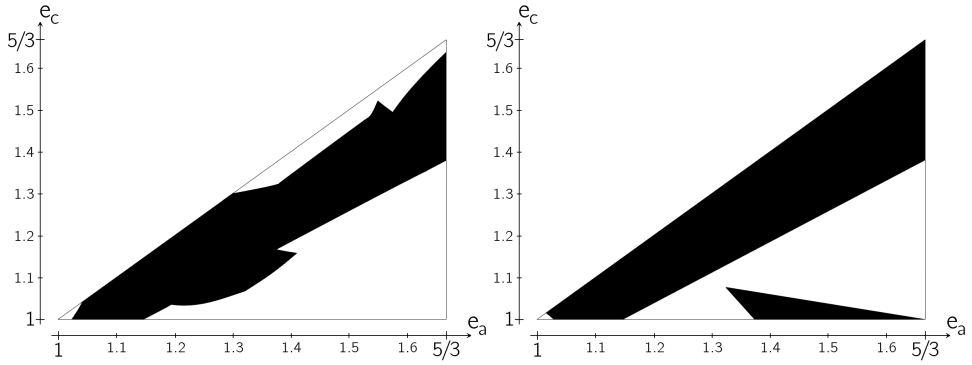
(B) Overall Stability without PTAs

FIGURE 21. Overall Stability with and without PTAs



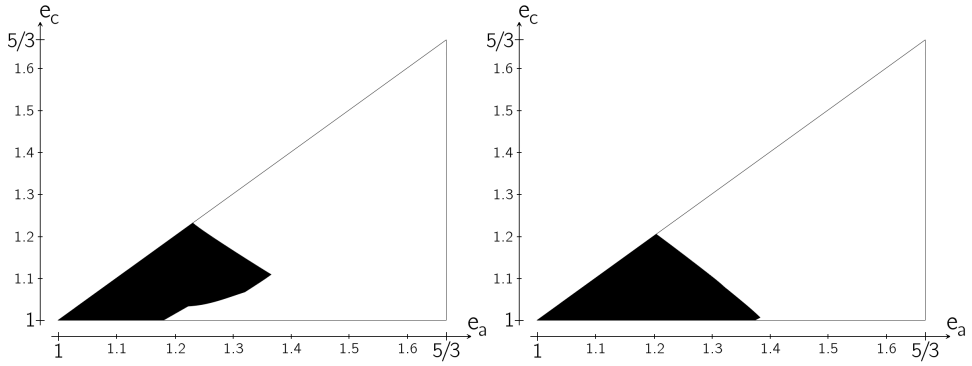
(A) with PTAs (B) without PTAs

FIGURE 22. Stability of MFN



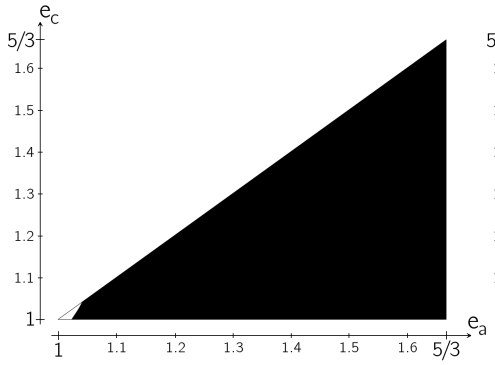
(A) with PTAs (B) without PTAs

FIGURE 23. Stability of MTA

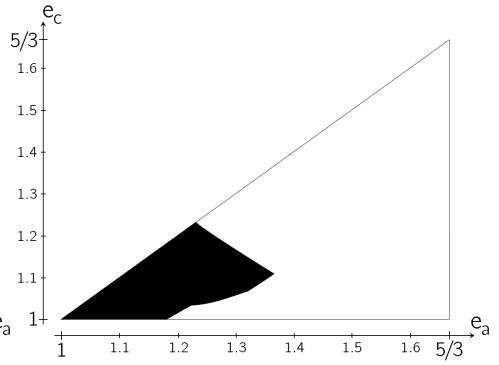


(A) with PTAs (B) without PTAs

FIGURE 24. Stability of MTAGFT

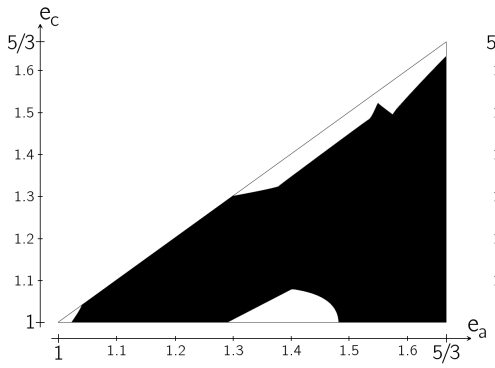


(A) CU

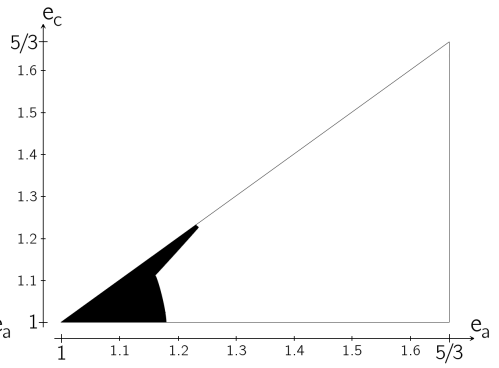


(B) CUGFT

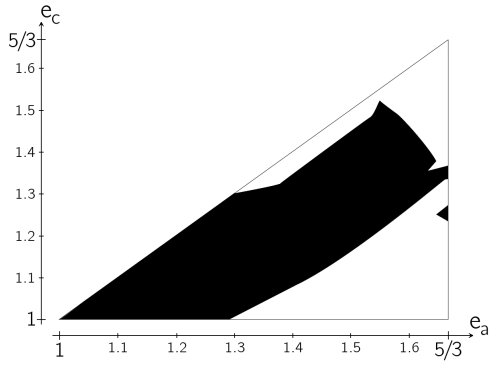
FIGURE 25. Stability of CU



(A) FTA



(B) FTAGFT



(C) FTAHub

FIGURE 26. Stability of FTA

C.2. **Exact Intervals.** The table here lists the exact intervals where each specific trade agreement is part of the stable set (for the border of the parameter space):

Trade Agreement	Exact Interval(s)
	$e_b = e_{\min} \leq e_c \leq e_{\max} = e_a$
CU(b,c)	[1.0000000000000000,1.3380093520374081]
CU(c,a)	[1.3259853039412157,1.6666666666666667]
FTA(b,c)	[1.0000000000000000,1.3807615230460921]
FTA(c,a)	[1.3353373413493654,1.6305945223780896]
FTAHub(c)	[1.2364729458917836,1.2698730794923179]
	[1.3353373413493654,1.3647294589178356]
MTA(c,a)	[1.379425517702071,1.635938543754175]
	$e_b = e_c = e_{\min} \leq e_a \leq e_{\max}$
CU(a,b)	[1.0240480961923848,1.1108884435537743]
CU(b,c)	[1.0240480961923848,1.6666666666666667]
CU(c,a)	[1.0240480961923848,1.1108884435537743]
CUGFT	[1.0000000000000000,1.1803607214428857]
FTA(a,b)	[1.0240480961923848,1.2404809619238477]
FTA(b,c)	[1.0240480961923848,1.29124916499666]
	[1.483633934535738,1.6666666666666667]
FTA(c,a)	[1.0240480961923848,1.2404809619238477]
FTAHub(b)	[1.0013360053440215,1.29124916499666]
FTAHub(c)	[1.0013360053440215,1.29124916499666]
FTAGFT	[1.0000000000000000,1.1803607214428857]
MTA(a,b)	[1.0240480961923848,1.1469605878423514]
MTA(c,a)	[1.0240480961923848,1.1469605878423514]
MTAGFT	[1.0000000000000000,1.1803607214428857]
	$e_b = e_{\min} \leq e_a = e_c \leq e_{\max}$
MFN	[1.2044088176352705,1.2297929191716768]
	[1.2752171008684035,1.276553106212425]
CU(a,b)	[1.0454241816967267,1.2498329993319974]
CU(b,c)	[1.0454241816967267,1.2498329993319974]
CU(c,a)	[1.0494321977287908,1.6666666666666667]
CUGFT	[1.0000000000000000,1.2297929191716768]
FTA(a,b)	[1.0494321977287908,1.2925851703406814]
FTA(b,c)	[1.0494321977287908,1.2925851703406814]
FTA(c,a)	[1.0454241816967267,1.2925851703406814]
FTAHub(a)	[1.0454241816967267,1.276553106212425]
FTAHub(b)	[1.0454241816967267,1.2925851703406814]
FTAHub(c)	[1.0454241816967267,1.276553106212425]
FTAGFT	[1.0000000000000000,1.2297929191716768]
MTA(a,b)	[1.0494321977287908,1.2244488977955912]
MTA(b,c)	[1.0494321977287908,1.2244488977955912]
MTA(c,a)	[1.0454241816967267,1.2925851703406814]
MTAGFT	[1.0000000000000000,1.2297929191716768]

TABLE 27. The exact intervals of stability with PTAs

Trade Agreement	Exact Interval(s)
	$e_b = e_{\min} \leq e_c \leq e_{\max} = e_a$
MFN	[1.0000000000000000,1.3780895123580494]
MTA(b,c)	[1.0000000000000000,1.0000000000000000]
MTA(c,a)	[1.379425517702071,1.6666666666666667]
	$e_b = e_c = e_{\min} \leq e_a \leq e_{\max}$
MFN	[1.6666666666666667,1.6666666666666667]
MTA(a,b)	[1.0307281229124916,1.1469605878423514]
MTA(b,c)	[1.0307281229124916,1.1469605878423514] [1.3754175016700068,1.6666666666666667]
MTA(c,a)	[1.0307281229124916,1.1469605878423514]
MTAGFT	[1.0000000000000000,1.3740814963259853]
	$e_b = e_{\min} \leq e_a = e_c \leq e_{\max}$
MTA(a,b)	[1.0160320641282565,1.1202404809619237]
MTA(b,c)	[1.0160320641282565,1.1202404809619237]
MTA(c,a)	[1.0160320641282565,1.6666666666666667]
MTAGFT	[1.0000000000000000,1.203072812291249]

TABLE 28. The exact intervals of stability without PTAs

CHAPTER 3

The Funding of Overconfident Entrepreneurs

1. Introduction

Every society has its share of people wanting to start a new business - despite a poor chance of survival and return on investment ('a private equity premium puzzle', Moskowitz and Vissing-Jorgensen (2002)). The two central issues for the aspiring entrepreneur in this context are the source of funding and the level of commitment. Usually, the two funding choices under consideration are bank and venture capitalist, i.e. debt and equity, while the commitment, i.e. the non-monetary investment, possibly takes many different forms (often simplified as effort). At the same time, every potential financier needs to assess how realistic (or optimistic) the underlying business plan actually is. The analysis of the overall process with a focus on a biased 'homo entreprenaurus' (Uusitalo (2001)) is at the heart of this paper.

The source of funding for entrepreneurs is a topic that has been and still is extensively investigated. A comprehensive analysis of the research that focuses on venture capital can be found in 'A survey of venture capital research' (2011) done by Da Rin, Hellmann, and Puri. The authors conclude in the end that despite the progress there are still open questions with respect to 'the choice between alternative sources of financing'. Furthermore, various empirical studies connect entrepreneurship with the bias of (unrealistic) optimism and overconfidence - take for example Cooper, Dunkelberg, and Woo (1988) or Camerer and Lovallo (1999).[1] Consequently, a number of papers study the effect of this self-illusion on different aspects of entrepreneurship (see the next section for a selection of this research). Among these, the investment level of the entrepreneur in the form of effort often plays a central role (see Malmendier and Tate (2005)[2]). However, to our knowledge the literature lacks a comprehensive model that features the investment of effort in a start-up under the assumption of the aforementioned perception bias while simultaneously answering the question about the source of financing put forth by Da Rin, Hellmann, and Puri. The model presented in this paper tries to fill this perceived gap in the literature and contribute to the academic discourse through the analysis of the effect of the existence and proportion of overconfident entrepreneurs on the overall choice problem and the induced social welfare.

[1] While psychologists clearly distinguish between (unrealistic) optimism and overconfidence, economists often group them together and use the terms interchangeably. The terms share the principle of overestimation, but (unrealistic) optimism affects the probability/magnitude of events while overconfidence relates to the ability/accuracy of the individual. Now, the entrepreneur in our model potentially exhibits both of these characteristics (at the same time). However, one could make the argument that the success/failure probability of a start-up depends at least in some form on the ability of the entrepreneur. Therefore, the wording in this paper follows the convention of the economic literature, i.e. mixing the two terms, but ultimately leans towards overconfidence.

[2] It is technically a paper concerned with corporate investment of funds, not effort.

In our framework, the entrepreneur faces a project of unknown quality that requires the investment of both funds and effort. The entrepreneur can choose to work with a venture capitalist, who provides effort and funding at the cost of sharing the profit, or resort to a bank, which provides only funding at the cost of a payment of interest. Furthermore, the entrepreneur invests effort both in advance and during the actual project - in case of the venture capitalist, the additional provision of effort by the venture capitalist takes place simultaneously (during the actual project). Finally, the perception of the quality of the project depends on the type of the entrepreneur, which is either optimistic or realistic. While a realistic entrepreneur receives a signal corresponding to the actual quality, an optimistic entrepreneur receives a distorted signal - perceiving the world through rose-tinted glasses. However, both types interpret the signal as an undistorted one.

In order to (analytically) solve the overall decision problem, our paper employs the concept of Perfect Bayesian Equilibrium, with a focus on pure strategies. Further, with respect to the underlying signaling game, our paper concentrates on studying separating equilibria - for the sake of a set-up where realistic and optimistic entrepreneurs potentially choose different strategies. The analysis contains two different perspectives. Each of them features a number of comparative static effects of the model parameters, but the first one concerns itself with the player's choices and their profits, while the second one examines the effect of other parameters on the comparative static effect of overconfidence on the social welfare (i.e. the sign of the cross-derivatives). Ultimately, our paper characterizes different constellations of the model parameters under which an increase in the share of overconfidence leads to an increase in social welfare.

This paper has six more sections, that is Section 2 features the related literature, Section 3 presents the model, Section 4 the equilibria, Section 5 the analysis, while Section 6 discusses the related literature again, and Section 7 concludes the paper.

2. Related Literature

As mentioned in the introduction, our model essentially combines three aspects, that of a possibly overconfident entrepreneur facing a project of unknown quality and the two related choices about the source of funding and the investment of effort. Let us start the selection of related literature with three papers that provided part of the foundation and justification for economists to deviate from the usual assumption of a fully rational individual when studying entrepreneurs. Alternatively, consult Taylor and Brown (1988) (and Taylor and Brown (1994)) for an extensive overview of the work of psychologists on the topic of self-illusions.

Analyzing the perceived chances of success (resp. failure) by new business owners, Cooper, Woo, and Dunkelberg (1988) discover significantly optimistic assessments. Contrary to the survival rate of start-ups and even when controlling for factors influencing the chances (personal background, nature of the firm), the authors observe that new entrepreneurs 'experience feelings of entrepreneurial euphoria'.

In a laboratory experiment, Camerer and Lovallo (1999) study the market entry of subjects in a setting where the payoff depends on relative skill. Consistent with the conjecture that overconfidence leads to excess entry, the majority of subjects who enter believe that 'the total profit earned by all entrants will be negative, but their own profit will be positive'.

Using panel data of Forbes 500 CEOs, Malmendier and Tate (2005) argue that the trait of overconfidence can explain distortions of corporate investments. Specifically, overconfident managers overinvest as long as internal funds are available but underinvest as soon as external funds are required.

In addition to these, consider the work by De Bondt and Thaler (1994) for a selective review of behavioral finance. The paper discusses the relevance of several key behavioral concepts, among these excessive self-confidence, to important topics in the finance literature.

With these papers on the aspect of an overconfident entrepreneur in mind, let us discuss the related literature on the two entrepreneurial choices (funds and effort). The papers of Manove and Padilla (1999) and de Bettignies and Brander (2007) deserve a special mention here as the main inspiration for the implementation of these ideas within our framework.

First, Manove and Padilla (1999) analyze the influence of optimistic entrepreneurs on the provision of funds through banks. In their model, the entrepreneur is either optimistic or realistic, but always believes to be realistic, and accordingly interprets a signal about the quality of the project. On that basis, the entrepreneur then chooses an investment level to be requested from the bank. As part of the analysis, the authors contrast the market equilibrium with the social planner's alternative in terms of the nature of the equilibria (pooling or separating) and other choices.

Second, de Bettignies and Brander (2007) analyze the competition of two sources of funding for an aspiring entrepreneur, bank and venture capital - with associated costs and benefits. Furthermore, the entrepreneur and the venture capitalist face the choice of investing (additional) effort in the project. In their analysis, the authors investigate comparative static effects of the model parameters on the respective strategies and sources of funding.

In addition to these two, let us discuss a collection of other papers that are closely related to ours, specifically in terms of model specifications but also in spirit. Note that it is only a small selection of papers from this large section of the literature. Consult the aforementioned survey by Da Rin, Hellmann, and Puri (2011) for more.

In the paper of Landier and Thesmar (2003), the entrepreneur faces a project with binary levels of quality and similar business strategies. The entrepreneur receives a corresponding signal about the quality of the project according to a hidden type, possibly optimistic, and then chooses a business strategy, essentially risky or safe. The authors focus on two financial contracts, specifically short- and long-term debt, and contrast their analysis with empirical evidence.

The framework by Ueda (2004) features an entrepreneur with private information about a project choosing between a bank and a venture capitalist for funding. When negotiating with the bank the information stays private, while it is fully revealed with the venture capitalist. Additionally, in the case of choosing the venture capitalist there is a risk of expropriation of the project. The author then studies the effect of various model parameters on the source of funding.

Next, the principal-agent model by de la Rosa (2011) focuses on the analysis of the incentive contracts in a moral-hazard framework with varying degrees of overconfidence of the agent. Similarly, Vilanova, Marchand, and Hichri (2015) use a laboratory setting to study the effect of entrepreneurial overconfidence on the nature of financing and advising contracts. Coelho, de Meza, and Reyniers (2004) construct

a simple borrower-lender model and utilize experimental evidence to investigate the implications of different policies on entrepreneurship in the context of optimism.

3. Model

An entrepreneur has access to a project that requires a fixed amount of initial investment I to be realized. In order to cover the cost, the entrepreneur considers two different sources of funding s. On one side, a bank offers a collateral-free loan with a fixed interest rate r. On the other side, a venture capitalist offers to trade services, including financing the project, in exchange for a (profit) share λ. Now, the outcome of the project $Y(Q, E)$ is determined by the quality Q of the project and the effort E that is put into it.

First, the project quality Q is either good G or bad B and actually realizes with probability γ and $1-\gamma$ respectively. Similarly, a corresponding signal q that is transmitted to the entrepreneur is either good g or bad b. The signal coincides with the actual project quality with probability $1-\epsilon$ (signal q for state Q). Thus, the term ϵ represents noise, i.e. imperfect information, implying that the signal deviates from the actual information with probability ϵ. Additionally, the perception $\tilde{q}$ of the noisy signal depends on the type of the entrepreneur A. Each entrepreneur is either realistic R or optimistic O with probabilities μ and $1-\mu$ respectively. Also, the perceived signal is either good $\tilde{g}$ or bad $\tilde{b}$. Now, if the entrepreneur is a realist, then the perceived signal corresponds to the noisy signal ($\tilde{q}$ matching q). However, if the entrepreneur is an optimist, then the perceived signal is always good ($\tilde{g}$).[3] Consequentially, in the case of good news the interpretation of both types coincides, while bad news yield a differing one.

Second, the effort $E(e_0, e_1, e_2)$ is composed of an initial as well as an additional effort provided by the entrepreneur - denoted by e_0 and e_1 respectively - and also of an effort provided by the venture capitalist - denoted by e_2.[4] The choice about the initial effort is actually the first step in the process while that of the others constitutes the last step. The effort provision by the venture capitalist is limited to a decision between the effort levels $\bar{e} > 0$ and zero depending on whether the project gets funded via the venture capitalist or not.[5] Contrary to this binary case, the entrepreneur chooses the optimal effort level from the range $[0, +\infty)$ each time. In any case, both the entrepreneur and the venture capitalist incur cost of $C(e_i)$ for their respective efforts e_i, $i = 1, 2, 3$.

The provision of funding, by the bank or the venture capitalist, entails cost of $\bar{r}I$, where $\bar{r}$ is the risk-free interest rate. The bank charges a fixed premium on top of the risk-free interest rate, i.e. $r = \bar{r} + \bar{p}$, while the venture capitalist demands the optimal share of the project from the range $[0, 1]$. Note that the venture capitalist can also choose to provide 'no funding' as the outside option. Similarly, the entrepreneur might decide to do 'no project'. Additionally, the initial investment is assumed to

[3] It essentially constitutes a type of ignorance about available information on the side of the optimistic entrepreneur. In other words, from the perspective of the optimistic entrepreneur there is no such thing as 'bad news', everything is 'good news'.

[4] The initial effort captures the idea of a 'planning' phase in the project, where the entrepreneur develops the business plan, and the additional effort captures the 'implementation' phase. Also, the effort of the venture capitalist could be interpreted as a 'package of services' that is provided to the entrepreneur.

[5] In other words, the 'package of services' provided by the venture capitalist to the entrepreneur corresponds to a fixed value independent of the project's properties and follows from other decisions.

be fully recoupable from the project in our model, which makes re-financing the only cost for the provision of funding.[6]

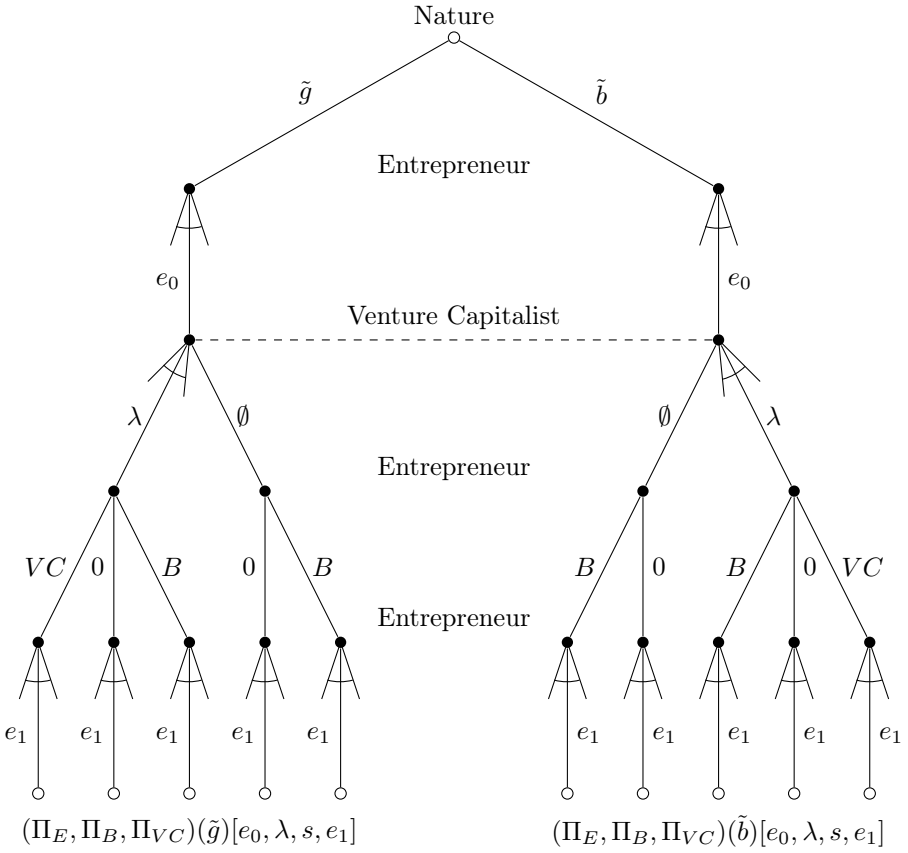


FIGURE 1. The extensive-form game of the decision process

The overall timing and structure of the game are depicted in detail in Figure 1.[7] For a detailed description of the probabilistic process that determines the signal given by nature see Section 3.1. Finally, the payoff functions of the players can be found as part of Section 3.2.

3.1. **Nature's Signal and Players' Beliefs.** A vital part of our model are the players' beliefs, specifically those of the entrepreneur, about the project's quality. All of these ultimately depend on the probabilistic process that determines the signal that nature sends to the entrepreneur (which in turn influences the initial effort of the entrepreneur, serving as the signal for the venture capitalist). An overview of this is presented in Figure 2.

[6] In certain scenarios this simply corresponds to the existence of efficient markets.

[7] It is a negligible but still noteworthy detail that the list of players whose actions are significant and that of those whose payoffs are relevant differ from each other. Technically, the complete list of players consists of nature, entrepreneur, venture capitalist, and bank. However, of these four only the actions of the first three are significant and only the payoffs of the last three are relevant. The bank does not face any decision but receives a payoff while the reverse applies to nature.

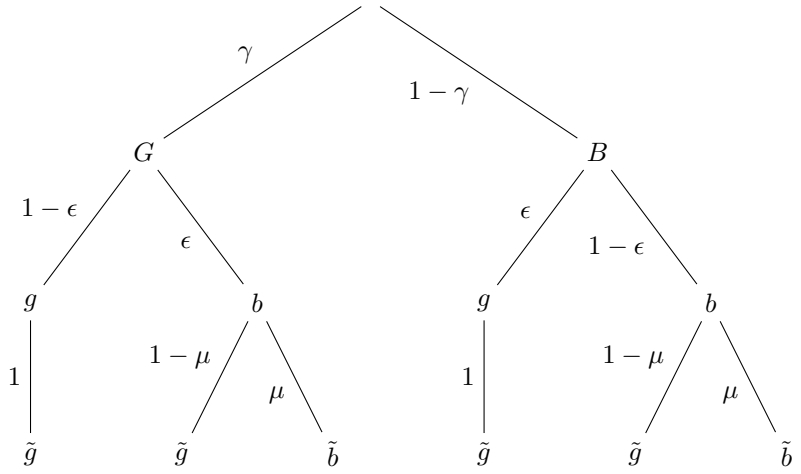


FIGURE 2. Overview of the probabilistic process which determines the project's quality and its signal.

As mentioned before, in the first two steps the project's quality and its corresponding (potentially noisy) signal are determined. Note that the probabilities of these steps are common knowledge and unanimous consensus among the players while the view on those of the next (and final) step differs due to the nature of the overconfidence. The signal that the entrepreneur obtains in the end depends on the (noisy) signal about the project's quality as well as on the confidence type of the entrepreneur. As long as the initial signal suggests 'good news' both types of entrepreneurs receive it accordingly. As soon as it means 'bad news' though only the realistic entrepreneur obtains the corresponding signal while the optimistic one receives the 'good news' one again. As a consequence, the probabilities in this case depend on the share of optimists (and realists) - again common knowledge. However, while initially the entrepreneur and the funding parties share the knowledge and the unanimous consensus, when it comes to actually interpreting their own signal, both types of entrepreneur believe themselves to be a realist and act accordingly.[8] Therefore, instead of the objective probabilities given by the shares, the entrepreneur employs subjective probabilities corresponding to $\mu = 1$ (resp. $1 - \mu = 0$). Following the nomenclature of Manove and Padilla (1999), the entrepreneur thus classifies as 'near-rational' agent. In the following, let us formally present the individual beliefs corresponding to this probabilistic process.

3.1.1. *Entrepreneur.* Let us denote the beliefs of a type of entrepreneur by π_E to better distinguish them from objective probabilities P. Independent of the type, the entrepreneur's belief about the project's quality conditioned on the (final) signal is given by $\pi_E(Q|\tilde{q}) = P(Q|q)$, where the objective probability is calculated using Bayes' rule.[9] Note that this objective probability includes the error probability.

[8]In a sense this means that the optimist also exhibits (extreme) overconfidence on a meta level. In addition to overconfidence the optimist exhibits denial about the possibility of overconfidence.

[9]In terms of the other parameters, $\pi_E(G|\tilde{g}) = \frac{(1-\epsilon)\gamma}{(1-\epsilon)\gamma+\epsilon(1-\gamma)}$ and $\pi_E(B|\tilde{b}) = \frac{(1-\epsilon)(1-\gamma)}{(1-\epsilon)(1-\gamma)+\epsilon\gamma}$ as well as $\pi_E(B|\tilde{g}) = \frac{\epsilon(1-\gamma)}{\epsilon(1-\gamma)+(1-\epsilon)\gamma}$ and $\pi_E(G|\tilde{b}) = \frac{\epsilon\gamma}{\epsilon\gamma+(1-\epsilon)(1-\gamma)}$.

Thus, both realist and optimist consider the possibility of distortion via noise, but not the potential for overconfidence.

3.1.2. *Bank.* Contrary to the entrepreneur who receives a direct signal about the quality of the project, the bank only receives an indirect one via the initial effort provided by the entrepreneur. However, as both volume and interest rate of the collateral-free loan are fixed, the bank does not need to make any decision - it serves the purpose of another outside option (besides the 'no project' one). Consequentially, it forgoes using the information of the initial effort of the entrepreneur and does not form any belief system about the (actual) quality of the project.[10]

3.1.3. *Venture Capitalist.* Similar to the bank, the venture capitalist receives the initial effort provided by the entrepreneur as signal. Denote the corresponding beliefs by π. As a component, the venture capitalist exhibits a belief $\pi(\tilde{q}|e_0)$ about the overall signal conditioned on the initial effort of the entrepreneur. Note, that this belief depends on the equilibrium strategy chosen by the entrepreneur and therefore becomes part of the equilibrium tuple itself. Using the information about the probabilistic process, the venture capitalist then computes the belief about the actual project quality given the initial effort via $\pi(Q|e_0) = \sum_{\tilde{q}\in\{\tilde{b},\tilde{g}\}} P(Q|\tilde{q})\pi(\tilde{q}|e_0)$, where the objective probabilities $P(Q|\tilde{q})$ follow from Bayes' rule.[11] The undetermined details of this belief system, $\pi(\tilde{q}|e_0)$, can be found in Section 4 as part of the general description of the model's equilibria.

3.2. **Payoff Functions.** In order to characterize the decision problem of each player, let us formulate the expected as well as the adjusted payoff function for each of them. Note that all payoff functions presented here ultimately depend on the choice tuple (e_0, λ, s, e_1), but for the sake of clarity it is omitted in the following descriptions. However, in order to stress the signal that the players act upon when maximizing their expected payoff, the corresponding signal is indicated via superscript (analogous to the corresponding player via subscript). Furthermore, for the welfare analysis, consider adjusted payoff functions that replace the subjective with objective beliefs. While those coincide with the expected payoff functions for the venture capitalist and for the bank, it is different from the perspective of the entrepreneur.

3.2.1. *Entrepreneur.* The entrepreneur bases the evaluation of the project on the potentially noisy and distorted signal $\tilde{q}$. The choice variables of the entrepreneur are the effort levels, e_0 and e_1, and the source of funding s - including the possibility of 'no project' - while always incurring the cost of $C(e)$ for each effort e. In conclusion, the entrepreneur maximizes the expected payoff given the available signal with respect to e_0, e_1, and s (note that the choice about the source of funding made by

[10]You could interpret this as the bank using a universal belief for these kinds of projects because of the relative (un-)importance of these credits for the overall business compared to other activities (specifically other markets). The universal belief then induces a universal interest rate.

[11]In detail, $P(Q|\tilde{q}) = \frac{\sum_{q\in\{b,g\}} P(\tilde{q}|q)P(q|Q)P(Q)}{\sum_{Q'\in\{B,G\}}\sum_{q\in\{b,g\}} P(\tilde{q}|q)P(q|Q')P(Q')}$.

the entrepreneur does not necessarily guarantee the corresponding outcome):

$$\Pi_E^{\tilde{q}} := \sum_{Q\in\{B,G\}} \pi_E(Q|\tilde{q})F_E(Q) - C(e_0) - C(e_1)$$

$$\text{with } F_E(Q) := \begin{cases} \max\{Y(Q,e_0,e_1,e_2(\lambda,s)) - rI, 0\} & s = B \\ (1-\lambda)Y(Q,e_0,e_1,e_2(\lambda,s)) & s = VC \wedge \lambda \in [0,1] \\ 0 & s = 0 \vee (s = VC \wedge \lambda = \emptyset) \end{cases}$$

The adjusted payoff of the entrepreneur given the available signal is then as follows:

$$\hat{\Pi}_E^{\tilde{q}} := \sum_{Q\in\{B,G\}} P(Q|\tilde{q})F_E(Q) - C(e_0) - C(e_1)$$

3.2.2. *Bank.* As the bank simply charges a fixed interest rate $\bar{r} + \bar{p}$ on the loan I while incurring the fixed cost of $\bar{r}I$, it does not require an objective function, i.e. expected payoff. Let us still state the adjusted payoff given the available signal:

$$\hat{\Pi}_B^{\tilde{q}} := \begin{cases} \sum_{Q\in\{B,G\}} P(Q|\tilde{q}) \min\{Y(Q,e_0,e_1,e_2(\lambda,s)), rI\} - \bar{r}I & s = B \\ 0 & s \neq B \end{cases}$$

3.2.3. *Venture Capitalist.* The venture capitalist reacts to the provided signal e_0 by setting the project share λ - potentially taking the outside option of 'no funding' - with matching determined effort $e_2(\lambda, s)$, while consequently incurring fixed cost of $\bar{r}I$ or zero and additional cost of $C(e_2(\lambda, s))$. In the end, the venture capitalist maximizes the following expected payoff given the initial effort with respect to λ:

$$\Pi_{VC}^{e_0} := \sum_{\tilde{q}\in\{\tilde{b},\tilde{g}\}} \pi(\tilde{q}|e_0)\Pi_{VC}^{\tilde{q}}$$

$$\text{where } \Pi_{VC}^{\tilde{q}} := \sum_{Q\in\{B,G\}} P(Q|\tilde{q})F_{VC}(Q) - C(e_2(\lambda,s))$$

$$\text{with } F_{VC}(Q) := \begin{cases} \lambda Y(Q,e_0,e_1,e_2(\lambda,s)) - \bar{r}I & \lambda \in [0,1] \wedge s = VC \\ 0 & \lambda = \emptyset \vee s \neq VC \end{cases}$$

$$\text{and } e_2(\lambda,s) = \begin{cases} \bar{e} & \lambda \in [0,1] \wedge s = VC \\ 0 & \lambda = \emptyset \vee s \neq VC \end{cases}$$

The adjusted payoff given the available signal follows via $\hat{\Pi}_{VC}^{\tilde{q}} = \Pi_{VC}^{\tilde{q}}$.

3.3. **Further Specifications.** Finally, let us further specify a couple of details about the model's functions, parameters, decisions, and equilibria.

3.3.1. *Functions.* The outcome of the project $Y(Q, E)$ is assumed to be of the form $\alpha(Q)E(e_0, e_1, e_2)$ with $E(e_0, e_1, e_2) = \beta_0 e_0 + \beta_1 e_1 + \beta_2 e_2$ for $\beta_i \in \mathbb{R}_{>0}$, $i = 1, 2, 3$, but with no specific values for $\alpha(Q)$. In other words, the outcome of the project scales multiplicatively with the quality level and additively with the effort levels (which are substitutes of each other). The cost of $C(e)$ for each effort e is assumed to be of the form $e^2/2$.

3.3.2. *Parameters.* Denote by ρ_q, $q \in \{g, b\}$, the probability of a quality conditioned on its corresponding signal and by $1 - \rho_q$ its complement, i.e. $\rho_q := P(Q|q)$ and $1 - \rho_q = P(Q^c|q)$ such that the upper case letter matches the lower case one.

Assume from now on that $\alpha(G) > \alpha(B) > 0$, i.e. the terms 'good' and 'bad' project are actually descriptive. Therefore, $Y(G, E) > Y(B, E) > 0$ for any $E > 0$. It is also necessary to assume that the noise (via its parameter ϵ) is not too extreme in its effect, i.e. $\epsilon < 1/2$. It will ensure that $\rho_g > 1-\rho_b$ (or $\rho_b > 1-\rho_g$). Moreover, with $\alpha(G) > \alpha(B)$ this will ensure $\sum_{Q\in\{B,G\}} P(Q|g)\alpha(Q) > \sum_{Q\in\{B,G\}} P(Q|b)\alpha(Q)$. Hence, a 'good' signal is actually good news and a 'bad' signal actually bad news.[12]

In the interest of a setting where both types of entrepreneurs and projects potentially exist, assume that $\mu \in (0, 1)$ resp. $\gamma \in (0, 1)$. Further, to avoid a trivial funding choice, assume that investment, risk-free interest rate, and risk premium are all non-zero, i.e. $I \neq 0$, $\bar{r} \neq 0$, resp. $\bar{p} \neq 0$. Finally, for a set-up where non-overconfidence distortions exist, assume $\epsilon \neq 0$.

3.3.3. *Decisions.* In order to always ensure properly defined equilibria, specifically for border cases, let us break possible ties in terms of (expected) profits as follows. For the source of funding the entrepreneur will always choose the venture capitalist, if it is part of the tie[13], or the bank, otherwise. Among different alternatives for effort the entrepreneur will always pick the lowest one out of those available. Concerning the terms of funding the venture capitalist will always prefer to fund the project with the highest share possible provided the project is feasible[14], or else not fund the project at all.

3.3.4. *Equilibria.* Finally, for studying the strategic interactions of this model, let us employ the concept of Perfect Bayesian Equilibrium, with a focus on pure strategies. Moreover, for the sake of a set-up where realistic and optimistic entrepreneurs potentially choose different strategies, let us limit our study to separating equilibria.

4. Equilibria

Let us now present an overview of the potential equilibria of our model, while the corresponding calculations can be found in Appendix C. Note that any equilibrium necessarily consists of two strategy tuples, one for each of the two perceived signals:

$$\sigma^* = \left(\sigma^*(\tilde{g}), \sigma^*(\tilde{b})\right)$$

where σ^* denotes the complete equilibrium tuple and $\sigma^*(\tilde{q})$ the individual ones. Now, each of the strategy tuples for the perceived signal comes from a list of nine potential strategy tuples that obviously depend on the perceived signal. Each of these nine strategy tuples falls into one of three categories with three elements each. Finally, each of those three categories corresponds to a different form of funding, namely via the bank, the venture capitalist, and the venture capitalist acting as a pseudo-bank. Note that the exact nature of the complete equilibrium tuple ultimately depends on the underlying parameter constellation.

[12] See Appendix B for the corresponding calculations.

[13] In the sense that the venture capitalist chose a project share in the unit interval.

[14] In the sense that the entrepreneur would choose the venture capitalist as the source of funding.

4.1. **Bank.** In terms of funding via the bank three potential strategy tuples exists - each of them corresponding to different expected levels of (re-)payment of the funds. In particular, in one scenario, B.3, the entrepreneur expects a full (re-)payment for the good and the bad project, while in another scenario, B.2, this is only the case for the good project, and in the other scenario, B.1, this is never the case.[15] Now, the elements of the three potential strategy tuples are listed in the following table:

	e_0	λ	s	e_1	e_2
B.1	0	$\emptyset$	B	0	0
B.2	$\pi_E(G\|\tilde{q})\alpha(G)\beta_0$	$\emptyset$	B	$\pi_E(G\|\tilde{q})\alpha(G)\beta_1$	0
B.3	$\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_0$	$\emptyset$	B	$\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	0

Note that the effort provided by the venture capitalist is actually listed in this table, even though it is technically not an element but a function of the strategy tuple. Technically, the respective perceived signal is also part of the strategy tuple.

The expected profit of the entrepreneur for each of these three strategy tuples ultimately determines (for a fixed parameter constellation) the dominant one:

$$\Pi_E^{\tilde{q}}(B.1) = 0$$
$$\Pi_E^{\tilde{q}}(B.2) = \frac{\pi_E(G|\tilde{q})^2\alpha(G)^2\left(\beta_0^2+\beta_1^2\right)}{2} - \pi_E(G|\tilde{q})rI$$
$$\Pi_E^{\tilde{q}}(B.3) = \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right)}{2} - rI$$

For all of the three alternatives the venture capitalist expects a profit of $\Pi_{VC}^{\tilde{q}}(B.\cdot) = 0$. On the other hand, the payoff for the bank naturally differs, but without any influence on the process that determines the equilibrium due to the deterministic nature of the bank. Thus, let us postpone the details on the perspective of the bank for now.[16]

It is also important to stress the fact that due to our rules on tie-breaking as well as due to the deterministic nature of the bank, the alternative B.1 essentially contains what would otherwise be the outside option of 'no funding'. It falls into the category of funding via the bank, but with all effort levels equal to zero it functions like 'no funding' (for the entrepreneur), which is important to keep in mind when analyzing the strategy tuples that emerge.

Finally, in contrast to the other (yet to be described) potential strategy tuples, these alternatives emerge whenever the corresponding expected profit dominates amongst the alternatives for the entrepreneur without any additional requirements for their existence (see Appendix C for the details).

4.2. **Venture Capitalist.** With respect to funding via the venture capitalist also three potential strategy tuples exists. Each of them corresponds to a different set-up of constraints limiting the optimal choices. In one scenario both the entrepreneur and the venture capitalist simply pick their respective optimal choices without any constraint being binding, referred to as VC.k. As special cases of this scenario, two other alternatives emerge. First, it is possible that mathematically the entrepreneur prefers a negative initial effort (in order to limit the project share). In that case, the technical limitation of the zero bound implies that the entrepreneur potentially

[15]Alternatively, the three scenarios correspond to different expected levels of profit, where the entrepreneur expects zero profit for both qualities (B.1), positive profit for both qualities (B.3), or a mixture of the two (B.2).

[16]The payoff for the bank necessarily plays a role in the welfare analysis later on.

chooses zero initial effort, referred to as VC.0. Second, in the case where the participation of the venture capitalist is relevant, the entrepreneur potentially chooses the appropriate participation-guaranteeing initial effort, referred to as VC.p. Note that all of these three strategy tuples depend on the participation constraint of the entrepreneur with respect to the bank not being binding. The elements of these three scenarios are listed in the following table (with some additional notations):

	e_0	λ	s	e_1	e_2
VC.0	0	λ_0	VC	$(1-\lambda_0)\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$
VC.p	e_p	λ_p	VC	$(1-\lambda_p)\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$
VC.k	e_k	λ_k	VC	$(1-\lambda_k)\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$

The initial effort levels of e_p and e_k are defined by the following two expressions

$$e_p = \frac{1}{\beta_0}\left(-\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - \beta_2\bar{e} + \beta_1\sqrt{2\frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})}{\mathbb{E}_P(\alpha|\tilde{q})}(\bar{e}^2+2\bar{r}I)}\right)$$

$$e_k = \beta_0\frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - 3\beta_2\bar{e}}{3\beta_0^2+4\beta_1^2}$$

and the three project shares λ_0, λ_p, and λ_k are given by the following three terms

$$\lambda_0 = \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + \beta_2\bar{e}}{2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}$$

$$\lambda_p = \sqrt{\frac{\bar{e}^2+2\bar{r}I}{2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}}$$

$$\lambda_k = \frac{2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+\beta_2\bar{e}\right)}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(3\beta_0^2+4\beta_1^2\right)}$$

which in turn imply the following three corresponding additional effort levels:

$$(1-\lambda_0)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 = \frac{1}{2\beta_1}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - \beta_2\bar{e}\right)$$

$$(1-\lambda_p)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 = \left(1-\sqrt{\frac{\bar{e}^2+2\bar{r}I}{2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}}\right)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1$$

$$(1-\lambda_k)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 = \beta_1\frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+2\beta_1^2\right)-2\beta_2\bar{e}}{3\beta_0^2+4\beta_1^2}$$

As with the bank scenario, the deciding factor for the dominant strategy tuple is the expected profit of the entrepreneur for each of those three scenarios:

$$\Pi_E^{\tilde{q}}(VC.0) = \frac{1}{8\beta_1^2}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2+2\beta_2\bar{e}\right)-3\beta_2^2\bar{e}^2\right)$$

$$\Pi_E^{\tilde{q}}(VC.p) = \frac{1}{4\mathbb{E}_P(\alpha|\tilde{q})\beta_0^2}\left(\Pi_{E,p_+}^{\tilde{q}} - \Pi_{E,p_-}^{\tilde{q}}\right)$$

$$\Pi_{E,p_+}^{\tilde{q}} = \sqrt{2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\bar{e}^2+2\bar{r}I\right)}\left(4\beta_1\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+\beta_2\bar{e}\right)\right)$$

$$\Pi_{E,p_-}^{\tilde{q}} = 2\mathbb{E}_P(\alpha|\tilde{q})\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+2\beta_2\bar{e}\right)+\beta_2^2\bar{e}^2\right) + \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(3\beta_0^2+4\beta_1^2\right)\left(\bar{e}^2+2\bar{r}I\right)$$

$$\Pi_E^{\tilde{q}}(VC.k) = \frac{1}{2\left(3\beta_0^2+4\beta_1^2\right)}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+2\beta_2\bar{e}\right)-3\beta_2^2\bar{e}^2\right)$$

Meanwhile, the respective expected profit of the venture capitalist is given as follows:

$$\Pi^{\tilde{q}}_{VC}(VC.0) = \frac{\mathbb{E}_P(\alpha|\tilde{q})}{4\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + 2\beta_2\bar{e}\right) + \beta_2^2\bar{e}^2\right) - \frac{\bar{e}^2}{2} - \bar{r}I$$

$$\Pi^{\tilde{q}}_{VC}(VC.p) = 0$$

$$\Pi^{\tilde{q}}_{VC}(VC.k) = \frac{4\mathbb{E}_P(\alpha|\tilde{q})\beta_1^2}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(3\beta_0^2 + 4\beta_1^2\right)^2}\Pi^{\tilde{q}}_{VC,k} - \frac{\bar{e}^2}{2} - \bar{r}I$$

$$\Pi^{\tilde{q}}_{VC,k} = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2 + \beta_1^2\right)\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2 + \beta_1^2\right) + 2\beta_2\bar{e}\right) + \beta_2^2\bar{e}^2$$

Note that the expected profit of zero for the venture capitalist in case of a binding participation constraint is due to the outside option of 'no project', which yields zero return, and our rules on tie-breaking.

Finally, the emergence of these alternatives depends not only on the comparison of the expected profit of the entrepreneur but also on the following conditions

$$\begin{aligned}
&\text{VC.0: } e_p \leq 0 < \xi_\lambda \quad \text{and } e_k < 0\\
&\text{VC.p: } 0 < e_p < \xi_\lambda \quad \text{and } e_k < e_p\\
&\text{VC.k: } e_p \leq e_k < \xi_\lambda \text{ and } 0 \leq e_k
\end{aligned}$$

where $\xi_\lambda = \left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - \beta_2\bar{e}\right)/\beta_0$ marks the critical value between partial and full transfer of the project payoff from the entrepreneur to the venture capitalist with respect to the initial effort, i.e. any initial effort equal to or above it implies that the venture capitalist prefers a project share equal to or above one.[17] Note that these conditions do not constitute a complete partition of the parameter space, instead any configuration violating all three conditions simply yields a strategy tuple in one of the other categories - if there exists an equilibrium for these values at all.

4.3. **Pseudo-Bank.** Essentially corresponding one-to-one to the bank scenarios there exist three pseudo-bank scenarios where the bank as an outside option motivates the venture capitalist to act as a pseudo-bank. In these cases, from the perspective of the entrepreneur the contract equals the corresponding bank scenario, but naturally differs for the venture capitalist. Due to the equivalence to the bank scenarios for the entrepreneur, refer to these three scenarios as VC.B.1, VC.B.2, resp. VC.B.3. Note that all of these three strategy tuples depend on the participation constraint of the entrepreneur with respect to the bank being binding, in other words it depends on $\Pi^{\tilde{q}}_E(B.*) > \Pi^{\tilde{q}}_E(VC.*)$, where $B.*$ resp. $VC.*$ refer to the optimal scenario in terms of funding via the bank resp. the venture capitalist.[18] The elements of these three scenarios are listed in the following table (with some additional notations):

	e_0	λ	s	e_1	e_2
VC.B.1	0	$\lambda_{B.1}$	VC	0	$\bar{e}$
VC.B.2	$\pi_E(G\|\tilde{q})\alpha(G)\beta_0$	$\lambda_{B.2}$	VC	$(1-\lambda_{B.2})\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$
VC.B.3	$\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_0$	$\lambda_{B.3}$	VC	$(1-\lambda_{B.3})\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$

[17] The nature of the critical value is ultimately responsible for the fact that the upper limit in contrast to the lower limit does not appear as a potential optimal choice. As soon as the upper limit of the interval becomes relevant, the profit for any alternative still within the interval approaches the profit of the border value from below (without reaching it), which itself equals that of the case of 'no funding'.

[18] In case the category of funding via the venture capitalist provided no potential strategy tuple, treat the optimal scenario of it like the 'no funding'-alternative.

The corresponding project shares are given by the following three expressions

$$\lambda_{B.1} = 1$$
$$\lambda_{B.2} = \frac{1}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}\left(\lambda_{B.2}^{+} - \lambda_{B.2}^{-}\right)$$
$$\lambda_{B.3} = \frac{1}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}\left(\lambda_{B.3}^{+} - \lambda_{B.3}^{-}\right)$$

with four additional auxiliary terms defined in the following (similar) manner:

$$\lambda_{B.2}^{+} = \pi_E(G|\tilde{q})\alpha(G)\beta_0^2 + \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + \beta_2\bar{e}$$
$$\lambda_{B.2}^{-} = \sqrt{\pi_E(G|\tilde{q})^2\alpha(G)^2\left(\beta_0^2+\beta_1^2\right)^2 + \beta_2^2\bar{e}^2 + 2\pi_E(G|\tilde{q})\alpha(G)\beta_0^2\beta_2\bar{e} - 2\pi_E(G|\tilde{q})\beta_1^2 rI}$$
$$\lambda_{B.3}^{+} = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right) + \beta_2\bar{e}$$
$$\lambda_{B.3}^{-} = \sqrt{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right)^2 + \beta_2^2\bar{e}^2 + 2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0^2\beta_2\bar{e} - 2\beta_1^2 rI}$$

Consequently, $(1-\lambda_{B.2})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 - (\lambda_{B.2}^{+} - \lambda_{B.2}^{-})/\beta_1$ as well as $(1-\lambda_{B.3})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 - (\lambda_{B.3}^{+} - \lambda_{B.3}^{-})/\beta_1$ represent the corresponding additional effort levels. Thus, while the initial effort corresponds 1-to-1 to that of the respective bank scenario, the additional effort clearly falls below its counterpart - except for the case of VC.B.1.

As mentioned, the expected profit of the entrepreneur matches the bank scenarios:

$$\Pi_E^{\tilde{q}}(VC.B.1) = 0$$
$$\Pi_E^{\tilde{q}}(VC.B.2) = \frac{\pi_E(G|\tilde{q})^2\alpha(G)^2\left(\beta_0^2+\beta_1^2\right)}{2} - \pi_E(G|\tilde{q})rI$$
$$\Pi_E^{\tilde{q}}(VC.B.3) = \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right)}{2} - rI$$

Obviously, the venture capitalist expects a different profit than in the bank scenarios:

$$\Pi_{VC}^{\tilde{q}}(VC.B.1) = \mathbb{E}_P(\alpha|\tilde{q})\beta_2\bar{e} - \frac{\bar{e}^2}{2} - \bar{r}I$$
$$\Pi_{VC}^{\tilde{q}}(VC.B.2) = \lambda_{B.2}\mathbb{E}_P(\alpha|\tilde{q})\lambda_{B.2}^{-} - \frac{\bar{e}^2}{2} - \bar{r}I$$
$$\Pi_{VC}^{\tilde{q}}(VC.B.3) = \lambda_{B.3}\mathbb{E}_P(\alpha|\tilde{q})\lambda_{B.3}^{-} - \frac{\bar{e}^2}{2} - \bar{r}I$$

It is noteworthy to point out that the first pseudo-bank scenario corresponds to a case where the venture capitalist essentially takes over the project completely without any compensation for the entrepreneur. However, it is a voluntary transfer as the entrepreneur ultimately makes the choice. Its occurrence depends on our rules on tie-breaking.

Finally, the existence of these alternatives depends not only on the relation of the expected profit of the entrepreneur but also on the following conditions

$$\text{VC.B.1: } \xi_\lambda \leq 0 \text{ and } \xi_{VC} \leq 0$$
$$\text{VC.B.2: } \lambda_{VC} \leq \lambda_{B.2}$$
$$\text{VC.B.3: } \lambda_{VC} \leq \lambda_{B.3}$$

where $\xi_{VC} = \left(-\beta_2\bar{e} + \left(\bar{e}^2 + 2\bar{r}I\right)/\left(2\mathbb{E}_P(\alpha|\tilde{q})\right)\right)/\beta_0$ marks the critical value for non-negative expected profit for the venture capitalist, when taking over the project, with

respect to the initial effort.[19] Furthermore, the term $\lambda_{VC} = \iota_{B.} - \sqrt{\iota_{B.}^2 - \psi}$ with $\psi = \left(\bar{e}^2 + 2\bar{r}I\right) / \left(2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\right)$ and $\iota_{B.} = \lambda_{B.}^+ / \left(2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\right)$ marks a similar critical value (in purpose) with respect to the project share.[20]

4.4. **Overall Optimality and Beliefs.** Let us now present the process by which the overall optimum gets determined and specify the corresponding belief system.

In order to formulate the overall optimum, use $B.*$ for the optimal bank scenario, $VC.*$ for the optimal not-bank-equivalent venture capitalist scenario, and also $VC.B.*$ for the optimal bank-equivalent venture capitalist scenario. Finally, with the corresponding conditions already determined, it is an intuitive two-step process. First, if $\Pi_E^{\tilde{q}}(VC.*) \geq \Pi_E^{\tilde{q}}(B.*)$, then $VC.*$ is simply the overall optimum. Otherwise, provided its requirements are met, $VC.B.*$ is the overall optimum, else, it is $B.*$.[21]

Lastly, while the focus on separating equilibria immediately fixes the belief system of the venture capitalist in equilibrium, it actually requires additional attention when not in equilibrium. Let us rectify this omission by specifying the belief system of the venture capitalist as a combination of indicator functions (with a singularity)

$$\pi(\tilde{q}|e_0) = \begin{cases} \mathbb{1}_{\{\tilde{q}=\tilde{d}\}}(\tilde{q}) & e_0 = e_0(\tilde{d}) \\ \mathbb{1}_{\{\tilde{q}\neq\tilde{d}\}}(\tilde{q}) & e_0 \neq e_0(\tilde{d}) \end{cases}$$

where $\tilde{d}$ refers to the specific signal which induces the more 'desirable' project share. Together with the non-pooling condition for the separating equilibrium given by $\Pi_E^{\tilde{q}}(e_0^*(\tilde{d})) < \Pi_E^{\tilde{q}}(e_0^*(\tilde{q}))$ for $\tilde{q} \neq \tilde{d}$ (abstracting from the remaining strategy tuple), this definition of the belief system prevents any deviation from the optimal choices, as the expected profit of the entrepreneur decreases with increasing project share and the bank's terms of funding are independent of the effort by the entrepreneur.

As a concluding remark, let us emphasize that it is a priori not clear whether there actually exists a separating equilibrium. In fact, as part of the analysis reveals, there definitely appear significant parts of the parameter space without the emergence of separating equilibria. However, in case there exists one, it is described by (a pair of) the previously presented strategy tuples.

5. Analysis

In the following, let us analyze the previously determined scenarios by studying various comparative static effects of the model parameters. In the first part, the focus lies on the impact with respect to the player's choices and their profits - providing general insights into the model dynamics. In the final part, the focus shifts towards the influence on the comparative static effect of overconfidence on the social welfare. In other words, this part determines the sign of the cross-derivatives and ultimately characterizes different constellations of the model parameters under which an increase in the share of overconfidence leads to an increase in social welfare. Between those two parts, let us properly define social welfare (for the final part) and then re-visit each of the potential scenarios presented before in order to determine

[19] As the presented relation of ξ_{VC} concerns zero, it is possible to re-formulate it as follows in order to stress the origin of this critical value: $\mathbb{E}_P(\alpha|\tilde{q})\beta_2\bar{e} - \frac{\bar{e}^2}{2} - \bar{r}I \geq 0$

[20] While it is not obvious through the notation, the critical value for the project share depends on the specific bank scenario under consideration.

[21] With the understanding that in case there exists no venture capitalist (pseudo-bank) scenario, then the category is automatically eliminated from the comparison.

the respective levels of social welfare. In addition, let us examine the choices of an unbiased hypothetical social planner in our model in order to understand the possible distortion effects of overconfidence.

5.1. **Model Dynamics.** As a step towards an intuition about the model dynamics present in the previously determined scenarios, let us study comparative static effects of the model parameters on optimal choice variables and expected profit functions. To keep it manageable, let us focus on the bank scenario with full re-payment, B.3, and the venture capitalist scenario unaffected by any of the constraints, VC.k.[22] Additionally, the analysis is restricted to unambiguous statements to further limit their numbers. The actual calculations for all of the comparative static effects can be found in Appendix D.

Proposition 1. *In the bank scenario with full re-payment, the*

(1) initial and additional effort e_0 and e_1 are
  (A) increasing in their respective effort productivities β_0 and β_1,
  (B) increasing in the expected project productivity $\mathbb{E}_{\pi_E}(\alpha|\tilde{q})$;
(2) expected profit of the entrepreneur Π_E is
  (A) increasing in the effort productivities β_0 and β_1,
  (B) increasing in the expected project productivity $\mathbb{E}_{\pi_E}(\alpha|\tilde{q})$,
  (C) decreasing in the interest rate of the bank r,
  (D) decreasing in the cost of the investment I.

At this point, it is important to remember the status of the bank as an additional outside option. As a consequence, the timing of the different forms of effort provided by the entrepreneur plays no role and there exists no strategic interaction between those efforts and the terms of funding. Thus, the productivities of effort and project positively influence the (respective) efforts and the (overall) project payoff without any adverse impact. Unsurprisingly, the two terms connected to the cost of funding negatively affect the profit of the entrepreneur.

In contrast to the bank, funding via the venture capitalist implies a relevance of the timing of the different forms of effort and introduces related strategic interactions. In particular, as about to be presented in the proposition about the project share, this potentially leads to a hold-up problem.

Proposition 2. *In the venture capitalist scenario without any constraints, the*

(1) initial and additional effort e_0 and e_1 are
  (A) decreasing in the effort productivity β_2,
  (B) decreasing in the effort provided by the venture capitalist $\bar{e}$,
  (C) increasing in the expected project productivity $\mathbb{E}_{\pi_E}(\alpha|\tilde{q})$,
  (D) less than they are in the bank scenario with full re-payment;
(2) additional effort e_1 is
  (A) increasing in the effort productivity β_1,
  (B) decreasing in the project share λ;
(3) expected profit of the entrepreneur Π_E is
  (A) increasing in the expected project productivity $\mathbb{E}_{\pi_E}(\alpha|\tilde{q})$;
(4) expected profit of the venture capitalist Π_{VC} is
  (A) decreasing in the risk-free interest rate $\bar{r}$,
  (B) decreasing in the cost of the investment I.

[22] Although a number of these statements follow for the other scenarios as well.

It is not surprising that with the venture capitalist, the influence of the parameters related to the cost of funding shift from the entrepreneur to the venture capitalist (but with the risk-free interest rate instead of the bank one). Also, the impact of the productivity of the project remains positive both on the efforts and the profit. However, due to the aforementioned strategic interactions, the positive influence of the effort productivity on the respective efforts only remains for the additional effort and its unambiguous nature vanishes for the impact on the overall project payoff.[23] Furthermore, the bargaining process, in the form of the project share, reduces the additional effort by the entrepreneur - both in this venture capitalist scenario and compared to the bank scenario with full re-payment. On top of that, the contribution of the venture capitalist ultimately reduces that of the entrepreneur, which is a development that is in part driven by the relationship between the contribution of the venture capitalist and the project share.

Proposition 3. *In the venture capitalist scenario without any constraints, the project share is*

(A) increasing in the effort productivity β_0,
(B) decreasing in the effort productivity β_1,
(C) increasing in the effort productivity β_2,
(D) increasing in the effort provided by the venture capitalist $\bar{e}$,
(E) decreasing in the expected project productivity $\mathbb{E}_{\pi_E}(\alpha|\tilde{q})$,
(F) increasing in the initial effort e_0,
(G) always at least $1/2$.

As indicated before, the contribution by the venture capitalist always increases the project share (and therefore reduces the additional effort by the entrepreneur). While the project share decreases with an increasing productivity of the project and the additional effort as expected, it increases with an increasing initial effort or via its productivity. This consequence of the strategic interaction between the entrepreneur and the venture capitalist captures the hold-up problem present in the framework. It showcases the mechanism by which the bargaining process potentially leads to an underinvestment of initial effort. Additionally, this illustrates the substitutability of the initial and the additional effort of the entrepreneur, in the way that an increasing initial effort leads to an increasing project share which then implies a decreasing additional effort. Finally, due to the bargaining power of the venture capitalist, who presents the entrepreneur with a take-it-or-leave-it offer, the project share is always at least $1/2$.

As a prepration for the coming analysis of social welfare, let us take a look at the impact of the share of realists on the involved probabilities and expectations. Also, let us compile a couple of statements on the relation between the biased and the unbiased expected productivity. Among the probabilities under consideration are the chances for a specific signal to occur, i.e. $P(\tilde{q})$ for $\tilde{q} \in \{\tilde{g}, \tilde{b}\}$, which follow from the probabilistic process presented in Figure 2.[24]

[23] Note that technically the proposition only provides statements about impacts that exist not about the lack of impacts. However, it is a straight forward calculation to show that for example the positive influence of the initial effort productivity on the initial effort depends on $\sqrt{3}\beta_0 < 2\beta_1$.

[24] In other words:

$$P(\tilde{g}) = \gamma\left((1-\epsilon) + \epsilon\,(1-\mu)\right) + (1-\gamma)\left(\epsilon + (1-\epsilon)\,(1-\mu)\right)$$
$$P(\tilde{b}) = \gamma\epsilon\mu + (1-\gamma)\,(1-\epsilon)\,\mu$$

Proposition 4. *The impact of an increasing share of realists μ is*

(A) positive on the probability for a good project given a good signal $P(G|\tilde{g})$ (negative on $P(B|\tilde{g})$),
(B) non-existent on the probability for a good project given a bad signal $P(G|\tilde{b})$ (also zero on $P(B|\tilde{b})$),
(C) positive on the unbiased project productivity given a good signal $\mathbb{E}_P(\alpha|\tilde{g})$,
(D) non-existent on the unbiased project productivity given a bad signal $\mathbb{E}_P(\alpha|\tilde{b})$,
(E) negative on the probability of a good signal $P(\tilde{g})$ (positive on $P(\tilde{b})$).

Furthermore, the biased expected project productivity is

(1) higher with a good signal than with a bad signal $\mathbb{E}_{\pi_E}(\alpha|\tilde{g}) > \mathbb{E}_{\pi_E}(\alpha|\tilde{b})$,
(2) higher than the unbiased one for a good signal $\mathbb{E}_{\pi_E}(\alpha|\tilde{g}) > \mathbb{E}_P(\alpha|\tilde{g})$,
(3) equal to the unbiased one for a bad signal $\mathbb{E}_{\pi_E}(\alpha|\tilde{b}) = \mathbb{E}_P(\alpha|\tilde{b})$.

Following intuition, a boost in realism increases the weight behind a good signal while having no similar influence for the case of a bad signal. An important factor here is the limit of 1/2 on the noise ϵ, which ensures a certain level of information present in the signals. It is also not surprising that an increase in the share of realists decreases the chance of a good signal in general, due to the different 'methods' of interpreting the signal. Finally, as expected, the entrepreneur overestimates the expected project productivity in the case of a good signal while correctly estimating it in the case of a bad signal, and a 'better' signal induces a similar expectation.

Note that the lack of analysis with respect to the impact of the share of realists on the perspective of the entrepreneur is due to the entrepreneur treating the situation as a realist (independent of the type). Thus, the share of realists plays no role in the probabilities, expectations, and choices of the entrepreneur.

5.2. **Social Welfare.** In order to analyze the impact of overconfident entrepreneurs on the social welfare, it is important to properly define social welfare in our model. As part of the pre-game, nature determines the signal transmitted to the entrepreneur. Each of the two cases yields a different strategy tuple, which then implies that the social welfare W takes the form of a weighted sum $\sum_{\tilde{q}\in\{\tilde{b},\tilde{g}\}} P(\tilde{q})W^{\tilde{q}}$ of case-specific welfare levels $W^{\tilde{q}}$.[25] Driven by our interest in all three involved parties, these $W^{\tilde{q}}$ enter as a sum $\hat{\Pi}^{\tilde{q}}_E + \hat{\Pi}^{\tilde{q}}_B + \hat{\Pi}^{\tilde{q}}_{VC}$ of all three adjusted expected profits $\hat{\Pi}^{\tilde{q}}$. Recall, that the adjusted expected profits simply rectify the error in judgment of the overconfident entrepreneurs with respect to the probability for the project quality given a signal.

Let us now re-visit each of the potential scenarios presented before in order to determine the respective levels of social welfare. Afterward, consider the perspective of an unbiased hypothetical social planner to examine the possible distortion effects of overconfidence.

5.2.1. *Bank.* Recall, that the bank scenarios correspond to different expected levels of (re-)payment of the funds from the perspective of the entrepreneur. In particular, B.3 implies a full (re-)payment, B.2 a partial one, B.1 none at all. Now, adjusting

[25]Where the $P(\tilde{q})$, $\tilde{q} \in \{\tilde{g}, \tilde{b}\}$, follow from the probabilistic process presented in Figure 2 again.

the expected profit of the entrepreneur with respect to the bias yields the following:

$$\hat{\Pi}_E^{\tilde{q}}(B.1) = 0$$
$$\hat{\Pi}_E^{\tilde{q}}(B.2) = \left(P(G|\tilde{q})\alpha(G) - \frac{\pi_E(G|\tilde{q})\alpha(G)}{2}\right)\pi_E(G|\tilde{q})\alpha(G)\left(\beta_0^2+\beta_1^2\right) - P(G|\tilde{q})rI$$
$$\hat{\Pi}_E^{\tilde{q}}(B.3) = \left(\mathbb{E}_P(\alpha|\tilde{q}) - \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})}{2}\right)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right) - rI$$

It follows immediately that the venture capitalist expects $\hat{\Pi}_{VC}^{\tilde{q}} = \Pi_{VC}^{\tilde{q}} = 0$ for all of the three constellations. Meanwhile, the adjusted expected profit of the bank yields:

$$\hat{\Pi}_B^{\tilde{q}}(B.1) = -\bar{r}I$$
$$\hat{\Pi}_B^{\tilde{q}}(B.2) = P(B|\tilde{q})\alpha(B)\pi_E(G|\tilde{q})\alpha(G)\left(\beta_0^2+\beta_1^2\right) + \left(P(G|\tilde{q})\bar{p} - P(B|\tilde{q})\bar{r}\right)I$$
$$\hat{\Pi}_B^{\tilde{q}}(B.3) = \bar{p}I$$

Note that similar to the venture capitalist the bank needs no adjustment, but remember that we had postponed the details on the perspective of the bank before.

Thus, the case-specific welfare level for each of the three scenarios is given by:

$$W^{\tilde{q}}(B.1) = -\bar{r}I$$
$$W^{\tilde{q}}(B.2) = \left(\mathbb{E}_P(\alpha|\tilde{q}) - \frac{\pi_E(G|\tilde{q})\alpha(G)}{2}\right)\pi_E(G|\tilde{q})\alpha(G)\left(\beta_0^2+\beta_1^2\right) - \bar{r}I$$
$$W^{\tilde{q}}(B.3) = \left(\mathbb{E}_P(\alpha|\tilde{q}) - \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})}{2}\right)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right) - \bar{r}I$$

5.2.2. *Venture Capitalist.* Recall, that the venture capitalist scenarios correspond to different set-ups of constraints limiting the optimal choices. While VC.k coincides with no binding restriction, the limit of zero restricts the optimal choice in VC.0 and the participation constraint of the venture capitalist matters for VC.p. Introduce $\tau_{\tilde{q}} := \mathbb{E}_P(\alpha|\tilde{q})/\mathbb{E}_{\pi_E}(\alpha|\tilde{q})$ in order to make the the adjusted expected profit clearer.[26] Now, adjusting the expected profit of the entrepreneur yields the following terms:

$$\hat{\Pi}_E^{\tilde{q}}(VC.0) = \frac{1}{8\beta_1^2}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\left(2\mathbb{E}_P(\alpha|\tilde{q}) - \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\right)\beta_1^2 + 2\beta_2\bar{e}\right) - \left(2\tau_{\tilde{q}}+1\right)\beta_2^2\bar{e}^2\right)$$
$$\hat{\Pi}_E^{\tilde{q}}(VC.p) = \frac{1}{4\mathbb{E}_P(\alpha|\tilde{q})\beta_0^2}\left(\hat{\Pi}_{E,p_+}^{\tilde{q}} - \hat{\Pi}_{E,p_-}^{\tilde{q}}\right)$$
$$\hat{\Pi}_{E,p_+}^{\tilde{q}} := \sqrt{2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\bar{e}^2+2\bar{r}I\right)}\big(4\beta_1\left(\mathbb{E}_P(\alpha|\tilde{q})\beta_0^2 + \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + \beta_2\bar{e}\right)$$
$$+ \left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) - \mathbb{E}_P(\alpha|\tilde{q})\right)2\beta_0^2\beta_1\big)$$
$$\hat{\Pi}_{E,p_-}^{\tilde{q}} := 2\mathbb{E}_P(\alpha|\tilde{q})\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right) + 2\beta_2\bar{e}\right) + \beta_2^2\bar{e}^2\right)$$
$$+ \left(\left(2\mathbb{E}_P(\alpha|\tilde{q}) + \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\right)\beta_0^2 + 4\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\right)\left(\bar{e}^2+2\bar{r}I\right)$$

[26] In a sense $\tau(\tilde{q})$ functions as the conversion factor between the expected project productivity of the entrepreneur and the venture capitalist

$$\begin{aligned}
\hat{\Pi}_E^{\tilde{q}}(VC.k) &= \frac{1}{2\left(3\beta_0^2+4\beta_1^2\right)^2}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\hat{\Pi}_{E,k_1}^{\tilde{q}} - 3\beta_2^2\bar{e}^2\hat{\Pi}_{E,k_2}^{\tilde{q}}\right)\\
\hat{\Pi}_{E,k_1}^{\tilde{q}} &:= \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\left(4\tau(\tilde{q})-1\right)\beta_0^4 + \left(12\tau(\tilde{q})-5\right)\beta_0^2\beta_1^2 + \left(8\tau(\tilde{q})-4\right)\beta_1^4\right)\\
&\quad + 2\beta_2\bar{e}\left(\left(5-2\tau(\tilde{q})\right)\beta_0^2 + 4\beta_1^2\right)\\
\hat{\Pi}_{E,k_2}^{\tilde{q}} &:= 3\beta_0^2 + \frac{1}{3}\left(8\tau(\tilde{q})+4\right)\beta_1^2
\end{aligned}$$

As before, it follows that $\hat{\Pi}_{VC}^{\tilde{q}} = \Pi_{VC}^{\tilde{q}}$ (with the previously provided expressions). Moreover, it holds that the bank expects the corresponding adjusted profit of $\hat{\Pi}_B^{\tilde{q}} = 0$ for all of the three constellations.

Thus, the case-specific welfare level for each of the three scenarios is given by:

$$\begin{aligned}
W^{\tilde{q}}(VC.0) &= \frac{1}{8\beta_1^2}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 W_{0_1}^{\tilde{q}} + W_{0_2}^{\tilde{q}}\right)\\
W_{0_1}^{\tilde{q}} &:= \left(4\mathbb{E}_P(\alpha|\tilde{q}) - \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\right)\beta_1^2 + \left(2\tau(\tilde{q})+1\right)2\beta_2\bar{e}\\
W_{0_2}^{\tilde{q}} &:= -\left(4\beta_1^2+\beta_2^2\right)\bar{e}^2 - 8\beta_1^2\bar{r}I\\
W^{\tilde{q}}(VC.p) &= \frac{1}{4\mathbb{E}_P(\alpha|\tilde{q})\beta_0^2}\left(\hat{\Pi}_{E,p_+}^{\tilde{q}} - \hat{\Pi}_{E,p_-}^{\tilde{q}}\right)\\
W^{\tilde{q}}(VC.k) &= \frac{1}{2\left(3\beta_0^2+4\beta_1^2\right)^2}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 W_{k_1}^{\tilde{q}} + W_{k_2}^{\tilde{q}}\right)\\
W_{k_1}^{\tilde{q}} &:= \left(12\mathbb{E}_P(\alpha|\tilde{q}) - \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\right)\beta_0^4 + \left(28\mathbb{E}_P(\alpha|\tilde{q}) - 5\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\right)\beta_0^2\beta_1^2\\
&\quad + \left(16\mathbb{E}_P(\alpha|\tilde{q}) - 4\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\right)\beta_1^4 + \left(\left(5+6\tau(\tilde{q})\right)\beta_0^2 + \left(4+8\tau(\tilde{q})\right)\beta_1^2\right)2\beta_2\bar{e}\\
W_{k_2}^{\tilde{q}} &:= -\bar{e}^2\left(\left(9\beta_0^2+4\beta_1^2\right)\beta_2^2 + \left(3\beta_0^2+4\beta_1^2\right)^2\right) - 2\left(3\beta_0^2+4\beta_1^2\right)^2\bar{r}I
\end{aligned}$$

5.2.3. *Pseudo-Bank.* Recall, that the pseudo-bank scenarios relate one-to-one to the bank scenarios where the bank as an outside option motivates the venture capitalist to act as a pseudo-bank. Even though the expected profit of the entrepreneur coincides with the respective bank scenario, this differs for the adjusted terms:

$$\begin{aligned}
\hat{\Pi}_E^{\tilde{q}}(VC.B.1) &= 0\\
\hat{\Pi}_E^{\tilde{q}}(VC.B.2) &= \frac{1}{2\beta_1^2}\Big(\left(2\tau(\tilde{q})-1\right)\left(\lambda_{B.2}^-\right)^2 + 2\left(1-\tau(\tilde{q})\right)\left(\lambda_{B.2}^+ - \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\right)\lambda_{B.2}^-\\
&\quad - \pi_E(G|\tilde{q})^2\alpha(G)^2\left(\beta_0^2\beta_1^2+\beta_1^4\right) - 2\pi_E(G|\tilde{q})\alpha(G)\beta_0^2\beta_2\bar{e} - \beta_2^2\bar{e}^2\Big)\\
\hat{\Pi}_E^{\tilde{q}}(VC.B.3) &= \frac{1}{2\beta_1^2}\Big(\left(2\tau(\tilde{q})-1\right)\left(\lambda_{B.3}^-\right)^2 + 2\left(1-\tau(\tilde{q})\right)\left(\lambda_{B.3}^+ - \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\right)\lambda_{B.3}^-\\
&\quad - \mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\left(\beta_0^2\beta_1^2+\beta_1^4\right) - 2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0^2\beta_2\bar{e} - \beta_2^2\bar{e}^2\Big)
\end{aligned}$$

Again, it follows that $\hat{\Pi}_{VC}^{\tilde{q}} = \Pi_{VC}^{\tilde{q}}$. Additionally, it holds that the bank expects the corresponding adjusted profit of $\hat{\Pi}_B^{\tilde{q}} = 0$ for all of the three constellations.

Thus, the case-specific welfare level for each of the three scenarios is given by:

$$W^{\tilde{q}}(VC.B.1) = \mathbb{E}_P(\alpha|\tilde{q})\beta_2\bar{e} - \frac{\bar{e}^2}{2} - \bar{r}I$$
$$W^{\tilde{q}}(VC.B.2) = \frac{1}{2\beta_1^2}\left(\kappa_{B.2} - \pi_E(G|\tilde{q})^2\alpha(G)^2\kappa_\beta - 4\pi_E(G|\tilde{q})\alpha(G)\beta_0^2\beta_2\bar{e} + 2\beta_1^2\bar{\kappa}I\right)$$
$$W^{\tilde{q}}(VC.B.3) = \frac{1}{2\beta_1^2}\left(\kappa_{B.3} - \mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\kappa_\beta - 4\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0^2\beta_2\bar{e} + 2\beta_1^2\bar{p}I\right)$$

with the help of $\kappa_{B.\cdot} := 2\lambda^-_{B.\cdot}\left(\lambda^+_{B.\cdot} + \left(\mathbb{E}_P(\alpha|\tilde{q}) - \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\right)\beta_1^2\right) - \bar{e}^2\left(\beta_1^2 + 2\beta_2^2\right)$ and $\kappa_\beta := 2\beta_0^4 + 3\beta_0^2\beta_1^2 + \beta_1^4$ as well as $\bar{\kappa} := \pi_E(G|\tilde{q})\bar{p} - \pi_E(B|\tilde{q})\bar{r}$.

5.2.4. *Social Planner Interlude.* As a preparation for the coming analysis of the impact of overconfidence on the social welfare, let us take a moment to consider the possible distortion effects of overconfidence. It is clear that overconfidence acts as a distortion in terms of the assessment of the project quality by the entrepreneur, specifically via the signal and the interpretation of it. In addition, as the discussion of the model dynamics revealed, the bargaining process between the entrepreneur and the venture capitalist introduces another distortion. While the interplay between these two distortions appears as part of the following investigation, the focus will be on studying a shift in overconfidence. In particular, let us present an intuition why an increase in overconfidence might not necessarily be detrimental for social welfare, which ultimately lays the foundation for the next analytical part of this paper. Now, in order to illustrate this, let us examine the perspective of a social planner.

The hypothetical social planner employed here chooses all strategies based on the previously introduced case-specific welfare levels $W^{\tilde{q}}$ given the distorted signal provided to the entrepreneur $\tilde{q}$. Thus, while the social planner takes the existence of overconfidence into consideration, its distortion on the level of the signals persist as another layer of noise. In other words, it is not an omniscient social planner for which the issue of overconfidence does not exist, but one who acts accordingly. Ultimately, the social planner maximizes with respect to the same strategy tuple

$$W^{\tilde{q}} = \begin{cases} \sum_{Q\in\{B,G\}} P(Q|\tilde{q})Y(Q, e_0, e_1, e_2(\lambda, s)) - C(e_0) - C(e_1) - C(e_2(\lambda, s)) - \bar{r}I \\ 0 \end{cases}$$

depending on whether the project gets funded (bank or venture capitalist) or not. As a consequence, three potential strategy tuples emerge for the social planner. Following the previously used notation, let us simply label them 0, B, resp. VC:

	e_0	λ	s	e_1	e_2
0	0	$\emptyset$	0	0	0
B	$\mathbb{E}_P(\alpha\|\tilde{q})\beta_0$	$\emptyset$	B	$\mathbb{E}_P(\alpha\|\tilde{q})\beta_1$	0
VC	$\mathbb{E}_P(\alpha\|\tilde{q})\beta_0$	λ	VC	$\mathbb{E}_P(\alpha\|\tilde{q})\beta_1$	$\bar{e}$

Note that here the numerical value of $\lambda \in [0, 1]$ plays no role for the scenario VC. Ultimately, the social planner chooses out of those three the maximal element based

on the following case-specific welfare levels (with similar rules on tie-breaking):

$$W^{\tilde{q}}(0) = 0$$
$$W^{\tilde{q}}(B) = \frac{\mathbb{E}_P(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right)}{2} - \bar{r}I$$
$$W^{\tilde{q}}(VC) = \frac{\mathbb{E}_P(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right)}{2} - \bar{r}I + \frac{\left(2\mathbb{E}_P(\alpha|\tilde{q})\beta_2 - \bar{e}\right)\bar{e}}{2}$$

Let this serve as the first-best benchmark case when studying the influence of the share of overconfident entrepreneurs on social welfare.[27] As mentioned before, in contrast to the social planner, the standard model contains two distortions, namely the bargaining process and the overconfidence of the entrepreneur. In the following, let us then study the impact of a shift in the level of overconfidence in comparison with the benchmark case. It is not a complete (technical) analysis but provides us with an intuition for the potential beneficial nature of an increase in overconfidence with respect to social welfare. As a consequence, an increase in overconfidence might actually result in the second-best scenario.[28] Finally, assume for the moment that the parameters, specifically the level of overconfidence, stay within a corridor such that the type of equilibrium remains the same. Also, assume that both elements of the equilibrium pair belong to the same category of funding for now.

First, let us compare the scenario of 'no project' of the social planner with the corresponding bank scenario (recall the established rules on tie-breaking). Clearly, due to the deterministic nature of the bank, the cost for the provision of funding acts as an externality. While it is a cost that the entrepreneur does not internalize (and neither does the bank), it is an effect independent of the level of overconfidence.

Second, let us consider the social planner's 'bank' scenario in contrast to the corresponding bank scenarios. Similar to the other case, there exists a distortion with respect to the cost for the provision of funding. However, there also exist additional distortions that affect the overall welfare on two levels, specifically via the involved probabilities and the induced choices for each of the two signals. First, as presented in Proposition 4, a shift in the level of overconfidence implies a similar one with respect to the weights used with the case-specific welfare levels.[29] Second, overconfidence also influences the provision of effort (initial and additional). Now, the extent of that effect depends on the dominant bank scenario and the parameters. Recall that $\mathbb{E}_{\pi_E}(\alpha|\tilde{g}) > \mathbb{E}_P(\alpha|\tilde{g})$ and $\mathbb{E}_{\pi_E}(\alpha|\tilde{b}) = \mathbb{E}_P(\alpha|\tilde{b})$ (see Proposition 4) while by definition $\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) > \pi_E(G|\tilde{q})\alpha(G)$. Thus, for the case of a full (re-)payment, overconfidence implies a deviation in case of the 'good' signal, $\mathbb{E}_{\pi_E}(\alpha|\tilde{g}) > \mathbb{E}_P(\alpha|\tilde{g})$, but no deviation for the 'bad' signal, $\mathbb{E}_{\pi_E}(\alpha|\tilde{b}) = \mathbb{E}_P(\alpha|\tilde{b})$. Moreover, a change in the level of overconfidence then amplifies this deviation as the social planner's choice decreases further with an increasing level of overconfidence, $\frac{\partial \mathbb{E}_P(\alpha|\tilde{g})}{\partial \mu} > 0 = \frac{\partial \mathbb{E}_P(\alpha|\tilde{b})}{\partial \mu}$. On the other hand, the case of a partial (re-)payment ultimately depends on the parameter constellation. Without specifications, it is unclear to what extent

[27] Alternatively, one could argue that the first-best scenario should actually be one without any overconfident entrepreneurs.

[28] Note that in terms of 'The General Theory of Second Best' (by Lipsey and Lancaster (1956)), this approach considers the two distortions as one and then analyzes another distortion via a shift of the level of overconfidence.

[29] In other words, a higher level of overconfidence implies a higher probability for a good signal and therefore a lower probability for a bad signal.

the entrepreneur deviates from the social planner in the case of the 'good' signal, $\pi_E(G|\tilde{g})\alpha(G) < \mathbb{E}_{\pi_E}(\alpha|\tilde{g}) > \mathbb{E}_P(\alpha|\tilde{g})$, while the 'bad' signal always yields a deviation, $\pi_E(G|\tilde{b})\alpha(G) < \mathbb{E}_{\pi_E}(\alpha|\tilde{b}) = \mathbb{E}_P(\alpha|\tilde{b})$. Here, increasing overconfidence potentially decreases the undetermined deviation. In a sense, in this scenario overconfidence might alleviate the issue of the possibility for a default on the loan.

Ultimately, combining all the presented insights, i.e. those about the probabilities and those about the effort levels, generally yields an inconclusive picture. In fact, the setting where full (re-)payment of the funds emerges as the equilibrium for both signals stands as the sole exception to this. In this case, a higher overconfidence leads to a higher weight on a higher deviation. Otherwise, an increase of overconfidence might actually be beneficial for social welfare, but it ultimately depends on the parameter constellation.[30]

Third, let us study the social planner's 'venture capitalist' scenario in contrast to the corresponding venture capitalist scenarios. While the previously described externality is internalized here, it still exhibits the same distortion of the probabilities and similar distortions in terms of the provided effort. With the distortion effect on the probabilities already discussed in the previous paragraph, let us turn to the similar but different distortion effect on the provision of effort. Specifically, let us focus on the additional effort for now. Here, the bargaining process introduces the project share, which essentially functions as a discount factor. Recall, that irrespective of the specific venture capitalist scenario the additional effort is equal to $(1-\lambda)\,\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1$. Contrast this to the choice of the social planner of $\mathbb{E}_P(\alpha|\tilde{q})\beta_1$. Again, the relations $\mathbb{E}_{\pi_E}(\alpha|\tilde{g}) > \mathbb{E}_P(\alpha|\tilde{g})$ and $\mathbb{E}_{\pi_E}(\alpha|\tilde{b}) = \mathbb{E}_P(\alpha|\tilde{b})$ (Proposition 4) yield a definite deviation for the 'bad' signal while the situation for the 'good signal' depends the parameter constellation. Also, increasing overconfidence potentially decreases the undetermined deviation - similar to the situation in the bank scenario. In a way, overconfidence might alleviate the distortion of the bargaining process. With respect to the initial effort, the complexity of the situation increases due to the strategic interactions between the initial effort and the project share. Further, the initial effort depends on the specific venture capitalist scenario, which includes the pseudo-bank cases. With our intention of simply providing an intuition in mind, let us therefore skip this aspect.

All in all, the venture capitalist scenario resembles the bank scenario (apart from its one exception) in that it is not straight forward to see the overall direction of the impact of a change in the level of overconfidence. In particular, an increase of overconfidence might not necessarily be detrimental for social welfare, provided an appropriate parameter constellation.

Finally, another avenue for the distortion effect of overconfidence to influence the social welfare is a shift in the type of the equilibrium. In the previous paragraphs, the assumption was that it would always remain the same. However, any shift of the level of overconfidence that changes the type of equilibrium would immediately affect the social welfare beyond the previously presented dynamics.[31] In particular, the

[30] This includes the 'mixed' bank scenario of partial and full repayment for different signals.

[31] It is important to point out that such a change will only occur in the case of a 'good' signal, because the share of optimists does not influence any choices or profits for a 'bad' signal - see (the comments on) Proposition 4. Furthermore, any change in the level of overconfidence leaves both the evaluation of the bank scenario(s) by the entrepreneur and the perspective of the bank completely untouched. Thus, any changes in the type of equilibrium follow from the development in the venture capitalist scenario(s).

direction of the distortion effect then depends on even more factors. Furthermore, without the assumption that both elements of the equilibrium pair belong to the same category of funding, the analysis becomes even more complicated. In any case, without any restrictions on the parameter constellation it is difficult to make any meaningful statement at this point. As a consequence, let us change our approach and attempt to solve this problem numerically (in the next part).

5.3. **Overconfidence Dynamics.** Let us now study the comparative static effect of overconfidence on the social welfare while varying the model parameters. However, due to the number of parameters, it is not feasible to vary all of them at once. Instead, let us analyze all possible pairs of model parameters while fixing the rest. It ultimately allows us to characterize relationships between the model parameters that facilitate the existence of a share of optimists such that an increase in its level increases social welfare - in other words, a local maximum of social welfare as a function of the share of realists, which differs from the respective maximum value.

With the formulae for the expected profit and the social welfare already calculated, this becomes a simple task of substituting parameters with the respective values and comparing the resulting expressions. This problem is solved numerically with the help of an algorithm - Appendix E contains the corresponding pseudocode. Note that this algorithm considers all possible combinations of the strategy tuples, using the previously presented expressions, provided their existence criterion is satisfied.

5.3.1. *Parameter Values.* As mentioned before, the number of parameters prohibits a simultaneous analysis and the focus will be on one pair of parameters at a time. Thus, for the other parameters (ultimately all of them) this requires default values. In this context, default value does not mean that the model is fitted to the data representing a default state of the world though. However, while listing all of the parameters under consideration with their respective intervals and default values here, it is still our intention to provide an intuition behind these default values. Combined with the analysis of all (pairwise) matchings of parameters, this provides robustness for the findings.

Before presenting the details for the parameters, let us assume a specific structure for two parameters in order to add further meaning to them. First of all, recall the fixed effort contribution by the venture capitalist. Let us turn this absolute value into a relative one instead. Assume for now that it is the product of a scalar and the general expected project productivity times the additional effort productivity of the venture capitalist, i.e. $\bar{e} = \theta_{\bar{e}} \mathbb{E}_P(\alpha)\beta_2$ for $\theta_{\bar{e}} > 0$. In a way, this corresponds to the notion that the package of services by the venture capitalist represents the result of an evolutionary process independent of the current project. Furthermore, approach the fixed cost of investment in a similar manner. As a reference, take the general expected payoff (without cost of funding) of a hypothetical bank scenario divided by the interest rate of the bank plus a correction for inflation and consider a scalar, i.e. $I = \theta_I \mathbb{E}_P(\alpha)^2 \left(\beta_0^2 + \beta_1^2\right)/2\,(r + 0.02)$ for $\theta_I > 0$. In other words, it is a combination of the concept of Return On Invested Capital as presented in the work by Damodaran (2007) and the policy of the European Central Bank displayed in 'The definition of price stability'. Note, while both of these assumptions yield two new parameters independent of the others, the original ones are now linked to the rest of the parameters - a point to keep in mind for the analysis.

In terms of the respective intervals and default values, let us first consider the different effort productivities, β_0, β_1, and β_2. Apart from the scenario where one of the parameters is equal to one, i.e. the case without any additional scaling for the respective effort, these values in itself contain only limited information. However, their relation with each other provides additional context. It seems reasonable to assume that the additional effort productivity of the entrepreneur is higher than the initial one and that of the venture capitalist - corresponding to the notion that effort is less productive in the planning phase than during the actual project and the project-independent package of services is less productive as well. In general, restrict the three parameters to the interval $(0, 10)$, which turns out to be sufficient to make meaningful statements. Finally, set the default values as $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$. Thus, no additional scaling for the initial effort of the entrepreneur and the additional effort of the venture capitalist, with twice the return on invested additional effort by the entrepreneur.

With the effort productivities attended to, let us examine the project ones, $\alpha(Q)$. As absolute values they share the issue of limited information within the parameter with the previous effort productivities. However, the difference between the two allows us an interpretation as the degree of 'project (in-)equality'. Let us use the valuation multiples for earnings used by the financial industry as a proxy. Looking at the corresponding data, for example Lie and Lie (2002), reveals a value in the range of 23 to 26 resp. 33 to 36 (EBITDA resp. EBIT) for the 75th percentile - depending on whether it is adjusted for cash and cash equivalents or not. Similarly, the paper finds 8 to 9 resp. 11 to 13 for the 25th percentile. Justified by these estimates, take the interval $(0, 100)$ for both types - with the restriction that $\alpha(B) < \alpha(G)$. Furthermore, take $\alpha(B) = 1$ and $\alpha(G) = 20$ as the default values. It might seem as though this stretches the findings of the paper, but remember that these values only serve as a reference point. In particular, the first default value simply assumes that the 'baseline' yields no (further) scaling in either direction.

Now, the quality of the project is ultimately determined via the corresponding probability for a good project γ. Without additional restrictions, these take values from the full interval $(0, 1)$. It is an established fact, that the majority of start-ups never make it past the first couple of years. As an example, take the work of Scarpetta, Hemmings, Tressel, and Woo (2002), which shows that about a third does not survive the first two years, while close to two-thirds fail to reach seven years. While survival of a start-up does not necessarily equal a good project quality, let us still use it as a proxy, specifically take the default value $\gamma = 1/3$.

Consider the noise ϵ, which represents the level of non-overconfidence distortions, as the next parameter. Recall the original assumption of a limit of 0.5 necessary to guarantee that the expressions 'good/bad news' are actually descriptive. For the sake of robustness, let us not restrict the interval any further, i.e. use $(0, 0.5)$. Also, as it is almost impossible to find a suitable proxy for it, simply settle for the middle ground of $\epsilon = 0.25$ as the default value. It represents a setting where noise plays a substantial role in the probabilistic process but without completely overshadowing other aspects.

Now, both the risk-free interest rate $\bar{r}$ and the interest rate premium $\bar{p}$ allow us to consult corresponding data. In particular, use United States Treasury securities as a guideline for the risk-free interest rate. As the 'Daily treasury yield curve rates' by the U.S. Department of the Treasury reveals, the interest rate in 2019 usually

varies between 2 and 3 percent (depending on maturity and date). Furthermore, take the findings of Arnott and Bernstein (2002) as a point of reference for the interest rate premium, which estimate a value of 2.4 percent. While each of them naturally varies over time, it seems reasonable to assume an interval of $(0, 0.35)$. Also, motivated by the data, let us use default values of $\bar{r} = \bar{p} = 0.025$. In a way, this corresponds to a time-frame close to a year. Finally, even though the intervals and default values coincide, it is important to stress that each of them still varies individually in the analysis.

Next, return to the scalar for the effort investment by the venture capitalist, $\theta_{\bar{e}}$. In general, let us vary the parameter within the interval $(0, 2)$. Thus, potentially the venture capitalist invests up to twice the project-independent marginal return. However, using the insights about the minimal value of the project share as a point of reference, let us set $\theta_{\bar{e}} = 0.5$ as the default value.

Finally, let us re-visit the scalar for the cost of investment, θ_I. Generally, limit the parameter to values in the interval $(0, 2)$. Therefore, when comparing it to a hypothetical bank scenario, the cost of investment possibly reach twice the amount associated with matching inflation. Using the fact that a company matching inflation is essentially on the verge of either destroying or creating value, let us set $\theta_I = 1$ as the default value.

5.3.2. *Varying Overconfidence.* With the default values for the parameters codified, let us move to the pairwise comparative static effects of overconfidence. In particular, let us study all parameters individually and characterize for each of them, where possible, a configuration that facilitates a beneficial impact of overconfidence. Now, beneficial impact in this context means that for this specific parameter constellation there exists a value for the share of realists such that decreasing it yields an increase in social welfare. Thus, it is a statement about a local maximum, not a global one. Formally, it is about the sign of $\frac{\partial W}{\partial \mu}(P)$ for a varying parameter constellation P in two dimensions.

As an appetizer, let us examine two parameter constellations in detail. Without the other parameters varying, it is possible to depict and describe the exact impact of the change in overconfidence on social welfare and on the equilibrium scenario, in particular on the source of funding. Let us first consider the parameter constellation which corresponds to all of the default values. In the associated figure, Figure 3, the share of realists μ is on the x-axis and social welfare is on the y-axis. Specifically, the black dots in the graphics correspond to the value of the adjusted social welfare, while the gray ones follow from the social welfare of the social planner.

First of all, as seen in Figure 3, in the case of the default parameter constellation an increase in overconfidence (from right to left) leads to a decrease in social welfare for all levels of overconfidence - both in the general model and for the social planner. Irrespective of the share of realists, the entrepreneur chooses the venture capitalist, either as a pseudo-bank in the good case or unconstrained in the bad case. Here, the hybrid nature of the venture capitalist as a pseudo-bank turns out to be a problem for overconfidence in two ways. First, while the imitated bank scenario corresponds to partial (re-)payment of the funds (either beneficial or detrimental), the default parameter constellation is such that an increase in overconfidence almost always increases the deviation from the social planner's choice (apart from a share of realists close to one). Second, in the scenario of a pseudo-bank, the effort of the entrepreneur only coincides with that of the bank scenario for the initial effort, while

the additional effort follows the dynamics of the venture capitalist scenario. Again, the values yield that an increase in overconfidence always increases the deviation. On top of that, an increase in overconfidence also puts more weight on this scenario. Meanwhile, the deviation in the other scenario remains constant, with less weight.

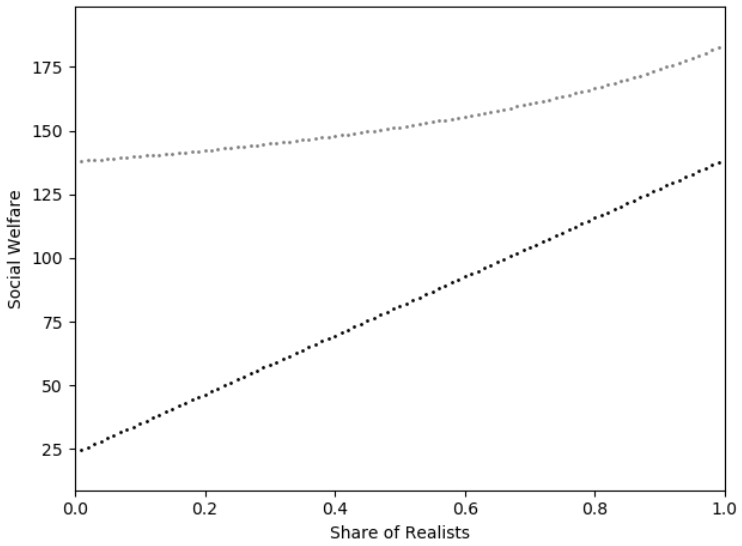


FIGURE 3. Influence of the share of realists on social welfare for the default parameter constellation

By contrast, in the case of a modified (default) parameter constellation, as presented in Figure 4 (with similar axes and functions), an increase in overconfidence implies different social welfare effects, depending on the degree of overconfidence. Note that this setting deviates in terms of two parameters from the default values, specifically $\beta_1 = 5$ and $\alpha(B) = 7$ (compared to $\beta_1 = 2$ and $\alpha(B) = 1$).[32] As the figure shows, there exists a critical level of overconfidence from which onwards the equilibrium scenario changes. As indicated before, this necessarily only occurs for the 'good' case while the 'bad' case stays the same. Here, the venture capitalist does not act as a pseudo-bank anymore and the entrepreneur chooses the bank proper, with full (re-) payment. The jump in social welfare originates from eliminating part of the aforementioned issues with the pseudo-bank's hybrid nature. Specifically, the change in the equilibrium scenario alleviates the impact of the bargaining process on the additional effort of the entrepreneur while the distortion effect of optimism on the initial effort persists. Meanwhile, independent of the level of overconfidence, the unconstrained venture capitalist scenario emerges in the other case. Finally, a change in overconfidence necessarily affects the probability for the two cases and therefore the weight on the development and the standstill respectively. Ultimately, the effect of overconfidence on social welfare is detrimental on a global scale here, but beneficial on a local level.

[32]In other words, the modified value for $\alpha(B)$ now resembles the proxy mentioned before, while the one for β_1 corresponds to a particularly productive entrepreneur (during the project).

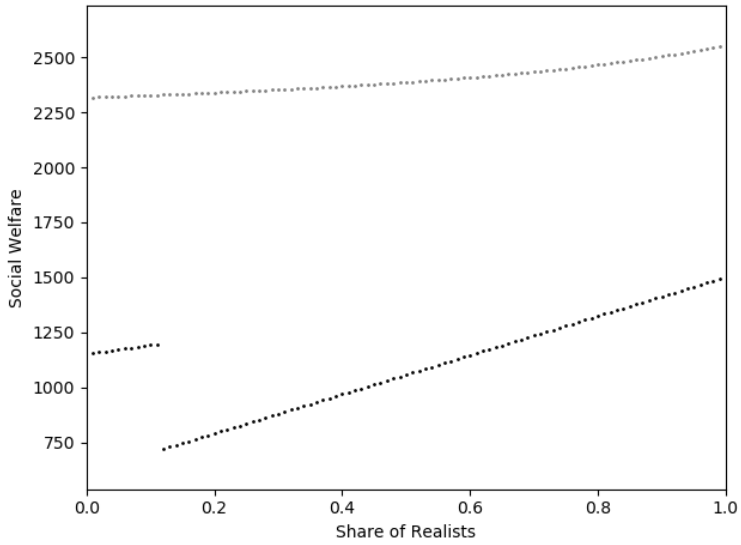


FIGURE 4. Influence of the share of realists on social welfare for the modified parameter constellation

While the modified parameter constellation seems to be a special one, it turns out that the underlying mechanics also appear for a variety of circumstances. Now, to elaborate on this and other statements, let us finally examine the pairwise comparative static effects of overconfidence. Appendix E contains a table listing all of the parameters under consideration together with the corresponding figures and an explanation of their design. In the following paragraphs, let us present the associated findings.

As mentioned before, the change in one of the two equilibrium scenarios from pseudo-bank to bank proper drives the beneficial impact of overconfidence for a number of parameter constellations. In fact, as the figures indicate, it is the cause (not just a correlation) for almost all of them. While the change might also occur under circumstances with no beneficial impact, in case there is a beneficial impact, it is almost always due to this change. Now, remember that the expected profit of both the pseudo-bank and the bank proper is identical for the entrepreneur and independent of the level of overconfidence. Thus, whenever overconfidence reaches a critical value, the venture capitalist not acting as a pseudo-bank anymore is due to a lack of (positive) expected profit (that acknowledges the existence of optimists) on the part of the venture capitalist. In other words, an abundance of optimists forces the venture capitalist out of the competition with the bank, which - in part due to its non-player nature - yields better conditions in terms of social welfare. Finally, let us go through all of the parameters and formulate constellations conducive to this mechanics.

First, focus on the three effort productivities β_0, β_1, and β_2, and their respective overconfidence dynamics, as presented in Figure 5, 6, and 7. It appears that in almost all comparisons with the other parameters, a change in β_0 neither increases nor decreases the chances of a potential beneficial impact of overconfidence. Now,

in those parts of the comparison where its value matters, it seems to be beneficial to be in the low to medium range - but it varies notably, depending on the counterpart. Meanwhile, almost all of the comparisons of β_1 and β_2 indicate a tendency towards medium to high and low values respectively. For all three of them, the lack of beneficial impact of overconfidence otherwise follows from a pseudo-bank scenario which turns out to be profitable for the venture capitalist independent of the level of optimists. Recall, that from the point of view of the entrepreneur the pseudo-bank resembles the bank in pay-off, which then implies that an increase in β_0 and β_1 automatically increases the expected profit, while a similar increase in β_2 does not. However, while the initial effort increases accordingly (identical to the bank one), the bargaining process interferes with the increase of the additional effort. Therefore, a low value for β_1 makes the expected profit of the venture capitalist resistant to an abundance of optimists, which otherwise threatens it, while β_0 may vary. Finally, remember our assumptions on the specific structure of I, which therefore implies that an increase of each of these two values increases the cost of the investment. However, due to the structure of the pseudo-bank, the entrepreneur mostly shares the impact of that development. As mentioned before, these underlying mechanics appear not for a shift in β_2, which simply provides benefits to the venture capitalist. Even the assumption on $\bar{e}$, which affects the cost of effort, does not change this.

Next, consult the figures 8 and 9 for the analysis with respect to the two overall project productivities $\alpha(B)$ and $\alpha(G)$. It is intuitive that a low value for $\alpha(B)$ and a high one for $\alpha(G)$ facilitate a setting where a critical value of overconfidence limits the venture capitalist. Otherwise, the lack of significant difference between the project qualities simply implies that the signal, therefore by extension overconfidence, plays no significant role in the whole process.

Moreover, the figures 10 and 11, which correspond to the probabilities γ and ϵ for the project being good and the signal being influenced by noise, reveal a picture that appears inconclusive at times but still illustrates tendencies. In terms of the value for γ it seems that it should be in the low to medium range in order to promote an environment where the pseudo-bank exists but remains sensitive to overconfidence. Intuitively, a value too high guarantees a pseudo-bank resistant to overconfidence, where (similar to the previous paragraph) the difference between good and bad project becomes negligible through the significantly different probabilities. Similarly, a value too low features a complete lack of pseudo-bank, irrespective of optimism. Meanwhile, the findings for ϵ follow an identical intuition but with reversed directions, i.e. a medium to high value appears conducive to the aforementioned environment.

Further, the risk-free interest rate $\bar{r}$ and the interest rate premium $\bar{p}$ resemble each other in terms of the conclusions from their figures 12 and 13. Across the board, the graphics suggest a low but not too low value for both of them. The underlying mechanics work in opposite ways though. While a high value for $\bar{r}$ creates a setting where the pseudo-bank never appears and a low value establishes circumstances where it never disappears, the complete opposite applies to $\bar{p}$. Now, the explanation follows from the structure of the pseudo-bank, whereby the interest rate premium and the risk-free interest rate constitute a potential transfer from the entrepreneur to the venture capitalist while the latter always pays the risk-free interest rate. Also, a higher interest rate premium automatically lowers the cost of re-financing due to our assumption on the specific structure of I, while - again - the complete opposite holds for the risk-free interest rate.

Finally, consider the scalars for service package $\theta_{\bar{e}}$ and cost of investment θ_I with their figures 14 and 15. While the findings in terms of the service package differ across the parameters, it appears that a configuration of either a low or a high value promotes a beneficial impact of overconfidence. Recall, that the service package incurs quadratic costs while the project supplies linear benefits, similar to the effort provided by the entrepreneur. Therefore, in the medium range, the fixed effort is such that the pseudo-bank emerges as a constant part of the equilibrium. Similarly, a low cost of investment yields a consistent pseudo-bank due to its structure (see the discussion in the previous paragraph). The graphics support this intuition, as a high value for the cost of investment facilitates a beneficial impact of overconfidence.

All in all, this whole analysis implies that a setting which facilitates a potential beneficial impact of overconfidence ideally features a medium to high β_1, a low β_2, a sufficiently low $\alpha(B)$ and high $\alpha(G)$, also a low but not too low $\bar{r}$ and $\bar{p}$, as well as either a low or a high $\theta_{\bar{e}}$ and a relatively high θ_I. While the picture remains somewhat inconclusive for the other parameter values, it shows a tendency towards a low to medium β_0 and γ, and a medium to high ϵ. Ultimately, all of these characterizations yield a potential beneficial impact of overconfidence through a change in one of the two equilibrium scenarios from pseudo-bank to bank proper.

A notable exception to this are the circumstances of a relatively high value for either γ or ϵ (or both), where overconfident entrepreneurs essentially counter the non-overconfidence distortions without any changes to the equilibrium scenario.[33] Here, optimists should almost be referred to as realists as well. Furthermore, whenever the delta between $\alpha(B)$ and $\alpha(G)$ remains relatively low, overconfident entrepreneurs incur only limited social cost with their biased judgement while simultaneously providing significant social benefits through a shift in the general probabilities, from the unconstrained venture capitalist to the bank scenario - but without changing the equilibrium.

6. Discussion

Finally, let us consider our model and its insights in the context of the related literature mentioned before. In general, it is unfortunately not a straight forward exercise to compare our model with those in the related literature, as it contains a variety of components from different streams of the literature. For example, while our model shares the implementation of the optimistic entrepreneur with the model of Manove and Padilla (1999), it differs substantially with respect to the other aspects of the model. However, our model shares enough similarities with the paper by de Bettignies and Brander (2007) to allow some basic comparisons.

In contrast to our model, theirs contains no initial effort of the entrepreneur and the effort provided by the venture capitalist is endogenous, not exogenous. Also, in their model the entrepreneur proposes the terms of funding and there does not exist any potential distortion in the form of optimism. Thus, to a certain degree their model functions as a (modified) benchmark model for the case without any optimistic entrepreneurs. When translating concepts from their to our setting, for example their 'chance of success' corresponds to our 'expected project productivity' and 'entrepreneurial effort' resembles 'additional effort', a number of features match:[34]

[33] It is a phenomenon more prevalent with γ and less so with ϵ.

[34] Note the fundamental difference between their 'chance of success' as an objective probability and our 'expected project productivity' as a subjective expectation.

First, their insights on the comparative static effects in the case of bank finance as presented in their Proposition 1 essentially matches (in part) our Proposition 1. Second, the case of venture capitalist finance yield a similar equivalence - compare their Proposition 2 to our Proposition 2. However, none of their insights with respect to the effort provided by the venture capitalist make sense in our set-up. Further, the relation between their findings in terms of the project share and ours varies - contrast their Proposition 3 with our Proposition 3. While in both models the influence of the effort productivity of the venture capitalist is positive on the project share and the effect of the effort productivity of the entrepreneur is negative, the overall limit on the project share is the other way around. While they find 1/2 to be the upper bound of the project share, we find it is the lower bound. Here, the results diverge due to the different set-ups with respect to the bargaining process (and the induced bargaining power). In our model the venture capitalist proposes a project share (after the entrepreneur already invested the initial effort), while in their model the entrepreneur proposes a project share (before any effort is invested). Thus, the power dynamic is completely the opposite. Finally, their paper contains a number of other statements, but due to the fundamental differences of the models, these become difficult to compare and translate.

7. Conclusion

The framework of this paper presents a tool for the analysis of the influence of overconfidence on an entrepreneur facing a project of unknown quality that requires the investment of both effort and funds. In this model, overconfidence manifests in the form of a distortion of the signal about the quality of the project. Using the concept of Perfect Bayesian Equilibrium, with a focus on pure strategies, and concentrating on studying separating equilibria, the calculations revealed a total of nine potential strategy tuples per each of the two signal types. Among these nine are three scenarios with funding via the bank, three with the venture capitalist, and three where the venture capitalist acts as a pseudo-bank.

In the analysis, this paper featured a study of the underlying model dynamics, specifically the comparative static effects of a selection of different model parameters on the player's choices and their profits. In particular, it allowed us to highlight the differences between the bank and the venture capitalist scenarios, specifically the impact of the strategic interactions and the bargaining process, and emphasize the hold-up problem present in the model. Moreover, introducing social welfare let us examine the impact of other models parameters on the comparative static effect of the level of overconfidence. A social planner interlude provided us with an intuition why an increase in overconfidence might actually be beneficial for social welfare. As it turns out, the distortion effect of optimism potentially alleviates other distortions, specifically the bargaining process of the entrepreneur and the venture capitalist, and the possibility for a default on the contract with the bank. The numerical part of our analysis then characterized different constellations of the model parameters under which an increase in overconfidence leads to an increase in social welfare. A significant portion of these featured a change in terms of the source of funding, specifically from the venture capitalist as a pseudo-bank to the bank proper. There, a degree of overconfidence beyond a critical level forced the venture capitalist out of the competition with the bank, which implied better conditions for social welfare.

While it is a local not a global statement, it does dispel the notion that an increase in overconfidence is necessarily detrimental for social welfare.

The contribution of this work to the literature is a comprehensive model that features the investment of effort in a project of unknown quality under the assumption of a perception bias on the side of the entrepreneur while simultaneously answering the question about the source of financing put forth by Da Rin, Hellmann, and Puri. As presented in the comparison with the paper by de Bettignies and Brander (2007), on a basic level our model produces results similar to that of a model without any form of overconfidence. Furthermore, the aforementioned (numerical) analysis of the comparative static effects, specifically that of a shift in the level of overconfidence, provides additional insights, which fills a perceived gap in the literature. While a number of papers already study the effect of a perception bias on different aspects of entrepreneurship, the literature lacks a multi-faceted model where various effects of the existence and proportion of overconfidence might be examined simultaneously. In particular, the framework of this paper permits an analysis of the impact via two different channels at the same time, namely in terms of the investment of effort and the source of finding, thereby providing additional robustness to the findings.

Looking ahead, a point in need of further investigation are the model's limitations. Including multiple aspects in a single model often comes with the necessity to otherwise restrict these facets in order to maintain manageability. Currently, the model lacks an endogenous bank and limits the endogeneity of the venture capitalist in terms of the effort. It is our opinion that rectifying this issue would not only bolster the general robustness of the model but specifically that of the characterization of the beneficial overconfidence. While it would probably require a modified methodology, it would make for a worthwhile undertaking.

Appendix A. Notation

	Type	Description
A	$\in \{O, R\}$	type of the entrepreneur
C	$: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$	cost of the individual effort
E	$: \mathbb{R}^3_{\geq 0} \to \mathbb{R}_{\geq 0}$	aggregate effort
e_0	$\in \mathbb{R}_{\geq 0}$	initial effort of the entrepreneur
e_1	$\in \mathbb{R}_{\geq 0}$	additional effort of the entrepreneur
e_2	$\in \{0, \bar{e}\}$	additional effort of the venture capitalist
$\bar{e}$	$\in \mathbb{R}_{>0}$	service package of the venture capitalist
I	$\in \mathbb{R}_{\geq 0}$	cost of the investment
P	$\to [0, 1]$	universal probability measure
$\bar{p}$	$\in \mathbb{R}_{\geq 0}$	interest rate premium
Q	$\in \{G, B\}$	quality of the project
q	$\in \{g, b\}$	signal of the quality
$\tilde{q}$	$\in \{\tilde{g}, \tilde{b}\}$	perceived signal of the quality
r	$\in \mathbb{R}_{\geq 0}$	interest rate of the bank
$\bar{r}$	$\in \mathbb{R}_{\geq 0}$	risk-free interest rate
s	$\in \{0, B, VC\}$	source of funding
Y	$: \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$	gross project payoff
α	$: \{G, B\} \to \mathbb{R}_{\geq 0}$	overall project productivity
β_0	$\in \mathbb{R}_{\geq 0}$	initial effort productivity of the entrepreneur
β_1	$\in \mathbb{R}_{\geq 0}$	additional effort productivity of the entrepreneur
β_2	$\in \mathbb{R}_{\geq 0}$	additional effort productivity of the venture capitalist
γ	$\in [0, 1]$	probability of the project being of good quality
ϵ	$\in [0, 1]$	probability of the signal being influenced by noise
λ	$\in [0, 1]$	share of the venture capitalist
μ	$\in [0, 1]$	probability of the entrepreneur being a realist
π	$\to [0, 1]$	belief system of the venture capitalist
π_E	$\to [0, 1]$	belief system of the entrepreneur
ρ_q	$\in [0, 1]$	probability of the project quality matching its signal

Note that this list only contains the notation used in the main part of the paper. Therefore, none of the expressions only used for the calculations in this appendix appear here in order to maintain usability of this table.

Appendix B. Auxiliary Calculations

As indicated in the main body of the paper, assume that the noise is not too extreme in its effect, i.e. its parameter satisfies $\epsilon < 1/2$. This assumption implies that a 'good' signal is good news and a 'bad' signal is bad news:

Remark 1. *If $\epsilon < 1/2$, then $\rho_g > 1 - \rho_b$ (or $\rho_b > 1 - \rho_g$) and $\mathbb{E}_P(\alpha|g) > \mathbb{E}_P(\alpha|b)$. Also, $\pi_E(G|\tilde{g}) > \pi_E(G|\tilde{b})$ (or $\pi_E(B|\tilde{b}) > \pi_E(B|\tilde{g})$) and $\mathbb{E}_{\pi_E}(\alpha|\tilde{g}) > \mathbb{E}_{\pi_E}(\alpha|\tilde{b})$.*

Proof. Recall the definitions of the ρ_q:

$$\rho_g = P(G|g) = \frac{(1-\epsilon)\gamma}{(1-\epsilon)\gamma + \epsilon(1-\gamma)} \qquad 1 - \rho_g = P(B|g) = \frac{\epsilon(1-\gamma)}{\epsilon(1-\gamma) + (1-\epsilon)\gamma}$$

$$\rho_b = P(B|b) = \frac{(1-\epsilon)(1-\gamma)}{(1-\epsilon)(1-\gamma) + \epsilon\gamma} \qquad 1 - \rho_b = P(G|b) = \frac{\epsilon\gamma}{\epsilon\gamma + (1-\epsilon)(1-\gamma)}$$

Then, by multiplying with the respective common denominator and simplifying (using in the process that $\gamma \neq 0, 1$) it ultimately follows that:

$$\begin{aligned} & 0 < \rho_g - (1 - \rho_b) \\ \Leftrightarrow\ & 0 < \frac{(1-\epsilon)\gamma}{(1-\epsilon)\gamma + \epsilon(1-\gamma)} - \frac{\epsilon\gamma}{\epsilon\gamma + (1-\epsilon)(1-\gamma)} \\ \Leftrightarrow\ & 0 < \gamma\,(1-\gamma)\,(1-2\epsilon) \\ \Leftrightarrow\ & \epsilon < \frac{1}{2} \end{aligned}$$

Alternatively, it is possible to use the expressions for $\rho_b - (1-\rho_g)$ provided beforehand. Furthermore, by using $\alpha(G) > \alpha(B)$ and the previous statement, $\rho_g + \rho_b - 1 > 0$:

$$\begin{aligned} \mathbb{E}_P(\alpha|g) - \mathbb{E}_P(\alpha|b) &= (\rho_g - (1-\rho_b))\,\alpha(G) + ((1-\rho_g) - \rho_b)\,\alpha(B) \\ &= (\rho_g + \rho_b - 1)\,(\alpha(G) - \alpha(B)) > 0 \end{aligned}$$

Finally, with $\pi_E(Q|\tilde{q}) = P(Q|q)$ it follows directly that $\pi_E(G|\tilde{g}) > \pi_E(G|\tilde{b})$ or alternatively $\pi_E(B|\tilde{b}) > \pi_E(B|\tilde{g})$. Furthermore, the previous statement then yields:

$$\mathbb{E}_{\pi_E}(\alpha|\tilde{g}) = \mathbb{E}_P(\alpha|g) > \mathbb{E}_P(\alpha|b) = \mathbb{E}_{\pi_E}(\alpha|\tilde{b})$$

□

Similarly, the limitation yields that the optimist is indeed optimistic:

Remark 2. *If $\epsilon < 1/2$, then $\pi_E(G|\tilde{g}) > P(G|\tilde{g})$ (or $P(B|\tilde{g}) > \pi_E(B|\tilde{g})$) as well as $\mathbb{E}_{\pi_E}(\alpha|\tilde{g}) > \mathbb{E}_P(\alpha|\tilde{g})$.*

Proof. Recall that $\pi_E(Q|\tilde{q}) = P(Q|q)$ (cf. the ρ_q expressions) and then use:

$$P(G|\tilde{g}) = \frac{((1-\epsilon) + \epsilon\,(1-\mu))\,\gamma}{((1-\epsilon) + \epsilon\,(1-\mu))\,\gamma + (\epsilon + (1-\epsilon)\,(1-\mu))\,(1-\gamma)}$$

Again, by multiplying with the respective common denominator and simplifying (using in the process that $\gamma \neq 0, 1$ and $\mu \neq 1$) it ultimately follows that:

$$\begin{aligned} & 0 < \pi_E(G|\tilde{g}) - P(G|\tilde{g}) \\ \Leftrightarrow\ & 0 < \frac{(1-\epsilon)\gamma}{(1-\epsilon)\gamma + \epsilon(1-\gamma)} - \frac{((1-\epsilon) + \epsilon\,(1-\mu))\,\gamma}{((1-\epsilon) + \epsilon\,(1-\mu))\,\gamma + (\epsilon + (1-\epsilon)\,(1-\mu))\,(1-\gamma)} \\ \Leftrightarrow\ & 0 < \gamma(1-\gamma)(1-\mu)\,(1-2\epsilon) \\ \Leftrightarrow\ & \epsilon < \frac{1}{2} \end{aligned}$$

Alternatively, it is possible to use the expressions for $P(B|\tilde{g}) - \pi_E(B|\tilde{g})$. Finally, with the assumption that $\epsilon < 1/2$ and also $\alpha(G) > \alpha(B)$:

$$\begin{aligned} & 0 < \mathbb{E}_{\pi_E}(\alpha|\tilde{g}) - \mathbb{E}_P(\alpha|\tilde{g}) \\ \Leftrightarrow\ & 0 < \gamma(1-\gamma)(1-\mu)\,(1-2\epsilon)\,(\alpha(G) - \alpha(B)) \\ \Leftrightarrow\ & 0 < \alpha(G) - \alpha(B) \end{aligned}$$

□

Furthermore, independent of the limitation the realist is indeed realistic:

Remark 3. *It holds that* $\pi_E(G|\tilde{b}) = P(G|\tilde{b})$ *and* $\pi_E(B|\tilde{b}) = P(B|\tilde{b})$ *as well as* $\mathbb{E}_{\pi_E}(\alpha|\tilde{b}) = \mathbb{E}_P(\alpha|\tilde{b})$.

Proof. Remember that $\pi_E(Q|\tilde{q}) = P(Q|q)$ (cf. the ρ_q expressions) and then use:

$$P(G|\tilde{b}) = \frac{\epsilon\mu\gamma}{\epsilon\mu\gamma + (1-\epsilon)\mu(1-\gamma)}$$

By simplifying (using in the process that $\mu \neq 0$) it directly follows that:

$$\pi_E(G|\tilde{b}) - P(G|\tilde{b}) = \frac{\epsilon\gamma}{\epsilon\gamma + (1-\epsilon)(1-\gamma)} - \frac{\epsilon\mu\gamma}{\epsilon\mu\gamma + (1-\epsilon)\mu(1-\gamma)} = 0$$

Alternatively, it is possible to use the expressions for $\pi_E(B|\tilde{b}) - P(B|\tilde{b})$. Finally, $\mathbb{E}_{\pi_E}(\alpha|\tilde{b}) - \mathbb{E}_P(\alpha|\tilde{b}) = \sum_{Q\in\{G,B\}} \alpha(Q)\left(\pi_E(Q|\tilde{b}) - P(Q|\tilde{b})\right) = 0$. □

Appendix C. Equilibria Calculations

As mentioned in the description of the equilibria, this part of the appendix contains the calculations for each step of the decision problem as well as the equilibria.[35] Considering that, let us consult the timing of the choices made by the players:

Step	Player	Choice
(0)	Nature	Signal
(1)	Entrepreneur	Initial Effort
(2)	Venture Capitalist	Terms of Funding
(3)	Entrepreneur	Source of Funding
(4)	Entrepreneur	Additional Effort

C.1. **Step (4).** Note that funding via the bank is a unilateral choice (in this model) while funding via the venture capitalist necessitates a matching strategy of both the entrepreneur and the venture capitalist. Therefore, the following choice by the entrepreneur about the additional effort depends in parts on both the terms and the source of funding as well as the initial effort:

If the entrepreneur chose the funding via the venture capitalist and vice versa, i.e. $s = VC$ and $\lambda \in [0,1]$, then $\Pi_E^{\tilde{q}} = (1-\lambda)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0 e_0 + \beta_1 e_1 + \beta_2 \bar{e}\right) - e_0^2/2 - e_1^2/2$ as the relevant objective function implies the optimal choice $e_1^* = (1-\lambda)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1$. In the scenario of a mismatch or the outside option, i.e. $s = VC$ and $\lambda = \emptyset$ or $s = 0$, the objective function $\Pi_E^{\tilde{q}} = -e_0^2/2 - e_1^2/2$ implies the optimal choice $e_1^* = 0$.

If the entrepreneur chose the bank, i.e. $s = B$, then the objective function $\Pi_E^{\tilde{q}} = \sum_{Q\in\{G,B\}} \pi_E(Q|\tilde{q}) \max\{\alpha(Q)\left(\beta_0 e_0 + \beta_1 e_1\right) - rI, 0\} - e_0^2/2 - e_1^2/2$ dictates that the optimization takes place on up to three intervals depending on the value of the two efforts. Define $J_Q(e_0) := (rI/\alpha(Q) - \beta_0 e_0)/\beta_1$ for $Q \in \{G,B\}$, which mark the minimal values for the additional effort by the entrepreneur required for a

[35] Specifically, Perfect Bayesian Equilibrium with a focus on pure strategies. Also, note that a number of statements in this appendix directly follow from FOCs and SOCs, which usually only require minimal explanation and justification - with the obvious ones actually being omitted.

non-negative return in the case of the respective project quality. With this definition, the objective function as a function of the additional effort is given by the following:

$$\Pi_E^{\tilde{q}} = -\frac{e_0^2}{2} - \frac{e_1^2}{2} + \begin{cases} 0 & e_1 \in [0, J_G(e_0)] \\ \pi_E(G|\tilde{q})\left(\alpha(G)\left(\beta_0 e_0 + \beta_1 e_1\right) - rI\right) & e_1 \in [J_G(e_0), J_B(e_0)] \\ \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0 e_0 + \beta_1 e_1\right) - rI & e_1 \in [J_B(e_0), +\infty) \end{cases}$$

It seems that this interval-wise objective function lacks a proper definition on the intermediate borders $J_G(e_0)$ and $J_B(e_0)$, but by definition of these values the respective functions actually coincide on these transition points. Note that without further restrictions it is possible that the optimal choice of an interval does not belong to that specific interval. It is also not obvious yet which of the intervals contains the overall optimal choice. In the following, let us therefore first determine the candidates for optimal choices on each interval with the specific requirements for them to exist within their interval:

$$e_1^{c_1} = \begin{cases} 0 & 0 < J_G(e_0) \\ J_G(e_0) & J_G(e_0) \leq 0 \end{cases}$$

$$e_1^{c_2} = \begin{cases} J_G(e_0) & \pi_E(G|\tilde{q})\alpha(G)\beta_1 \leq J_G(e_0) \\ \pi_E(G|\tilde{q})\alpha(G)\beta_1 & J_G(e_0) < \pi_E(G|\tilde{q})\alpha(G)\beta_1 < J_B(e_0) \\ J_B(e_0) & J_B(e_0) \leq \pi_E(G|\tilde{q})\alpha(G)\beta_1 \end{cases}$$

$$e_1^{c_3} = \begin{cases} J_B(e_0) & \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 \leq J_B(e_0) \\ \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 & J_B(e_0) < \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 \end{cases}$$

Again, a number of local optimal choices (on the boundaries) are not global ones. Note that the objective function takes the form of interval-wise quadratic functions of the additional effort while the corresponding intervals are such that each pair of functions is identical on the border of those intervals. Any local optimal choice on the border can therefore only be a global one when it is also the local one on the other corresponding interval. Consequently, $J_G(e_0)$ only appears when $J_G(e_0) \leq 0$ and $\pi_E(G|\tilde{q})\alpha(G)\beta_1 \leq J_G(e_0)$, thus $\pi_E(G|\tilde{q})\alpha(G)\beta_1 \leq J_G(e_0) \leq 0$, which is a contradiction due to our assumption on the parameters. Similarly, $J_B(e_0)$ only matters when $J_B(e_0) \leq \pi_E(G|\tilde{q})\alpha(G)\beta_1$ and $\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 \leq J_B(e_0)$, thus $\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 \leq J_B(e_0) \leq \pi_E(G|\tilde{q})\alpha(G)\beta_1$, which is another contradiction. Finally, by combining the conditions for the remaining three possible optimal choices, the following three candidates emerge (with still potentially overlapping conditions)

$$\begin{aligned} e_1^{c_1} &= 0 && \text{for } 0 < J_G(e_0) \\ e_1^{c_2} &= \pi_E(G|\tilde{q})\alpha(G)\beta_1 && \text{for } J_G(e_0) < \pi_E(G|\tilde{q})\alpha(G)\beta_1 < J_B(e_0) \\ e_1^{c_3} &= \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 && \text{for } J_B(e_0) < \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 \end{aligned}$$

or alternatively, in terms of the initial effort, the three candidates are given by

$$\begin{aligned} e_1^{c_1} &= 0 && \text{for } e_0 < \xi_1^{\max} \\ e_1^{c_2} &= \pi_E(G|\tilde{q})\alpha(G)\beta_1 && \text{for } \xi_2^{\min} < e_0 < \xi_2^{\max} \\ e_1^{c_3} &= \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 && \text{for } \xi_3^{\min} < e_0 \end{aligned}$$

with $\xi_1^{\max} := \frac{rI}{\alpha(G)\beta_0}$, $\xi_2^{\min} := \frac{rI}{\alpha(G)\beta_0} - \frac{\pi_E(G|\tilde{q})\alpha(G)\beta_1^2}{\beta_0}$, $\xi_2^{\max} := \frac{rI}{\alpha(B)\beta_0} - \frac{\pi_E(G|\tilde{q})\alpha(G)\beta_1^2}{\beta_0}$, and $\xi_3^{\min} := \frac{rI}{\alpha(B)\beta_0} - \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}{\beta_0}$, and refer to the candidates as B.1, B.2, and B.3 for

the remainder of this paper. Note that so far this only characterizes potential optimal choices and the conditions for their existence. In order to find the unique optimal choice, let us take a look at the deltas of the expected profits corresponding to the three candidates while using the notation $\Delta(X,Y) := \Pi_E^{\tilde{q}}(e_1 = X) - \Pi_E^{\tilde{q}}(e_1 = Y)$

- $\Delta(B.2, B.1) = \pi_E(G|\tilde{q})\left(\alpha(G)\beta_0 e_0 - rI\right) + \frac{\left(\pi_E(G|\tilde{q})\alpha(G)\beta_1\right)^2}{2}$
- $\Delta(B.3, B.1) = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0 e_0 - rI + \frac{\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1\right)^2}{2}$
- $\Delta(B.3, B.2) = \pi_E(B|\tilde{q})\left(\alpha(B)\beta_0 e_0 - rI\right) + \frac{\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2 - \pi_E(G|\tilde{q})^2\alpha(G)^2\right)\beta_1^2}{2}$

and with this formulate conditions for a ranking depending on the initial effort

- $\Delta(B.2, B.1) > 0$ if and only if $e_0 > \xi_{2,1}^{\min}$
- $\Delta(B.3, B.1) > 0$ if and only if $e_0 > \xi_{3,1}^{\min}$
- $\Delta(B.3, B.2) > 0$ if and only if $e_0 > \xi_{3,2}^{\min}$

with $\xi_{2,1}^{\min} = \xi_{1,2}^{\max} := \frac{rI}{\alpha(G)\beta_0} - \frac{\pi_E(G|\tilde{q})\alpha(G)\beta_1^2}{2\beta_0}$, $\xi_{3,1}^{\min} = \xi_{1,3}^{\max} := \frac{rI}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0} - \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}{2\beta_0}$, and $\xi_{3,2}^{\min} = \xi_{2,3}^{\max} := \frac{rI}{\alpha(B)\beta_0} - \frac{\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) + \pi_E(G|\tilde{q})\alpha(G)\right)\beta_1^2}{2\beta_0}$. Comparing all of these critical values with each other yields the following relations[36]:

(A) $0 < \xi_1^{\max}$
(B) $\xi_2^{\min} < \xi_{2,1}^{\min} < \xi_1^{\max}$
(C) $\xi_2^{\min} < \xi_2^{\max}$
(D) $\xi_3^{\min} < \xi_{3,2}^{\min} < \xi_2^{\max}$
(E) If $\xi_{3,2}^{\min} \leq \xi_{2,1}^{\min}$, then $\xi_3^{\min} < \xi_{3,1}^{\min} < \xi_1^{\max}$

For these five statements let us compute the differences between the expressions and then show the claims using that all (involved) parameters are strictly positive together with the fact that $\alpha(G) > \alpha(B)$ and (thus) $\alpha(G) > \mathbb{E}_{\pi_E}(\alpha|\tilde{q}) > \alpha(B)$:

$$\text{(Ad A)} \qquad \xi_1^{\max} = \frac{rI}{\alpha(G)\beta_0} > 0$$

$$\text{(Ad B)} \qquad \xi_{2,1}^{\min} - \xi_2^{\min} = \xi_1^{\max} - \xi_{2,1}^{\min} = \frac{\pi_E(G|\tilde{q})\alpha(G)\beta_1^2}{2\beta_0} > 0$$

$$\text{(Ad C)} \qquad \xi_2^{\max} - \xi_2^{\min} = \frac{rI}{\beta_0}\left(\frac{1}{\alpha(B)} - \frac{1}{\alpha(G)}\right) > 0$$

$$\text{(Ad D)} \qquad \xi_{3,2}^{\min} - \xi_3^{\min} = \xi_2^{\max} - \xi_{3,2}^{\min} = \frac{\pi_E(B|\tilde{q})\alpha(B)\beta_1^2}{2\beta_0} > 0$$

$$\text{(Ad E.I)} \qquad \xi_{3,1}^{\min} - \xi_3^{\min} = \xi_{2,1}^{\min} - \xi_{3,2}^{\min} + \frac{rI}{\beta_0}\left(\frac{1}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})} - \frac{1}{\alpha(G)}\right) > 0$$

$$\text{(Ad E.II)} \qquad \xi_1^{\max} - \xi_{3,1}^{\min} = \xi_{2,1}^{\min} - \xi_{3,2}^{\min} + \frac{rI}{\beta_0}\left(\frac{1}{\alpha(B)} - \frac{1}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})}\right) > 0$$

[36]Each of those relations contains an intuitive interpretation, stating that the interval where

(A) B.1 exists contains other elements besides zero;
(B) B.1 and B.2 exist overlap and B.2 starts dominating B.1 before B.1 stops existing;
(C) B.2 exists contains some elements (as long as these are positive);
(D) B.2 and B.3 exist overlap and B.3 starts dominating B.2 before B.2 stops existing;
(E) B.1 and B.3 exist overlap and B.3 starts dominating B.1 before B.1 stops existing in the case that B.3 starts dominating B.2 before B.2 starts dominating B.1.

Using these five statements in combination with the definitions of the critical values implies that there always exists a partition of the set of potential initial effort consisting of up to three sets corresponding to the three previously introduced candidates for the additional effort and let us denote these sets by $X_{B.1}, X_{B.2}, X_{B.3}$. While the exact partition necessarily depends on the parameters, the set of partitions is ultimately limited to a set of four elements (using all the previous statements)

$$\begin{aligned}
\mathcal{P} :=& \{\{X_{B.3}\}, \{X_{B.1}, X_{B.3}\}, \{X_{B.2}, X_{B.3}\}, \{X_{B.1}, X_{B.2}, X_{B.3}\}\} \\
\text{where } \{X_{B.3}\} =& \{[0, +\infty)\} \\
\{X_{B.1}, X_{B.3}\} =& \{[0, \xi_{3,1}^{\min}], (\xi_{3,1}^{\min}, +\infty)\} \\
\{X_{B.2}, X_{B.3}\} =& \{[0, \xi_{3,2}^{\min}], (\xi_{3,2}^{\min}, +\infty)\} \\
\{X_{B.1}, X_{B.2}, X_{B.3}\} =& \{[0, \xi_{2,1}^{\min}], (\xi_{2,1}^{\min}, \xi_{3,2}^{\min}], (\xi_{3,2}^{\min}, +\infty)\}
\end{aligned}$$

with the understanding that an $X_{B.\cdot}$ not part of the partition equals the empty set. Thus, the optimal choice for additional effort depending on the initial effort is given by the following:

$$e_1^* = \begin{cases} 0 & e_0 \in X_{B.1} \\ \pi_E(G|\tilde{q})\alpha(G)\beta_1 & e_0 \in X_{B.2} \\ \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 & e_0 \in X_{B.3} \end{cases}$$

C.2. **Step (3).** The entrepreneur now compares three options when choosing the source of funding, namely bank, venture capitalist, and no funding, i.e. no project. As the previous step of our calculation revealed, a total of five classes, not cases, of possible equilibria based on these three options emerge (three classes for the bank, one for the venture capitalist, and one for no funding - including the mismatch).[37] Consequently, the following five expressions of expected profit matter for this step:

$$\begin{aligned}
&(0)\ \Pi_E^{\tilde{q}} = -\tfrac{e_0^2}{2} \\
&(B.1)\ \Pi_E^{\tilde{q}} = -\tfrac{e_0^2}{2} \\
&(B.2)\ \Pi_E^{\tilde{q}} = \pi_E(G|\tilde{q})\left(\alpha(G)\beta_0 e_0 - rI\right) - \tfrac{e_0^2}{2} + \tfrac{(\pi_E(G|\tilde{q})\alpha(G)\beta_1)^2}{2} \\
&(B.3)\ \Pi_E^{\tilde{q}} = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0 e_0 - rI - \tfrac{e_0^2}{2} + \tfrac{\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1\right)^2}{2} \\
&(VC)\ \Pi_E^{\tilde{q}} = (1-\lambda)\,\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0 e_0 + \beta_2 \bar{e}\right) - \tfrac{e_0^2}{2} + \tfrac{\left((1-\lambda)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1\right)^2}{2}
\end{aligned}$$

However, similar to before, the choice about the source of funding depends in part on the terms of funding as well as the initial effort:

In particular, the terms of funding determine which classes of equilibria are viable. If the venture capitalist offered funding, i.e. $\lambda \in [0, 1]$, then three of them qualify - one for the bank, one for the venture capitalist, and one for no funding. Furthermore, if the venture capitalist did not offer funding, i.e. $\lambda = \emptyset$, then only two remain - without the one for the venture capitalist. Finally, the different partitions together with the initial effort determine the relevant bank class, as shown before.

Also, in terms of (expected) profit for the entrepreneur, the option of no funding corresponds to both the scenario of funding by the bank in case the entrepreneur expects no (re-) payment of the funds for the two project qualities as well as that of funding by the venture capitalist in case the venture capitalist demands everything. While the three belong to different classes of equilibria, it does not matter for the

[37] A class of equilibria in this context refers to equilibria with equivalent pay-off for each player.

optimal choice by the entrepreneur. As a consequence, the relations between the critical values shown before guarantee that the analysis in the previous step with respect to the partition and the initial effort determined not only the candidate with respect to the bank but includes as a byproduct the comparison with the option of no funding and with the venture capitalist in the special case of a full transfer. Therefore, it suffices to only consider up to two expressions in this step, namely one for the bank (including the two bank-equivalent scenarios) and potentially another one for the venture capitalist. In the case where the venture capitalist offered funding, one of the following three expressions determines the optimal choice, using the notation for the delta as before (while also extending it to the candidate $(1-\lambda)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1$, where entrepreneur and venture capitalist chose each other, simply referred to as VC):

$$\Delta(B.1, VC) = -(1-\lambda)\,\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\,(\beta_0 e_0 + \beta_2 \bar{e}) - \frac{\left((1-\lambda)\,\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1\right)^2}{2}$$

$$\Delta(B.2, VC) = \left(\lambda \mathbb{E}_{\pi_E}(\alpha|\tilde{q}) - \pi_E(B|\tilde{q})\alpha(B)\right)\beta_0 e_0 + (\lambda-1)\,\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_2\bar{e}$$
$$- \pi_E(G|\tilde{q})rI + \frac{1}{2}\left(\lambda\,(2-\lambda)\,\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2 - \left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2 - \pi_E(G|\tilde{q})^2\alpha(G)^2\right)\right)\beta_1^2$$

$$\Delta(B.3, VC) = \lambda \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0 e_0 + (\lambda-1)\,\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_2\bar{e} - rI + \frac{\lambda\,(2-\lambda)\,\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\beta_1^2}{2}$$

Note that without further assumptions $\Delta(B.1, VC) \leq 0$ and $\Delta(B.1, VC) = 0$ if and only if $\lambda = 1$. Furthermore, in the case of $\lambda = 1$ both $\Delta(B.2, VC) = \Delta(B.2, B.1)$ and $\Delta(B.3, VC) = \Delta(B.3, B.1)$ using the equivalence described before.

Finally, let us formulate the optimal choice depending on the terms of funding as well as the initial effort. If the venture capitalist chose $\lambda = \emptyset$, then clearly $s^* = B$. Otherwise, with $\lambda \in [0, 1]$, the partition and the initial effort tell us which options to compare:

$$s^* = \begin{cases} VC & \text{for } \begin{cases} e_0 \in X_{B.1} \\ e_0 \in X_{B.2} \text{ and } \Delta(B.2, VC) \leq 0 \\ e_0 \in X_{B.3} \text{ and } \Delta(B.3, VC) \leq 0 \end{cases} \\ B & \text{for } \begin{cases} e_0 \in X_{B.2} \text{ and } \Delta(B.2, VC) > 0 \\ e_0 \in X_{B.3} \text{ and } \Delta(B.3, VC) > 0 \end{cases} \end{cases}$$

Moreover, in the special case of $\lambda = 1$ the equalities with respect to the difference in expected profit, $\Delta(B.2, VC) = \Delta(B.2, B.1)$ and $\Delta(B.3, VC) = \Delta(B.3, B.1)$, imply that the venture capitalist only materializes as the optimal choice when the class of no funding dominates among the bank alternatives.

C.3. **Step (2).** While in the previous steps the players never needed to take the participation of the other player into consideration, this changes in this step with the choice by the venture capitalist. In order to avoid complicating the computation, the analysis is essentially split into two parts. First, the optimal choice is derived while ignoring any participation constraint. Second, this first candidate is modified in those scenarios where these conditions play a role.

As determined before, if the venture capitalist would choose $\lambda = \emptyset$, then $s^* = B$ (with corresponding e_1^*) and therefore $\Pi_{VC}^{e_0} = 0$. This outside option sets the bar for the potential alternatives $\lambda \in [0, 1]$.

In order to compare the outside option with the other alternatives it is necessary to first find the alternatives' optimal choice. The objective function for $\lambda \in [0,1]$ is given by $\Pi_{VC}^{e_0} = \lambda \left(\mathbb{E}_\pi(\alpha|e_0)\left(\beta_0 e_0 + \beta_2 \bar{e}\right) + \mathbb{E}_\pi(\alpha \mathbb{E}_{\pi_E}(\alpha)|e_0)\left(1-\lambda\right)\beta_1^2\right) - \bar{r}I - \bar{e}^2/2$ which yields $\left(\mathbb{E}_\pi(\alpha|e_0)\left(\beta_0 e_0 + \beta_2 \bar{e}\right) + \mathbb{E}_\pi(\alpha \mathbb{E}_{\pi_E}(\alpha)|e_0)\right)\beta_1^2\right) / \left(2\mathbb{E}_\pi(\alpha \mathbb{E}_{\pi_E}(\alpha)|e_0)\beta_1^2\right)$ as the corresponding candidate (originating from the first-order-condition). By using the assumptions on the parameters it follows that this candidate is greater than zero, but it is only less or equal to one with $\mathbb{E}_\pi(\alpha|e_0)\left(\beta_0 e_0 + \beta_2 \bar{e}\right) \leq \mathbb{E}_\pi(\alpha \mathbb{E}_{\pi_E}(\alpha)|e_0))\beta_1^2$ (it is also always greater than 1/2 due to the aforementioned assumption). Thus:

$$\lambda^c = \begin{cases} \lambda_s & \frac{\mathbb{E}_\pi(\alpha|e_0)(\beta_0 e_0 + \beta_2 \bar{e})}{\mathbb{E}_\pi(\alpha \mathbb{E}_{\pi_E}(\alpha)|e_0)\beta_1^2} < 1 \\ 1 & 1 \leq \frac{\mathbb{E}_\pi(\alpha|e_0)(\beta_0 e_0 + \beta_2 \bar{e})}{\mathbb{E}_\pi(\alpha \mathbb{E}_{\pi_E}(\alpha)|e_0)\beta_1^2} \end{cases}$$

$$\text{with } \lambda_s := \frac{\mathbb{E}_\pi(\alpha|e_0)\left(\beta_0 e_0 + \beta_2 \bar{e}\right) + \mathbb{E}_\pi(\alpha \mathbb{E}_{\pi_E}(\alpha)|e_0)\beta_1^2}{2\mathbb{E}_\pi(\alpha \mathbb{E}_{\pi_E}(\alpha)|e_0)\beta_1^2}$$

Up until this point, it was not yet necessary to limit the category of equilibria under consideration. However, the following analysis depends (among other things) on the perceived signal of the quality by the entrepreneur. Let us therefore act on our initial commitment to separating equilibria to avoid unnecessary calculations. Additionally, this simplifies the project share candidate when simultaneously re-formulating its cases in terms of the initial effort:

$$\lambda^c = \begin{cases} \lambda_s & e_0 < \xi_{s,1}^{\max} \\ 1 & \xi_{s,1}^{\max} \leq e_0 \end{cases}$$

$$\text{with } \lambda_s = \frac{\beta_0 e_0 + \beta_2 \bar{e} + \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}{2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}$$

$$\text{and } \xi_{s,1}^{\max} := \frac{1}{\beta_0}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - \beta_2 \bar{e}\right)$$

Next, let us check the participation of the venture capitalist while also preparing the set of possible concessions to the entrepreneur for the next step by setting the expected profit equal to zero and then solving for the project share, which yields

$$\lambda_\pm^{VC} = \lambda_s \pm \sqrt{\Delta_{VC}}$$

$$\text{with } \Delta_{VC} := \lambda_s^2 - \psi$$

$$\text{and with } \psi := \frac{\bar{e}^2 + 2\bar{r}I}{2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}.$$

As a consequence, depending on the parameters three different scenarios might arise:

(1) $\Delta_{VC} < 0$:
   - $\Pi_{VC}^{e_0} < 0$ for all λ

(2) $\Delta_{VC} = 0$:
   - $\Pi_{VC}^{e_0} < 0$ for $\lambda \neq \lambda_s$
   - $\Pi_{VC}^{e_0} = 0$ for $\lambda = \lambda_s$

(3) $\Delta_{VC} > 0$:
   - $\Pi_{VC}^{e_0} < 0$ for $\lambda \notin [\lambda_-^{VC}, \lambda_+^{VC}]$
   - $\Pi_{VC}^{e_0} = 0$ for $\lambda \in \{\lambda_-^{VC}, \lambda_+^{VC}\}$
   - $\Pi_{VC}^{e_0} > 0$ for $\lambda \in (\lambda_-^{VC}, \lambda_+^{VC})$

In other words, in the first case the expected profit is always negative implying that the outside option dominates any candidate in the interval - removing the need for

a comparison with the participation of the entrepreneur. Next, in the second case only the optimal choice is on par with the outside option, making it the sole element in the set of possible concessions, which therefore necessitates only one comparison. Finally, in the last of the three cases there exists a proper interval where its elements dominate the outside option; thereby forming a proper set of possible concessions that requires further comparisons.

Note that in the second case, the comparison is only necessary for those scenarios where the optimal choice given by the first-order-condition lies in the unit interval. Otherwise, the expected profit on the interval is always strictly negative and therefore less than that of the outside option. Similarly, in the third case, a lower limit λ_{-}^{VC} greater than one removes the need for comparison. Furthermore, the upper limit λ_{+}^{VC} never actually becomes relevant as the participation constraint of the entrepreneur itself comes in the form of an upper limit. Therefore, as long as it is greater than the optimal choice, the venture capitalist simply sticks to that. Define $\lambda_{VC} := \lambda_{-}^{VC}$ for the actual comparison.

Let us now present the initial candidate for the optimal choice of this step, completely ignoring the participation constraint of the entrepreneur for a moment,

(1) $\lambda_s < 1$
  (1) $\Delta_{VC} < 0$
    $\lambda^c = \emptyset$
  (2) $\Delta_{VC} \geq 0$
    $\lambda^c = \lambda_s$
(2) $\lambda_s \geq 1$
  (1) $\Delta_{VC} < 0 \vee \lambda_{VC} > 1$
    $\lambda^c = \emptyset$
  (2) $\Delta_{VC} \geq 0 \wedge \lambda_{VC} \leq 1$
    $\lambda^c = 1$

and let us furthermore formulate this candidate in terms of the initial effort by employing a number of critical values (with the familiar notation of $\xi_{xyz}^{\min / \max}$)[38]

(1) $e_0 \in [0, \xi_{s,1}^{\max})$
  (1) $e_0 \in [0, \xi_{VC}^{\min_1})$
    $\lambda^c = \emptyset$
  (2) $e_0 \in [\xi_{VC}^{\min_1}, +\infty)$
    $\lambda^c = \lambda_s$
(2) $e_0 \in [\xi_{s,1}^{\max}, +\infty)$
  (1) $e_0 \in [0, \xi_{VC}^{\min_2})$
    $\lambda^c = \emptyset$
  (2) $e_0 \in [\xi_{VC}^{\min_2}, +\infty)$
    $\lambda^c = 1$

with the two critical values for the initial effort given by the following expressions:

$$\xi_{VC}^{\min_1} := \frac{1}{\beta_0}\left(-\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - \beta_2\bar{e} + \beta_1\sqrt{2\frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})}{\mathbb{E}_P(\alpha|\tilde{q})}(\bar{e}^2 + 2\bar{r}I)}\right)$$

$$\xi_{VC}^{\min_2} := \frac{1}{\beta_0}\left(-\beta_2\bar{e} + \frac{\bar{e}^2 + 2\bar{r}I}{2\mathbb{E}_P(\alpha|\tilde{q})}\right)$$

[38]Note that the equivalence of $\Delta_{VC} \geq 0 \wedge \lambda_{VC} \leq 1$ and $e_0 \geq \xi_{VC}^{\min_2}$ uses $\lambda_s \geq 1$.

In order to present the initial candidate for the optimal choice in a compact manner, let us define yet another partition of the initial effort space into three sets:

$$\begin{aligned}
Z_s^c &:= [0, \xi_{s,1}^{\max}) \cap [\xi_{VC}^{\min_1}, +\infty) \\
Z_1^c &:= [\xi_{s,1}^{\max}, +\infty) \cap [\xi_{VC}^{\min_2}, +\infty) \\
Z_\emptyset^c &:= \left([0, \xi_{s,1}^{\max}) \cap [0, \xi_{VC}^{\min_1})\right) \cup \left([\xi_{s,1}^{\max}, +\infty) \cap [0, \xi_{VC}^{\min_2})\right)
\end{aligned}$$

Finally, ignoring the participation constraint of the entrepreneur, the initial candidate for the optimal choice for the project share depending on the initial effort is given by:

$$\lambda^c = \begin{cases} \lambda_s & e_0 \in Z_s^c \\ 1 & e_0 \in Z_1^c \\ \emptyset & e_0 \in Z_\emptyset^c \end{cases}$$

With the varying set of possible concessions by the venture capitalist in mind, let us turn to the critical values for the participation of the entrepreneur. Now, the comparison necessarily depends on the available alternative for the entrepreneur, i.e. the respective class of funding by the bank, which itself depends on the partition. Note that in the case of B.1 the entrepreneur always chooses the venture capitalist, following the previously determined delta. For the cases B.2 and B.3 let us compute the set of project shares that guarantee the participation of the entrepreneur by setting the delta equal to zero and then solving for the project share, which yields

$$\begin{aligned}
\lambda_\pm^{B.\cdot} =& 2\lambda_s \pm \sqrt{\Delta_{B.\cdot}} \\
\text{with } \Delta_{B.\cdot} :=& 4\lambda_s^2 - \eta_{B.\cdot} \\
\text{and with } \eta_{B.\cdot} :=& \frac{D_{B.\cdot}}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2 \beta_1^2}
\end{aligned}$$

with B.$\cdot$ denoting the relevant one of the bank scenarios, i.e. either B.2 or B.3, where $D_{B.2} := D_{B.3} + 2\pi_E(B|\tilde{q})\left(\alpha(B)\beta_0 e_0 - rI\right) + \left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2 - \pi_E(G|\tilde{q})^2\alpha(G)^2\right)\beta_1^2$ and $D_{B.3} := 2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_2\bar{e} + rI\right)$. Note, as shown previously, it holds that $2\lambda_s > 1$. Similar to before, depending on the parameters three different scenarios might arise:

(1) $\Delta_{B.\cdot} < 0$:
   - $\Delta(B.\cdot, VC) < 0$ for all λ

(2) $\Delta_{B.\cdot} = 0$:
   - $\Delta(B.\cdot, VC) < 0$ for $\lambda \neq 2\lambda_s$
   - $\Delta(B.\cdot, VC) = 0$ for $\lambda = 2\lambda_s$

(3) $\Delta_{B.\cdot} > 0$:
   - $\Delta(B.\cdot, VC) < 0$ for $\lambda \notin [\lambda_-^{B.\cdot}, \lambda_+^{B.\cdot}]$
   - $\Delta(B.\cdot, VC) = 0$ for $\lambda \in \{\lambda_-^{B.\cdot}, \lambda_+^{B.\cdot}\}$
   - $\Delta(B.\cdot, VC) > 0$ for $\lambda \in (\lambda_-^{B.\cdot}, \lambda_+^{B.\cdot})$

Consequently, in the first case the expected profit with the venture capitalist always trumps that with the bank, which always guarantees the entrepreneur's participation. Also, the same applies to the second case due to our assumption on breaking ties. However, in the third case there exists a proper interval where this relation reverses. This case and interval require further comparison with the previously determined critical values for the venture capitalist.

In case $\lambda_-^{B.\cdot} \geq 1$ the third case also requires no further comparison. Moreover, the upper limit $\lambda_+^{B.\cdot}$ stays irrelevant due to the relation $2\lambda_s > 1$. Define $\lambda_{B.\cdot} := \lambda_-^{B.\cdot}$ for the actual comparison.

Let us now present the modified candidate for the optimal choice of this step, specifically incorporating the participation constraint of the entrepreneur this time:[39]

(1) $\lambda^c = \lambda_s$
  (1) $B.\cdot = B.1$
    $\lambda^* = \lambda_s$
  (2) $B.\cdot \neq B.1$
    (1) $\Delta_{B.\cdot} \leq 0$
      $\lambda^* = \lambda_s$
    (2) $\Delta_{B.\cdot} > 0$
      (1) $\lambda_s \leq \lambda_{B.\cdot}$
        $\lambda^* = \lambda_s$
      (2) $\lambda_{B.\cdot} < \lambda_s$
        (1) $\lambda_{B.\cdot} < \lambda_{VC}$
          $\lambda^* = \emptyset$
        (2) $\lambda_{VC} \leq \lambda_{B.\cdot}$
          $\lambda^* = \lambda_{B.\cdot}$

(2) $\lambda^c = 1$
  (1) $B.\cdot = B.1$
    $\lambda^* = 1$
  (2) $B.\cdot \neq B.1$
    (1) $\Delta_{B.\cdot} \leq 0$
      $\lambda^* = 1$
    (2) $\Delta_{B.\cdot} > 0$
      (1) $1 \leq \lambda_{B.\cdot}$
        $\lambda^* = 1$
      (2) $\lambda_{B.\cdot} < 1$
        (1) $\lambda_{B.\cdot} < \lambda_{VC}$
          $\lambda^* = \emptyset$
        (2) $\lambda_{VC} \leq \lambda_{B.\cdot}$
          $\lambda^* = \lambda_{B.\cdot}$

(3) $\lambda^c = \emptyset$
  $\lambda^* = \emptyset$

Note that the relation between $\lambda_{B.\cdot}$ and λ_{VC} yields two different optimal choices for the venture capitalist but without any effect on (the profit of) the entrepreneur. Whenever this relation becomes actually relevant, the venture capitalist either provides no funding, making the bank the only option (barring the outside option), or chooses a project share that emulates (the profit of) the bank option. Thus, while the two scenarios differ significantly from the perspective of the venture capitalist, the corresponding objective functions of the entrepreneur coincide. As a consequence, it is not necessary to distinguish between the two scenarios for the computation of the initial effort in the next step. The full strategy tuple of the equilibria will differ, but not the entry with respect to the initial effort.

[39] It is noteworthy to point out that for $\lambda_s < 1$ it holds that $\Delta_{VC} \geq 0$ implies $\lambda_{VC} < 1$.

Recall now that the partition and the initial effort determined the class of funding by the bank that competes with the venture capitalist and that in the case of B.1 the comparison always favors the venture capitalist. With this in mind, let us formulate the candidate in terms of the initial effort by using the respective partitions as well as employing a number of critical values again (also using the notation of $\xi_{xyz}^{\min\,/\,\max}$)

(1) $e_0 \in Z_s^c$
  (1) $e_0 \in X_{B.1}$
    $\lambda^* = \lambda_s$
  (2) $e_0 \in X_{B.\cdot} \neq X_{B.1}$
    (1) $e_0 \in [0, \xi_{B.\cdot}^{\max}]$
      $\lambda^* = \lambda_s$
    (2) $e_0 \in (\xi_{B.\cdot}^{\max}, +\infty)$
      (1) $e_0 \in [0, \xi_{B.\cdot,s}^{\max}]$
        $\lambda^* = \lambda_s$
      (2) $e_0 \in (\xi_{B.\cdot,s}^{\max}, +\infty)$
        (1) $e_0 \in Y_{\neg B.\cdot}$
          $\lambda^* = \emptyset$
        (2) $e_0 \in Y_{B.\cdot}$
          $\lambda^* = \lambda_{B.\cdot}$

(2) $e_0 \in Z_1^c$
  (1) $e_0 \in X_{B.1}$
    $\lambda^* = 1$
  (2) $e_0 \in X_{B.\cdot} \neq X_{B.1}$
    (1) $e_0 \in [0, \xi_{B.\cdot}^{\max}]$
      $\lambda^* = 1$
    (2) $e_0 \in (\xi_{B.\cdot}^{\max}, +\infty)$
      (1) $e_0 \in [0, \xi_{B.\cdot,1}^{\max}]$
        $\lambda^* = 1$
      (2) $e_0 \in (\xi_{B.\cdot,1}^{\max}, +\infty)$
        (1) $e_0 \in Y_{\neg B.\cdot}$
          $\lambda^* = \emptyset$
        (2) $e_0 \in Y_{B.\cdot}$
          $\lambda^* = \lambda_{B.\cdot}$

(3) $e_0 \in Z_\emptyset^c$
  $\lambda^* = \emptyset$

with the critical values for the initial effort defined using additional auxiliary terms

$$\begin{aligned}
\xi_E &:= \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + \beta_2\bar{e} \\
\xi_{E,Q} &:= \pi_E(Q|\tilde{q})\alpha(Q)\beta_1^2 + \beta_2\bar{e} \\
\xi_M &:= \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_2\bar{e} + rI \\
\xi_{M,Q} &:= \pi_E(Q|\tilde{q})\left(\alpha(Q)\beta_2\bar{e} + rI\right)
\end{aligned}$$

- in order to simplify some of the convoluted expressions - in the following manner:[40]

[40]While it appears that there might be a reason to consider a counterpart to $\xi_{B.2,s}^{\max}$, i.e. $\xi_{B.2,s}^{\min}$, it is actually strictly negative using $\sqrt{x_1 + x_2} > \sqrt{x_1} + \sqrt{x_2}$ for $x_1, x_2 > 0$ and thus irrelevant.

$$\xi^{\max}_{B.2} := \frac{1}{\beta_0}\left(-\xi_{E,G} + \beta_1\sqrt{2\xi_{M,G}}\right)$$
$$\xi^{\max}_{B.3} := \frac{1}{\beta_0}\left(-\xi_{E} + \beta_1\sqrt{2\xi_{M}}\right)$$
$$\xi^{\max}_{B.2,s} := \frac{1}{\beta_0}\left(-\xi_{E,G} + \frac{1}{3}\pi_E(B|\tilde{q})\alpha(B)\beta_1^2 + \beta_1\sqrt{\frac{8}{3}\left(\xi_{M,G} + \frac{1}{6}\pi_E(B|\tilde{q})^2\alpha(B)^2\beta_1^2\right)}\right)$$
$$\xi^{\max}_{B.3,s} := \frac{1}{\beta_0}\left(-\xi_{E} + \beta_1\sqrt{\frac{8}{3}\xi_{M}}\right)$$
$$\xi^{\max}_{B.2,1} := \frac{1}{\beta_0}\left(\frac{rI}{\alpha(G)} - \frac{\pi_E(G|\tilde{q})\alpha(G)\beta_1^2}{2}\right)$$
$$\xi^{\max}_{B.3,1} := \frac{1}{\beta_0}\left(\frac{rI}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})} - \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}{2}\right)$$

Recall, in the overview $X_{B.\cdot}$ refers to the interval with respect to the initial effort where $B.\cdot$ dominates the other bank options. It is important to remember that $X_{B.\cdot} \neq X_{B.1}$ actually contains two cases of its own, $B.2$ and $B.3$ respectively. Also, for the relation between $\lambda_{B.\cdot}$ and λ_{VC} let $Y_{B.\cdot}$ resp. $Y_{\neg B.\cdot}$ denote the sets where $\lambda_{VC} \leq \lambda_{B.\cdot}$ resp. $\lambda_{B.\cdot} < \lambda_{VC}$. While the nature of the sets $X^P_{B.\cdot}$ follows from the previously presented partition with respect to the initial effort, the sets $Y_{B.\cdot}$ and $Y_{\neg B.\cdot}$ elude an elegant presentation, specifically an explicit specification with respect to the initial effort, and remain implicit conditions. At best, it is possible to re-phrase the condition $\lambda_{VC} \leq \lambda_{B.\cdot}$ for the numerical analysis as $\lambda_s \geq \sqrt{4\lambda_s^2 - \eta_{B.\cdot}} - \sqrt{\lambda_s^2 - \psi}$ (and similarly for $\lambda_{B.\cdot} < \lambda_{VC}$).

Finally, in order to properly present the modified candidate for the optimal choice, let us define yet another partition of the initial effort space into five sets:

$$\begin{aligned}
Z_s &:= Z_s^c \cap Y_{\neg B|s}\\
Z_1 &:= Z_1^c \cap Y_{\neg B|1}\\
Z_{B.2} &:= \cup_{i\in\{s,1\}}\left(Z_i^c \cap Y_{B.2|i}\right) \cap Y_{B.2}\\
Z_{B.3} &:= \cup_{i\in\{s,1\}}\left(Z_i^c \cap Y_{B.3|i}\right) \cap Y_{B.3}\\
Z_\emptyset &:= Z_\emptyset^c \cup \left(\cup_{i\in\{s,1\}}\left(Z_i^c \cap Y_{B|i}\right)\right)
\end{aligned}$$

using all previously defined sets, including the aforementioned $Y_{B.\cdot}$ and $Y_{\neg B.\cdot}$, and

$$\begin{aligned}
Y_{\neg B|s} &:= \cup_{i=1,2,3}\left(X_{B.i} \cap X_{\neg B.i|s}\right)\\
Y_{\neg B|1} &:= \cup_{i=1,2,3}\left(X_{B.i} \cap X_{\neg B.i|1}\right)\\
Y_{B.2|s} &:= X_{B.2} \cap (\xi^{\max}_{B.2}, +\infty) \cap (\xi^{\max}_{B.2,s}, +\infty)\\
Y_{B.2|1} &:= X_{B.2} \cap (\xi^{\max}_{B.2}, +\infty) \cap (\xi^{\max}_{B.2,1}, +\infty)\\
Y_{B.3|s} &:= X_{B.3} \cap (\xi^{\max}_{B.3}, +\infty) \cap (\xi^{\max}_{B.3,s}, +\infty)\\
Y_{B.3|1} &:= X_{B.3} \cap (\xi^{\max}_{B.3}, +\infty) \cap (\xi^{\max}_{B.3,1}, +\infty)\\
Y_{B|s} &:= \cup_{i=2,3}\left(X_{B.i} \cap (\xi^{\max}_{B.i}, +\infty) \cap (\xi^{\max}_{B.i,s}, +\infty) \cap Y_{\neg B.i}\right)\\
Y_{B|1} &:= \cup_{i=2,3}\left(X_{B.i} \cap (\xi^{\max}_{B.i}, +\infty) \cap (\xi^{\max}_{B.i,1}, +\infty) \cap Y_{\neg B.i}\right)
\end{aligned}$$

with the previously introduced $X_{B.\cdot}$ and the following additional collection of set:

$$\begin{aligned}
X_{\neg B.1|s} &:= [0, +\infty) \\
X_{\neg B.2|s} &:= [0, \xi^{\max}_{B.2}] \cup \left((\xi^{\max}_{B.2}, +\infty) \cap [0, \xi^{\max}_{B.2,s}]\right) \\
X_{\neg B.3|s} &:= [0, \xi^{\max}_{B.3}] \cup \left((\xi^{\max}_{B.3}, +\infty) \cap [0, \xi^{\max}_{B.3,s}]\right) \\
X_{\neg B.1|1} &:= [0, +\infty) \\
X_{\neg B.2|1} &:= [0, \xi^{\max}_{B.2}] \cup \left((\xi^{\max}_{B.2}, +\infty) \cap [0, \xi^{\max}_{B.2,1}]\right) \\
X_{\neg B.3|1} &:= [0, \xi^{\max}_{B.3}] \cup \left((\xi^{\max}_{B.3}, +\infty) \cap [0, \xi^{\max}_{B.3,1}]\right)
\end{aligned}$$

Finally, incorporating the participation constraint of the entrepreneur, the candidate for the optimal choice for the project share depending on the initial effort is given by:

$$\lambda^* = \begin{cases} \lambda_s & e_0 \in Z_s \\ 1 & e_0 \in Z_1 \\ \lambda_{B.2} & e_0 \in Z_{B.2} \\ \lambda_{B.3} & e_0 \in Z_{B.3} \\ \emptyset & e_0 \in Z_\emptyset \end{cases}$$

C.4. **Step (1).** As demonstrated in the previous step, the five classes of equilibria require refinement, specifically the one concerning the venture capitalist actually contains four sub-classes. In addition to the project shares originating from the objective function of the venture capitalist, with either partial or complete transfer of the project payoff, the venture capitalist potentially acts as a pseudo-bank, essentially mimicking the scheme of either partial or complete re-payment. Now, ultimately the optimal choice of the initial effort depends on a comparison of the respective objective functions with the corresponding initial effort plugged in. Thus, let us now compute the initial effort for the different objective functions.

First, the objective function of the entrepreneur for the three bank scenarios as an interval-wise function of the initial effort is given by the following expression(s):

$$\Pi^{\tilde{q}}_E = -\frac{e_0^2}{2} + \begin{cases} 0 & e_0 \in X_{B.1} \\ \pi_E(G|\tilde{q})\left(\alpha(G)\beta_0 e_0 - rI\right) + \frac{(\pi_E(G|\tilde{q})\alpha(G)\beta_1)^2}{2} & e_0 \in X_{B.2} \\ \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0 e_0 - rI + \frac{\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1\right)^2}{2} & e_0 \in X_{B.3} \end{cases}$$

As in the step of the additional effort, without further restrictions it is possible that the optimal choice of an interval does not belong to that specific interval. Similarly, it is also not obvious yet which of the intervals contains the overall optimal choice. In the following, let us first therefore determine the candidates for optimal choices on each interval with the specific requirements for them to exist within their interval:

$$\begin{aligned}
e_0^{c_1} &= 0 && \text{for } 0 \in X_{B.1} \\
e_0^{c_2} &= \pi_E(G|\tilde{q})\alpha(G)\beta_0 && \text{for } \pi_E(G|\tilde{q})\alpha(G)\beta_0 \in X_{B.2} \\
e_0^{c_3} &= \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0 && \text{for } \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0 \in X_{B.3}
\end{aligned}$$

Note, similar to the computation of the additional effort, either the optimal choice lies within the corresponding interval or the final comparison between the different bank scenarios eliminates this option. The profit functions with respect to these bank scenarios all take the form of quadratic functions of the initial effort while the corresponding intervals are such that each pair of functions is identical on the border of those intervals. Thus, if the optimal choice on an interval is on the border, then it is not possible that that of another interval also falls onto the border using arguments similar to before. Therefore, any optimal choice on the border is always dominated by a non-border one of the other interval. As a consequence, refer to the candidates as B.1, B.2, and B.3 from now on (essentially extending the notation). Now, alternatively, in terms of the expected profit, the three candidates are given by

$$\begin{aligned}
&e_0^{c_1} = 0 && \text{for } \Delta(B.1, B.1; B.1, B.i) \geq 0 \text{ and } i = 2, 3 \\
&e_0^{c_2} = \pi_E(G|\tilde{q})\alpha(G)\beta_0 && \text{for } \begin{matrix} \Delta(B.2, B.2; B.2, B.1) > 0 \\ \Delta(B.2, B.2; B.2, B.3) \geq 0 \end{matrix} \\
&e_0^{c_3} = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0 && \text{for } \Delta(B.3, B.3; B.3, B.i) > 0 \text{ and } i = 1, 2
\end{aligned}$$

with $\Delta(X, Y; X', Y') := \Pi_E^{\tilde{q}}(e_0 = X, e_1 = X') - \Pi_E^{\tilde{q}}(e_0 = Y, e_1 = Y')$ and by using the definition of the sets $X_{B..}$.[41] Again, note that so far this only characterizes potential optimal choices and the conditions for their existence. In order to find the unique optimal choice, let us a look at the deltas of the expected profits corresponding to the three candidates while re-using (and abusing) the notation $\Delta(X, Y) := \Delta(X, Y; X, Y)$

- $\Delta(B.2, B.1) = \frac{\pi_E(G|\tilde{q})^2\alpha(G)^2\left(\beta_0^2+\beta_1^2\right)}{2} - \pi_E(G|\tilde{q})rI$
- $\Delta(B.3, B.1) = \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right)}{2} - rI$
- $\Delta(B.3, B.2) = \frac{\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2 - \pi_E(G|\tilde{q})^2\alpha(G)^2\right)\left(\beta_0^2+\beta_1^2\right)}{2} - \pi_E(B|\tilde{q})rI$

and furthermore note that whenever one of the bank scenarios dominates the others, the (two) conditions for its existence are always satisfied due to following relation:

$$\Delta(B.j, B.j; B.j, B.i) \geq \Delta(B.j, B.i) \text{ for } i, j \in \{1, 2, 3\}$$

Thus, the optimal choice for initial effort depending on the specific scenario is given by the following

$$e_0^* = \begin{cases} 0 & \Delta(B.1, B.2) \geq 0 \text{ and } \Delta(B.1, B.3) \geq 0 \\ \pi_E(G|\tilde{q})\alpha(G)\beta_0 & \Delta(B.2, B.1) > 0 \text{ and } \Delta(B.2, B.3) \geq 0 \\ \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0 & \Delta(B.3, B.1) > 0 \text{ and } \Delta(B.3, B.2) > 0 \end{cases}$$

which in turn implies the following overall expected profit for these three alternatives:

$$\Pi_E^{\tilde{q}} = \begin{cases} 0 & \Delta(B.1, B.2) \geq 0 \text{ and } \Delta(B.1, B.3) \geq 0 \\ \frac{\pi_E(G|\tilde{q})^2\alpha(G)^2\left(\beta_0^2+\beta_1^2\right)}{2} - \pi_E(G|\tilde{q})rI & \Delta(B.2, B.1) > 0 \text{ and } \Delta(B.2, B.3) \geq 0 \\ \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right)}{2} - rI & \Delta(B.3, B.1) > 0 \text{ and } \Delta(B.3, B.2) > 0 \end{cases}$$

Furthermore, it is obvious that for all of the three constellations $\Pi_{VC}^{\tilde{q}} = 0$.

[41] Recall our rules on tie-breaking, which influence the specific nature of these inequalities.

Consider the objective function of the entrepreneur as an interval-wise function of the initial effort again but this time for the three alternative non-bank scenarios:

$$\Pi_E^{\tilde{q}} = -\frac{e_0^2}{2} + \begin{cases} (1-\lambda_s)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0 e_0 + \beta_2 \bar{e}\right) + \frac{\left((1-\lambda_s)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1\right)^2}{2} & e_0 \in Z_s^c \\ 0 & e_0 \in Z_1^c \\ 0 & e_0 \in Z_\emptyset^c \end{cases}$$

Note that due to our two-step approach (of treating bank and non-bank scenarios separately in the first step and then comparing them in the second one) these sets appear in their pre-'participation constraint' form. Also, due to the different nature of these three scenarios, let us work through them one by one instead of following the all-in-one method of the bank scenarios. Furthermore, let us then work our way from the bottom to the top. In the last scenario, i.e. the case of no funding, the objective function implies the following single candidate and its existence conditions:

$$e_0^c = 0 \qquad \text{for } 0 \in Z_\emptyset^c$$

However, without even investigating the corresponding conditions, it is obvious that as a consequence $\Pi_E^{\tilde{q}} = 0$. In other words, from the perspective of the entrepreneur this scenario is identical to the first bank scenario. While it is actually not yet time in our two-step approach for the comparison with the bank scenario(s), this insight eventually eliminates the scenario of no funding from the pool of potential equilibria due to our convention on tie-breaking and the nature of the bank as a (non-)player. Let us therefore eliminate this scenario at this point already. Note that this extends naturally from $Z_\emptyset^c$ to $Z_\emptyset$ when introducing the participation constraint later on. Also, this whole scenario initially contains the possibility of a non-zero initial effort, but any non-zero initial effort necessarily implies strictly negative expected profit, which eliminates these options from the pool as well. Now, in the middle scenario, i.e. the case of complete transfer, the objective function implies a similar situation:

$$e_0^c = 0 \qquad \text{for } 0 \in Z_1^c$$

As before, it is obvious that $\Pi_E^{\tilde{q}} = 0$. Again, from the perspective of the entrepreneur this scenario is identical to the first bank scenario, but here it differs from the point of view of the venture capitalist. However, while the equivalence in expected profit for the entrepreneur eliminated the other scenario from the pool of potential equilibria, the same does not apply to the case of a complete transfer because of our convention on tie-breaking. Also, similar to before, this case initially contains the possibility of a non-zero initial effort, which then gets eliminated from the pool as well. Now, alternatively, in terms of parameter relations the single candidate is given by

$$e_0^c = 0 \qquad \text{for } \xi_{s,1}^{\max} \leq 0 \text{ and } \xi_{VC}^{\min_2} \leq 0$$

and refer to this candidate as VC.B.1 (analogous to the equivalent B.· notation). Finally, in the first scenario, i.e. the case of partial transfer, the different values for the interval borders actually matter as potential optimal choices for once. Therefore, let us provide the inner case first and expand the analysis to the borders afterwards. By plugging the corresponding project share into the objective function and also

using $\partial^2\Pi_E^{\tilde{q}}/\partial^2 e_0 < 0$, the inner case candidate for the initial effort is given by:

$$e_k := \beta_0 \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - 3\beta_2\bar{e}}{3\beta_0^2 + 4\beta_1^2}$$

Now, what is left to do is to check whether this candidate actually belongs to the corresponding interval. Recall, the definition of $Z_s^c = [0, \xi_{s,1}^{\max}) \cap [\xi_{VC}^{\min_1}, +\infty)$ and note that as soon as the upper limit of this interval becomes relevant, the profit for any alternative still within the interval approaches the profit of the border value from below (without reaching it), which itself equals that of the complete transfer, or no funding. As a consequence, this constellation gets eliminated from the pool. Now, what remains are these three candidates and their existence conditions:

$$\begin{aligned} e_0^{c_1} &= 0 && \text{for } 0 \in Z_s^c \text{ and } e_k < 0 \\ e_0^{c_2} &= \xi_{VC}^{\min_1} && \text{for } \xi_{VC}^{\min_1} \in Z_s^c \text{ and } e_k < \xi_{VC}^{\min_1} \\ e_0^{c_3} &= e_k && \text{for } e_k \in Z_s^c \end{aligned}$$

Alternatively, in terms of parameter relations the three candidates are given by

$$\begin{aligned} e_0^{c_1} &= 0 && \text{for } \xi_{VC}^{\min_1} \leq 0 < \xi_{s,1}^{\max} \text{ and } e_k < 0 \\ e_0^{c_2} &= \xi_{VC}^{\min_1} && \text{for } 0 < \xi_{VC}^{\min_1} < \xi_{s,1}^{\max} \text{ and } e_k < \xi_{VC}^{\min_1} \\ e_0^{c_3} &= e_k && \text{for } \xi_{VC}^{\min_1} \leq e_k < \xi_{s,1}^{\max} \text{ and } 0 \leq e_k \end{aligned}$$

and refer to these candidates as VC.0, VC.p, and VC.k from now on. Furthermore, let us introduce a change in notation of $e_p := \xi_{VC}^{\min_1}$, $\xi_{VC} := \xi_{VC}^{\min_2}$, and $\xi_\lambda := \xi_{s,1}^{\max}$ for the sake of clarity. Then, the optimal choice for the initial effort is given by:

$$e_0^* = \begin{cases} 0 & \xi_\lambda \leq 0 \text{ and } \xi_{VC} \leq 0 \\ & e_p \leq 0 < \xi_\lambda \text{ and } e_k < 0 \text{ and } \Pi_E^{\tilde{q}}(VC.0) > 0 \\ e_p & 0 < e_p < \xi_\lambda \text{ and } e_k < e_p \text{ and } \Pi_E^{\tilde{q}}(VC.p) > 0 \\ e_k & e_p \leq e_k < \xi_\lambda \text{ and } 0 \leq e_k \text{ and } \Pi_E^{\tilde{q}}(VC.k) > 0 \end{cases}$$

Note that due to the exclusive nature of the required parameter relations, the three single conditions on the expected profits are not only necessary but in fact sufficient. Let us now also provide for each initial effort the project share and additional effort, as well as the corresponding expected profit of entrepreneur and venture capitalist. For the scenario VC.B.1 with $e_0^* = 0$ it follows that $\lambda^* = 1$ and $e_1^* = 0$, which then implies expected profits of $\Pi_E^{\tilde{q}} = 0$ as well as $\Pi_{VC}^{\tilde{q}} = \mathbb{E}_P(\alpha|\tilde{q})\beta_2\bar{e} - \frac{\bar{e}^2}{2} - \bar{r}I$. Next, in the case of VC.0 with $e_0^* = 0$ the project share and additional effort are given by

$$\begin{aligned} \lambda^* &= \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + \beta_2\bar{e}}{2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2} \\ e_1^* &= \frac{1}{2\beta_1}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - \beta_2\bar{e}\right) \end{aligned}$$

which all together yield the following overall expected profit for the two players of

$$\begin{aligned} \Pi_E^{\tilde{q}} &= \frac{1}{8\beta_1^2}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + 2\beta_2\bar{e}\right) - 3\beta_2^2\bar{e}^2\right) \\ \Pi_{VC}^{\tilde{q}} &= \frac{\mathbb{E}_P(\alpha|\tilde{q})}{4\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + 2\beta_2\bar{e}\right) + \beta_2^2\bar{e}^2\right) - \frac{\bar{e}^2}{2} - \bar{r}I \end{aligned}$$

For VC.p with $e_0^* = \left(-\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - \beta_2\bar{e} + \beta_1\sqrt{2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})(\bar{e}^2+2\bar{r}I)/\mathbb{E}_P(\alpha|\tilde{q})}\right)/\beta_0$, the corresponding project share and additional effort are given by the two expressions

$$\lambda^* = \sqrt{\frac{\bar{e}^2+2\bar{r}I}{2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}}$$

$$e_1^* = \left(1-\sqrt{\frac{\bar{e}^2+2\bar{r}I}{2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}}\right)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1$$

which in turn then imply an overall expected profit for the respective parties of

$$\begin{aligned}
\Pi_E^{\tilde{q}} =& \frac{1}{4\mathbb{E}_P(\alpha|\tilde{q})\beta_0^2}\left(\Pi_{E,p_+}^{\tilde{q}} - \Pi_{E,p_-}^{\tilde{q}}\right)\\
\Pi_{E,p_+}^{\tilde{q}} :=& \sqrt{2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\bar{e}^2+2\bar{r}I\right)}\left(4\beta_1\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+\beta_2\bar{e}\right)\right)\\
\Pi_{E,p_-}^{\tilde{q}} :=& 2\mathbb{E}_P(\alpha|\tilde{q})\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+2\beta_2\bar{e}\right)+\beta_2^2\bar{e}^2\right)\\
&+\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(3\beta_0^2+4\beta_1^2\right)\left(\bar{e}^2+2\bar{r}I\right)\\
\Pi_{VC}^{\tilde{q}} =& 0
\end{aligned}$$

Finally, consider the scenario VC.k with $e_0^* = \beta_0\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - 3\beta_2\bar{e}\right)/\left(3\beta_0^2+4\beta_1^2\right)$, which corresponds to an optimal share and additional effort given by the terms

$$\lambda^* = \frac{2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+\beta_2\bar{e}\right)}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(3\beta_0^2+4\beta_1^2\right)}$$

$$e_1^* = \beta_1\frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+2\beta_1^2\right)-2\beta_2\bar{e}}{3\beta_0^2+4\beta_1^2}$$

resulting in an overall expected profit in this specific case for the two parties of

$$\begin{aligned}
\Pi_E^{\tilde{q}} =& \frac{1}{2\left(3\beta_0^2+4\beta_1^2\right)}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+2\beta_2\bar{e}\right)-3\beta_2^2\bar{e}^2\right)\\
\Pi_{VC}^{\tilde{q}} =& \frac{4\mathbb{E}_P(\alpha|\tilde{q})\beta_1^2}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(3\beta_0^2+4\beta_1^2\right)^2}\Pi_{VC,k}^{\tilde{q}} - \frac{\bar{e}^2}{2} - \bar{r}I\\
\Pi_{VC,k}^{\tilde{q}} :=& \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+2\beta_2\bar{e}\right)+\beta_2^2\bar{e}^2
\end{aligned}$$

As a consequence of our two-step approach, two scenarios escaped the analysis so far, namely the scenarios where the venture capitalist acts as a pseudo-bank.[42] Now, the relevance of the pseudo-bank scenarios depends on $\Pi_E^{\tilde{q}}(B.*) > \Pi_E^{\tilde{q}}(VC.*)$, where $B.*$ resp. $VC.*$ refer to the optimal bank resp. non-bank scenario. Then, the corresponding pseudo-bank scenarios appear whenever it holds that $\lambda_{VC} \leq \lambda_{B.*}$.[43] Under these conditions and the relevance criterion, the payoff equivalence with the corresponding bank scenario for the entrepreneur implies an identical initial effort:

$$e_0^* = \begin{cases} \pi_E(G|\tilde{q})\alpha(G)\beta_0 & B.* = B.2 \\ \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0 & B.* = B.3 \end{cases}$$

Note that the payoff necessarily differs from the perspective of the venture capitalist. Also, recall the insight that whenever one of the bank scenarios dominates the others,

[42] Note that the scenario with complete transfer acts as another pseudo-bank scenario.

[43] Implicitly, $\Pi_E^{\tilde{q}}(B.*) > \Pi_E^{\tilde{q}}(VC.*)$ requires $\Pi_E^{\tilde{q}}(B.*) > 0$, $\Delta_{B.*} > 0$ and $\Delta_{VC.*} > 0$, where $\Delta_{VC.*}$ refers to the already defined Δ_{VC} corresponding to $VC.*$.

the conditions for its existence are always satisfied. As before, let us now provide for each of the two initial efforts the project share and additional effort, as well as the corresponding expected profit of entrepreneur and venture capitalist. Specifically, for the case with $e_0^* = \pi_E(G|\tilde{q})\alpha(G)\beta_0$ it follows that the project share is given by

$$\begin{aligned}\lambda^* =& \frac{1}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}\left(\lambda_{B.2}^+ - \lambda_{B.2}^-\right)\\ \lambda_{B.2}^+ :=& \pi_E(G|\tilde{q})\alpha(G)\beta_0^2 + \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + \beta_2\bar{e}\\ \lambda_{B.2}^- :=& \sqrt{\pi_E(G|\tilde{q})^2\alpha(G)^2\left(\beta_0^2+\beta_1^2\right)^2 + \beta_2^2\bar{e}^2 + 2\pi_E(G|\tilde{q})\alpha(G)\beta_0^2\beta_2\bar{e} - 2\pi_E(G|\tilde{q})\beta_1^2 rI}\end{aligned}$$

implying the corresponding additional effort of $e_1^* = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 - (\lambda_{B.2}^+ - \lambda_{B.2}^-)/\beta_1$, which yields the expected profit of $\Pi_E^{\tilde{q}} = \pi_E(G|\tilde{q})^2\alpha(G)^2\left(\beta_0^2+\beta_1^2\right)/2 - \pi_E(G|\tilde{q})rI$ for the entrepreneur (matching the corresponding bank scenario) and furthermore $\Pi_{VC}^{\tilde{q}} = \lambda^*\mathbb{E}_P(\alpha|\tilde{q})\lambda_{B.2}^- - \frac{\bar{e}^2}{2} - \bar{r}I$ for the venture capitalist. Following the other examples in terms of notation, let us denote this scenario by VC.B.2 going forward. In the scenario with $e_0^* = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0$ it follows that the project share is given by

$$\begin{aligned}\lambda^* =& \frac{1}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}\left(\lambda_{B.3}^+ - \lambda_{B.3}^-\right)\\ \lambda_{B.3}^+ :=& \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right) + \beta_2\bar{e}\\ \lambda_{B.3}^- :=& \sqrt{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right)^2 + \beta_2^2\bar{e}^2 + 2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0^2\beta_2\bar{e} - 2\beta_1^2 rI}\end{aligned}$$

yielding the corresponding additional effort of $e_1^* = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 - (\lambda_{B.3}^+ - \lambda_{B.3}^-)/\beta_1$, which consequently implies the expected profit of $\Pi_E^{\tilde{q}} = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right)/2 - rI$ for the entrepreneur (matching again the corresponding bank scenario) and also $\Pi_{VC}^{\tilde{q}} = \lambda^*\mathbb{E}_P(\alpha|\tilde{q})\lambda_{B.3}^- - \frac{\bar{e}^2}{2} - \bar{r}I$ for the venture capitalist. Finally, let us denote this scenario by VC.B.3. Furthermore, it holds that $\lambda_{VC} = \iota_{B.\cdot} - \sqrt{\iota_{B.\cdot}^2 - \psi}$ using $\psi = \left(\bar{e}^2 + 2\bar{r}I\right)/\left(2\mathbb{E}_P(\alpha|\tilde{q})\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\right)$ and $\iota_{B.\cdot} := \lambda_{B.\cdot}^+/\left(2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\right)$.[44]

In order to formulate the overall optimum, use $B.*$ for the optimal bank scenario, $VC.*$ for the optimal not-bank-equivalent venture capitalist scenario, and also $VC.B.*$ for the optimal bank-equivalent venture capitalist scenario.[45] Finally, with the corresponding conditions already determined, it is an intuitive two-step process. First, if $\Delta(VC.*, B.*) \geq 0$, then $VC.*$ is simply the overall optimum. Otherwise, provided its conditions apply, $VC.B.*$ is the overall optimum, else, it is $B.*$.

In the interest of providing a concluding overview of all the scenarios, let us use:

$$\begin{aligned}\lambda_{B.1} :&= 1\\ \lambda_0 :&= \lambda_s(0)\\ \lambda_p :&= \lambda_s(e_p)\\ \lambda_k :&= \lambda_s(e_k)\end{aligned}$$

By combining everything, the following complete list of potential scenarios emerges:

[44] While it is not obvious through the notation, the critical value for the project share depends on the specific bank scenario under consideration.

[45] Contrary to the previous notation the non-bank scenarios get split into two notations here.

	e_0	λ	s	e_1	e_2
B.1	0	$\emptyset$	B	0	0
B.2	$\pi_E(G\|\tilde{q})\alpha(G)\beta_0$	$\emptyset$	B	$\pi_E(G\|\tilde{q})\alpha(G)\beta_1$	0
B.3	$\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_0$	$\emptyset$	B	$\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	0
VC.0	0	λ_0	VC	$(1-\lambda_0)\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$
VC.p	e_p	λ_p	VC	$(1-\lambda_p)\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$
VC.k	e_k	λ_k	VC	$(1-\lambda_k)\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$
VC.B.1	0	$\lambda_{B.1}$	VC	0	$\bar{e}$
VC.B.2	$\pi_E(G\|\tilde{q})\alpha(G)\beta_0$	$\lambda_{B.2}$	VC	$(1-\lambda_{B.2})\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$
VC.B.3	$\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_0$	$\lambda_{B.3}$	VC	$(1-\lambda_{B.3})\mathbb{E}_{\pi_E}(\alpha\|\tilde{q})\beta_1$	$\bar{e}$

It is important to point out that this whole analysis ultimately depended on the perceived signal by the entrepreneur. The complete equilibrium therefore contains two elements from the complete list of potential scenarios - one for each of the two perceived signals (which play the role of types in terms of the signaling game)

$$\sigma^* = \left(\sigma^*(\tilde{g}), \sigma^*(\tilde{b})\right)$$

where σ^* denotes the complete equilibrium strategy and $\sigma^*(\tilde{q})$ the individual ones. Now, the nature of the separating equilibrium eliminates a number of combinations, specifically the ones with identical initial effort in both cases, but would require further parameter specification to significantly limit the set of potential tuple.

Lastly, while the focus on separating equilibria immediately fixes the belief system of the venture capitalist in equilibrium, it actually requires additional attention when not in equilibrium. Let us rectify this omission by specifying the belief system of the venture capitalist as a combination of indicator functions (with a singularity)

$$\pi(\tilde{q}|e_0) = \begin{cases} \mathbb{1}_{\{\tilde{q}=\tilde{d}\}}(\tilde{q}) & e_0 = e_0(\tilde{d}) \\ \mathbb{1}_{\{\tilde{q}\neq\tilde{d}\}}(\tilde{q}) & e_0 \neq e_0(\tilde{d}) \end{cases}$$

where $\tilde{d}$ refers to the specific signal which induces the more 'desirable' project share. Together with the non-pooling condition for the separating equilibrium given by $\Pi_E^{\tilde{q}}(e_0^*(\tilde{d})) < \Pi_E^{\tilde{q}}(e_0^*(\tilde{q}))$ for $\tilde{q} \neq \tilde{d}$ (abstracting from the remaining strategy tuple), this definition of the belief system prevents any deviation from the optimal choices, as the expected profit of the entrepreneur decreases with increasing project share.

Appendix D. Model Dynamics Calculations

The following appendix contains all the calculations for the model dynamics presented in the main part of the paper. Almost all of the arguments are based on our assumption of non-zero parameters.

Proof of Proposition 1. Recall, $e_0(B.3) = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0$ and $e_1(B.3) = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1$ as well as $\Pi_E(B.3) = \frac{1}{2}\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\left(\beta_0^2+\beta_1^2\right) - rI$. By simply taking partial derivatives it follows that for $i \in \{0,1\}$

$$\text{(Ad 1.A)} \qquad \frac{\partial e_i(B.3)}{\partial \beta_i} = \mathbb{E}_{\pi_E}(\alpha|\tilde{q}) > 0$$

$$\text{(Ad 1.B)} \qquad \frac{\partial e_i(B.3)}{\partial \mathbb{E}_{\pi_E}(\alpha|\tilde{q})} = \beta_i > 0$$

$$\text{(Ad 2.A)} \qquad \frac{\partial \Pi_E(B.3)}{\partial \beta_i} = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2\beta_i > 0$$

and

$$\text{(Ad 2.B)} \qquad \frac{\partial \Pi_E(B.3)}{\partial \mathbb{E}_{\pi_E}(\alpha|\tilde{q})} = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right) > 0$$

$$\text{(Ad 2.C)} \qquad \frac{\partial \Pi_E(B.3)}{\partial r} = -I < 0$$

$$\text{(Ad 2.D)} \qquad \frac{\partial \Pi_E(B.3)}{\partial I} = -r < 0$$

□

Proof of Proposition 2. Recall the terms for the entrepreneurial effort and profit

$$\begin{aligned}
e_0(VC.k) &= \beta_0 \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - 3\beta_2\bar{e}}{3\beta_0^2+4\beta_1^2} \\
e_1(VC.k) &= \beta_1 \frac{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+2\beta_1^2\right) - 2\beta_2\bar{e}}{3\beta_0^2+4\beta_1^2} \\
&= \left(1-\lambda(VC.k)\right)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 \\
\Pi_E(VC.k) &= \frac{1}{2\left(3\beta_0^2+4\beta_1^2\right)}\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2\left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\left(\beta_0^2+\beta_1^2\right)+2\beta_2\bar{e}\right) - 3\beta_2^2\bar{e}^2\right)
\end{aligned}$$

where the alternative representation of the additional effort originates from the calculations in Appendix C. By taking partial derivatives it then follows that:

$$\text{(Ad 1.A)} \quad \frac{\partial e_0(VC.k)}{\partial \beta_2} = -\frac{3\beta_0\bar{e}}{3\beta_0^2+4\beta_1^2} < 0 \quad \frac{\partial e_1(VC.k)}{\partial \beta_2} = -\frac{2\beta_1\bar{e}}{3\beta_0^2+4\beta_1^2} < 0$$

$$\text{(Ad 1.B)} \quad \frac{\partial e_0(VC.k)}{\partial \bar{e}} = -\frac{3\beta_0\beta_2}{3\beta_0^2+4\beta_1^2} < 0 \quad \frac{\partial e_1(VC.k)}{\partial \bar{e}} = -\frac{2\beta_1\beta_2}{3\beta_0^2+4\beta_1^2} < 0$$

$$\text{(Ad 1.C)} \quad \frac{\partial e_0(VC.k)}{\partial \mathbb{E}_{\pi_E}(\alpha|\tilde{q})} = \frac{\beta_0\beta_1^2}{3\beta_0^2+4\beta_1^2} > 0 \quad \frac{\partial e_1(VC.k)}{\partial \mathbb{E}_{\pi_E}(\alpha|\tilde{q})} = \frac{\beta_1\left(\beta_0^2+2\beta_1^2\right)}{3\beta_0^2+4\beta_1^2} > 0$$

Furthermore, recall $e_0(B.3) = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0$ and $e_1(B.3) = \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1$ as well as anticipate $\frac{\partial\lambda(VC.k)}{\partial\beta_1} < 0$ and $\lambda(VC.k) > 0$ (two insights of Proposition 3). Then, by re-formulating the relations resp. taking partial derivatives again it follows that:

(Ad 1.D) $e_0(VC.k) < e_0(B.3) \Leftrightarrow 3\beta_0 \left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) \left(\beta_0^2 + \beta_1^2\right) + \beta_2 \bar{e}\right) > 0$
$e_1(VC.k) < e_1(B.3) \Leftrightarrow \lambda(VC.k)\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 > 0$

(Ad 2.A) $$\frac{\partial e_1(VC.k)}{\partial \beta_1} = (1 - \lambda(VC.k))\, \mathbb{E}_{\pi_E}(\alpha|\tilde{q}) - \frac{\partial \lambda(VC.k)}{\partial \beta_1} \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 > 0$$

(Ad 2.B) $$\frac{\partial e_1(VC.k)}{\partial \lambda(VC.k)} = -\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1 < 0$$

(Ad 3.A) $$\frac{\partial \Pi_E(VC.k)}{\partial \mathbb{E}_{\pi_E}(\alpha|\tilde{q})} = \frac{\beta_1^2 \left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) \left(\beta_0^2 + \beta_1^2\right) + \beta_2 \bar{e}\right)}{3\beta_0^2 + 4\beta_1^2} > 0$$

(Ad 4.A) $$\frac{\partial \Pi_{VC}(VC.k)}{\partial \bar{r}} = -I < 0$$

(Ad 4.B) $$\frac{\partial \Pi_{VC}(VC.k)}{\partial I} = -\bar{r} < 0$$

□

Proof of Proposition 3. Recall two alternative expressions for the project share

$$\lambda(VC.k) = \frac{2 \left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) \left(\beta_0^2 + \beta_1^2\right) + \beta_2 \bar{e}\right)}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) \left(3\beta_0^2 + 4\beta_1^2\right)} = \frac{\beta_0 e_0 + \mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 + \beta_2 \bar{e}}{2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2}$$

where the alternative representation originates from Appendix C again. Hence, by taking partial derivatives resp. using the alternative representation it follows that:

(Ad A) $$\frac{\partial \lambda(VC.k)}{\partial \beta_0} = \frac{4\beta_0 \left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2 - 3\beta_2 \bar{e}\right)}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) \left(3\beta_0^2 + 4\beta_1^2\right)^2} > 0 \text{ via } e_0(VC.k) > 0$$

(Ad B) $$\frac{\partial \lambda(VC.k)}{\partial \beta_1} = -\frac{4\beta_1 \left(\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_0^2 + 4\beta_2 \bar{e}\right)}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) \left(3\beta_0^2 + 4\beta_1^2\right)^2} < 0$$

(Ad C) $$\frac{\partial \lambda(VC.k)}{\partial \beta_2} = \frac{2\bar{e}}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) \left(3\beta_0^2 + 4\beta_1^2\right)} > 0$$

(Ad D) $$\frac{\partial \lambda(VC.k)}{\partial \bar{e}} = \frac{2\beta_2}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q}) \left(3\beta_0^2 + 4\beta_1^2\right)} > 0$$

(Ad E) $$\frac{\partial \lambda(VC.k)}{\partial \mathbb{E}_{\pi_E}(\alpha|\tilde{q})} = -\frac{2\beta_2 \bar{e}}{\mathbb{E}_{\pi_E}(\alpha|\tilde{q})^2 \left(3\beta_0^2 + 4\beta_1^2\right)} < 0$$

(Ad F) $$\frac{\partial \lambda(VC.k)}{\partial e_0} = \frac{\beta_0}{2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2} > 0$$

(Ad G) $$\lambda(VC.k) = \frac{1}{2} + \frac{\beta_0 e_0 + \beta_2 \bar{e}}{2\mathbb{E}_{\pi_E}(\alpha|\tilde{q})\beta_1^2} > \frac{1}{2}$$

□

Proof of Proposition 4. First of all, it holds that $\frac{\partial P(G|\tilde{g})}{\partial \mu} = \frac{1}{P(\tilde{g})^2}\gamma(1-\gamma)(1-2\epsilon) > 0$ as $\epsilon < \frac{1}{2}$. Consequently, using the fact that $\frac{\partial P(B|\tilde{g})}{\partial \mu} = -\frac{\partial P(G|\tilde{g})}{\partial \mu}$, it follows that $\frac{\partial \mathbb{E}_P(\alpha|\tilde{g})}{\partial \mu} = \frac{\partial P(G|\tilde{g})}{\partial \mu} (\alpha(G) - \alpha(B)) > 0$ as $\alpha(G) > \alpha(B)$. With $P(G|\tilde{b})$, resp. $P(B|\tilde{b})$, independent of μ it is clear that $\frac{\partial P(G|\tilde{b})}{\partial \mu} = \frac{\partial P(B|\tilde{b})}{\partial \mu} = 0$ and thus $\frac{\partial \mathbb{E}_P(\alpha|\tilde{b})}{\partial \mu} = 0$. Finally, $\frac{\partial P(\tilde{g})}{\partial \mu} = -\frac{\partial P(\tilde{b})}{\partial \mu} = -(\gamma\epsilon + (1-\gamma)(1-\epsilon)) < 0$. The other statements follow from the remarks in Appendix B. □

Appendix E. Overconfidence Dynamics

E.1. **Pseudocode.** The following represents a minimalistic pseudocode for the procedure used in the analysis of the overconfidence dynamics:

Algorithm Pairwise Comparative Static Effect of Overconfidence

Input: Pair (i, j), Expected Profit Π, Equilibrium Condition E, Social Welfare W
Output: List of Pairs (p_i, p_j) with Beneficial resp. Detrimental Impact (L_{ij}^B, L_{ij}^D)

procedure PairCompStatEff$((i, j), \Pi, E, W)$
  for $(p_i, p_j) \in P_{ij}$ **do**
    $p = (p_i, p_j) + (\bar{p_1}, \ldots, \hat{p_i}, \ldots, \hat{p_j}, \ldots, \bar{p_n})$
    for $\mu \in (0, 1)$ **do**
      $(S_{p,\tilde{g}}(\mu), S_{p,\tilde{b}}(\mu)) = \arg\max\{\Pi(p, \mu)(S_{\tilde{g}}, S_{\tilde{b}}) | E(S_{\tilde{g}}, S_{\tilde{b}})\}$
      $W_p(\mu) = W(p, \mu)(S_{p,\tilde{g}}(\mu), S_{p,\tilde{b}}(\mu))$
    if $\exists \mu_1 < \mu_2 \in (0, 1)$: $W_p(\mu_1) > W_p(\mu_2)$ **then**
      $L_{ij}^B += (p_i, p_j)$
    else
      $L_{ij}^D += (p_i, p_j)$
  return (L_{ij}^B, L_{ij}^D)

E.2. **Graphics.** The following part of the appendix contains all the graphics for the pairwise comparative static effect analysis of overconfidence. Each pair of parameters appears twice in the process, once for each of the two elements. Also, the next table lists all of the parameters under consideration again, together with a reference to the figure where it is the focus of the analysis (and a description as a reminder).

Parameter	Figure	Description
β_0	5	initial effort productivity of the entrepreneur
β_1	6	additional effort productivity of the entrepreneur
β_2	7	additional effort productivity of the venture capitalist
$\alpha(B)$	8	overall project productivity for the bad project
$\alpha(G)$	9	overall project productivity for the good project
γ	10	probability of the project being of good quality
ϵ	11	probability of the signal being influenced by noise
$\bar{r}$	12	risk-free interest rate
$\bar{p}$	13	interest rate premium
$\theta_{\bar{e}}$	14	scalar for the service package of the venture capitalist
θ_I	15	scalar for the cost of the investment

Finally, for each of the parameters there exist a figure with ten (sub-)figures. While the 'main' parameter under consideration is on the y-axis for all ten of them, the other parameter on the x-axis varies. Each point in the (scatter) plot corresponds to exactly one parameter constellation (with the remaining parameters according to their default value) and a varying degree of overconfidence. In total, 99 different levels of overconfidence, with equal distance to each other, are considered each time. Furthermore, each parameter varies 199 times. In the graphics, a black dot indicates that there exists a beneficial impact of overconfidence purely driven by a change in one of the two equilibrium scenarios from pseudo-bank to bank proper, dark gray implies a beneficial influence on social welfare through other channels, light gray entails a detrimental effect, while the remaining white area corresponds to a lack of a (separating) equilibrium or parameter restrictions ($\alpha(B) < \alpha(G)$).

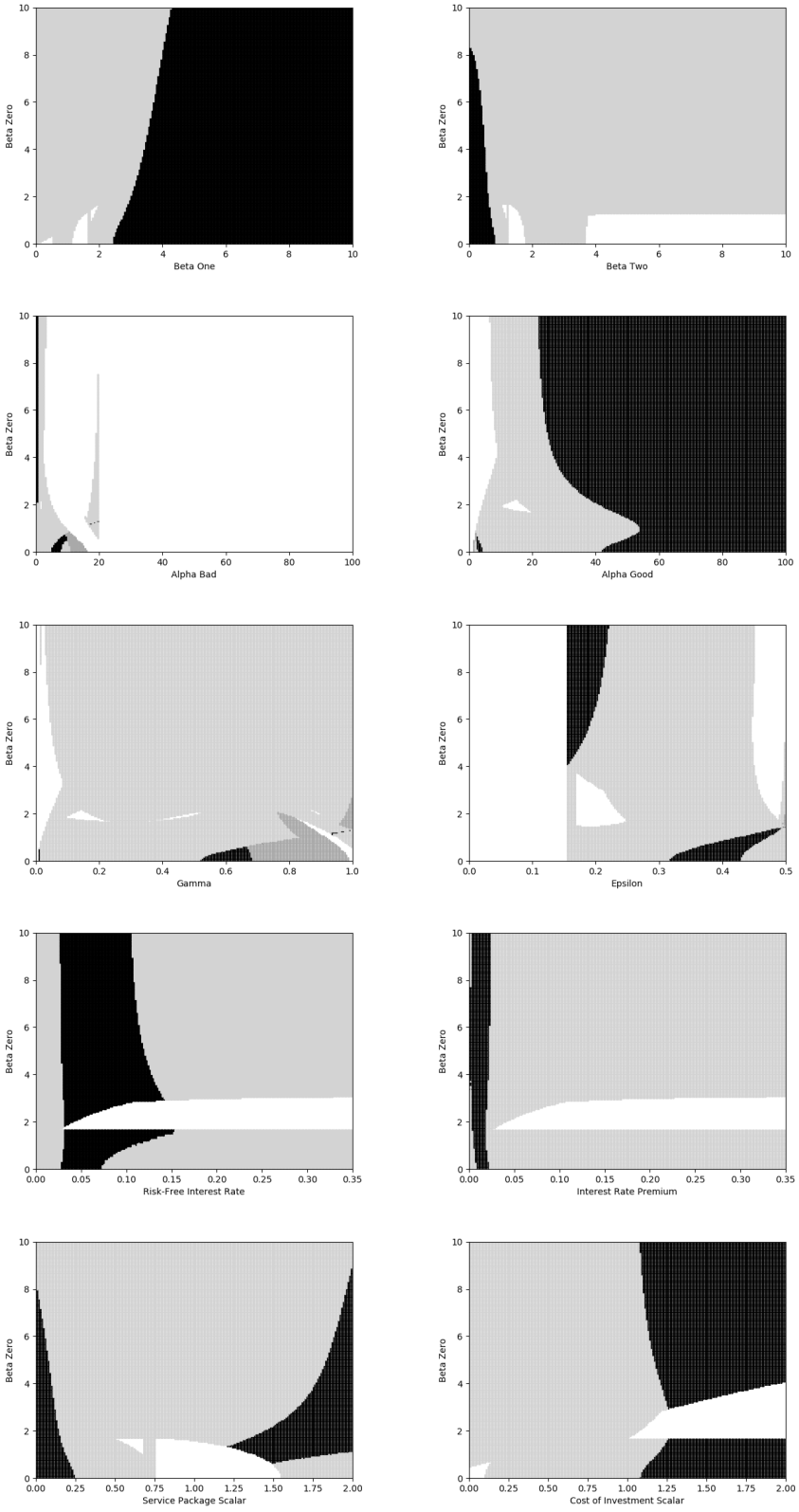


FIGURE 5. The overconfidence dynamics with a focus on β_0

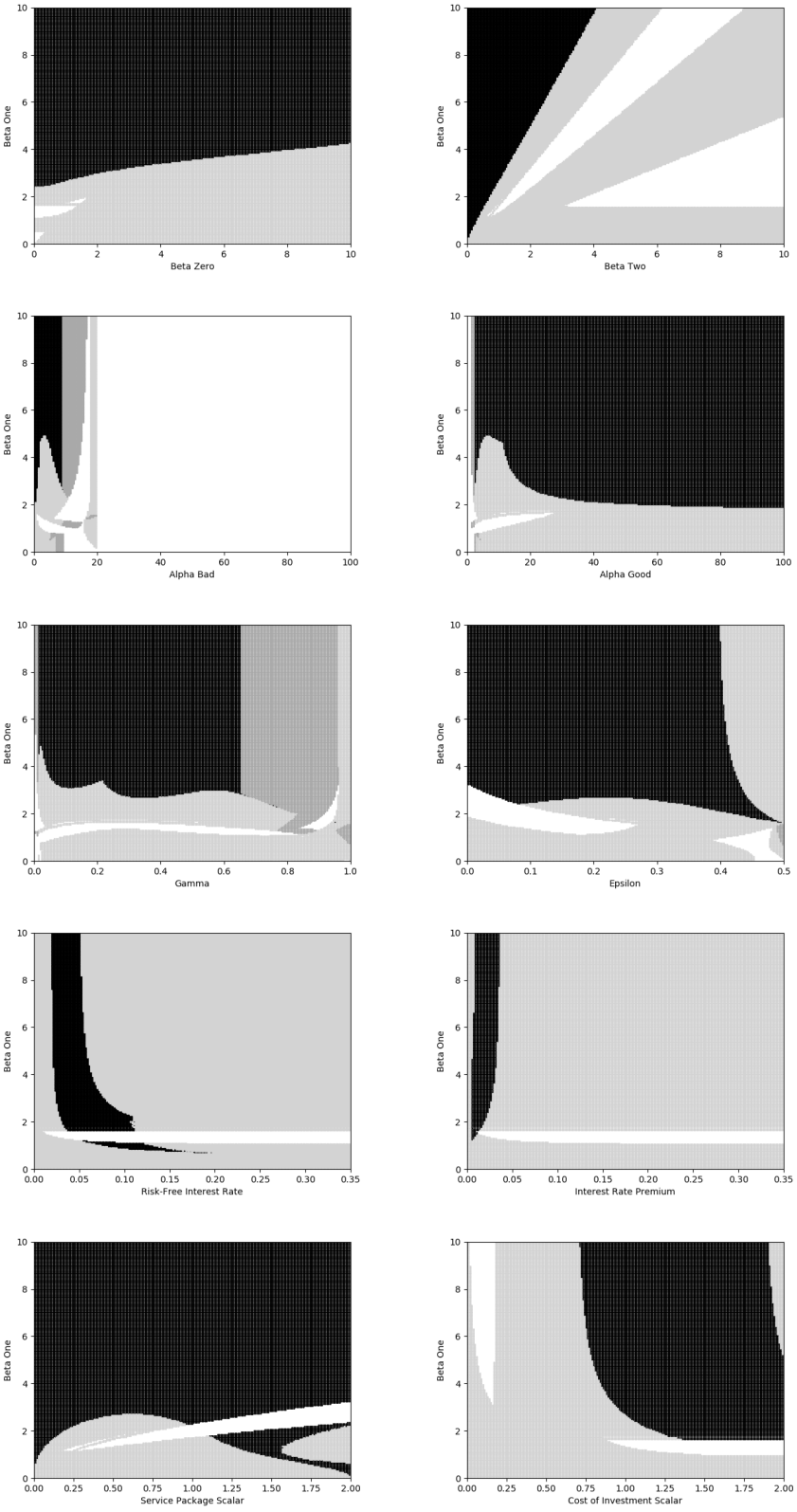


FIGURE 6. The overconfidence dynamics with a focus on β_1

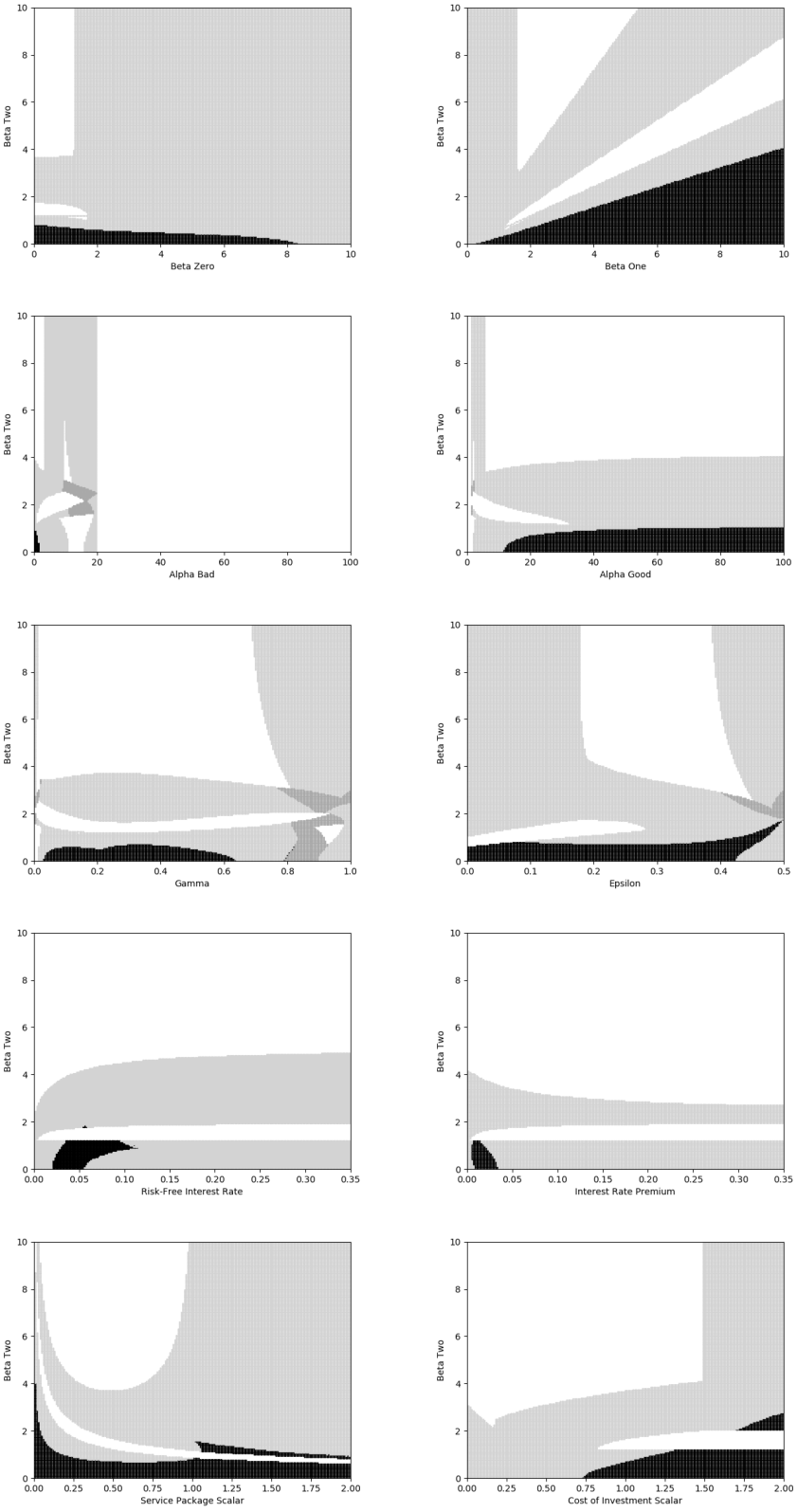


FIGURE 7. The overconfidence dynamics with a focus on β_2

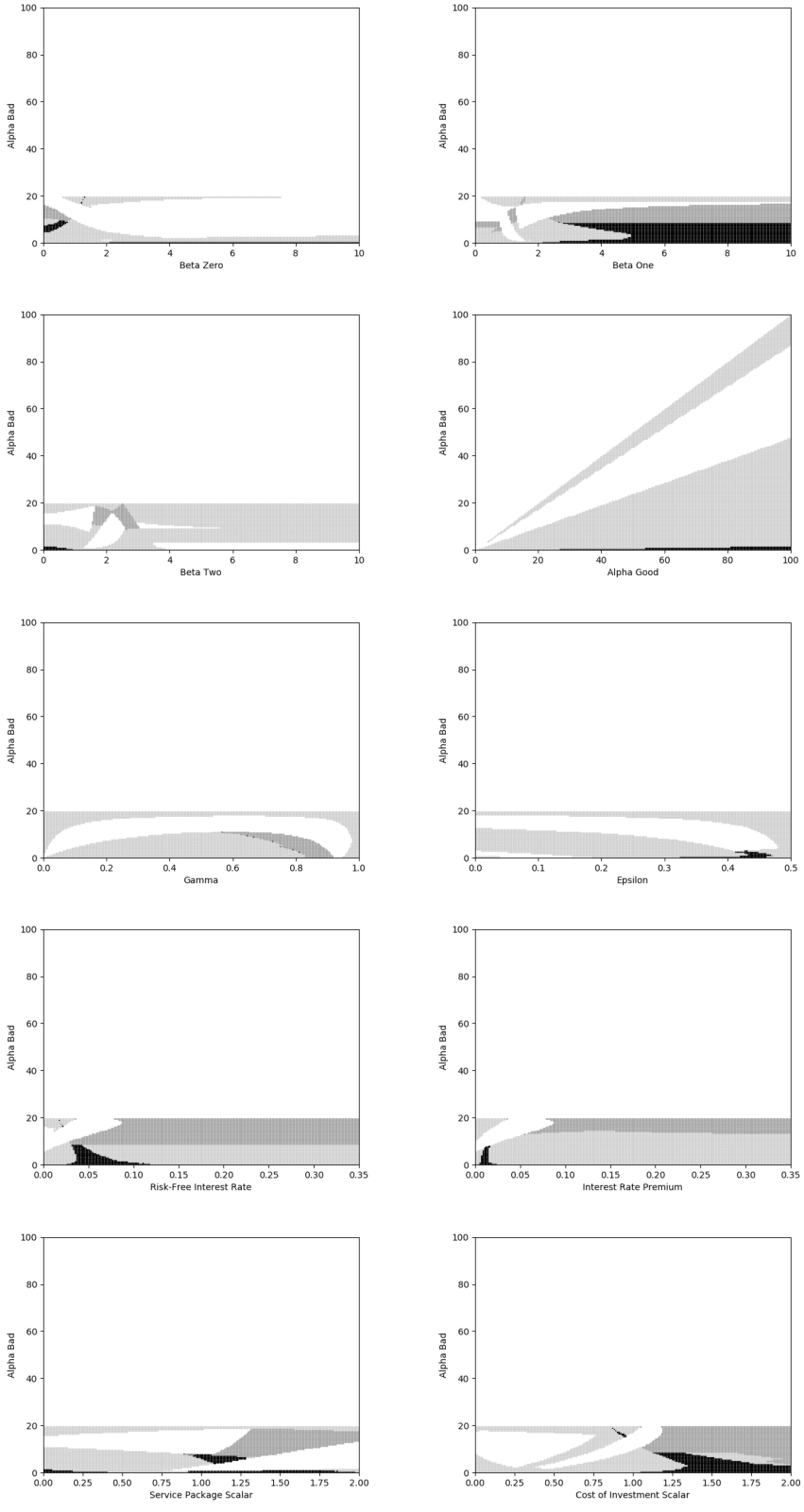


FIGURE 8. The overconfidence dynamics with a focus on $\alpha(B)$

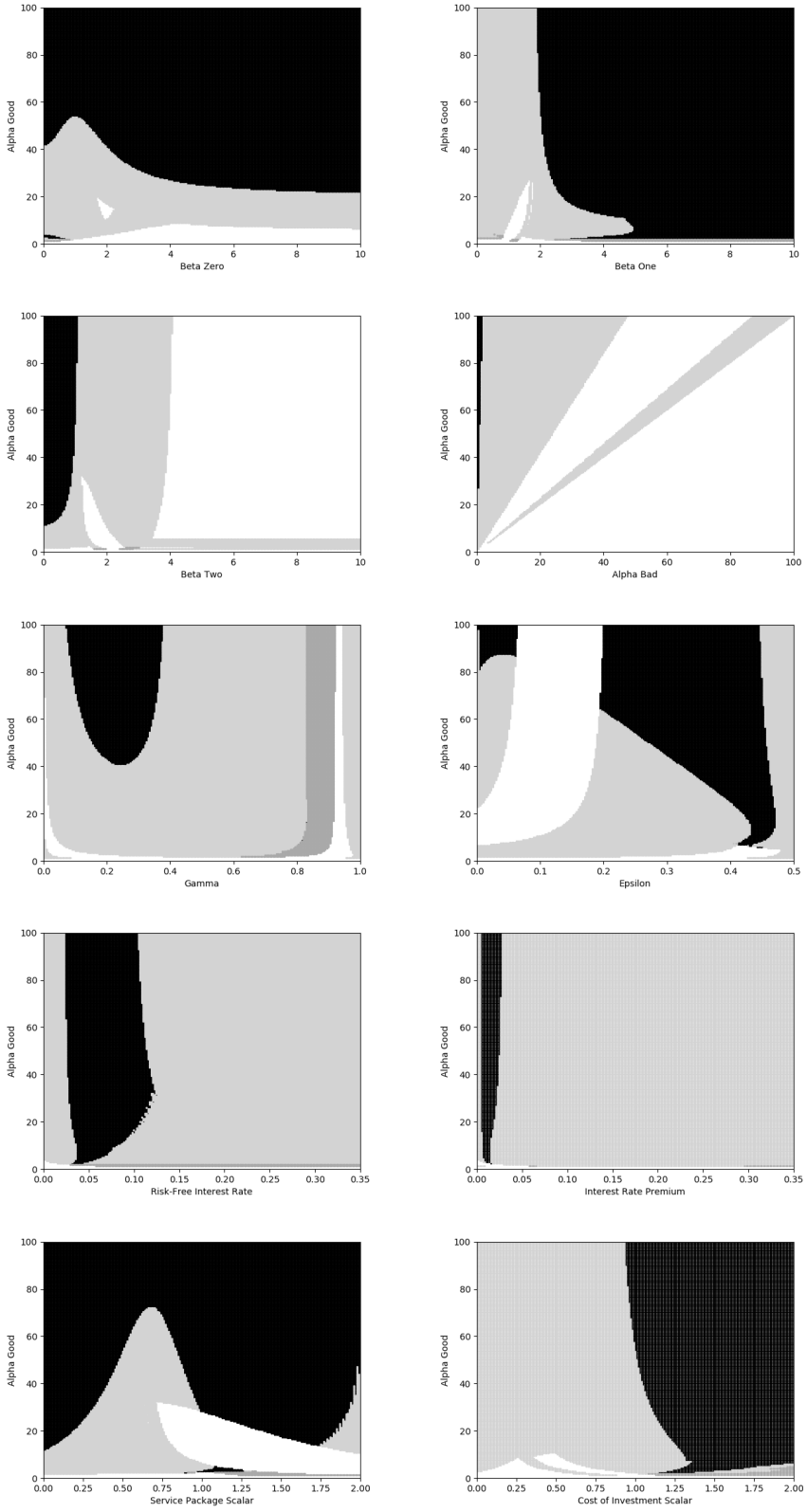


FIGURE 9. The overconfidence dynamics with a focus on $\alpha(G)$

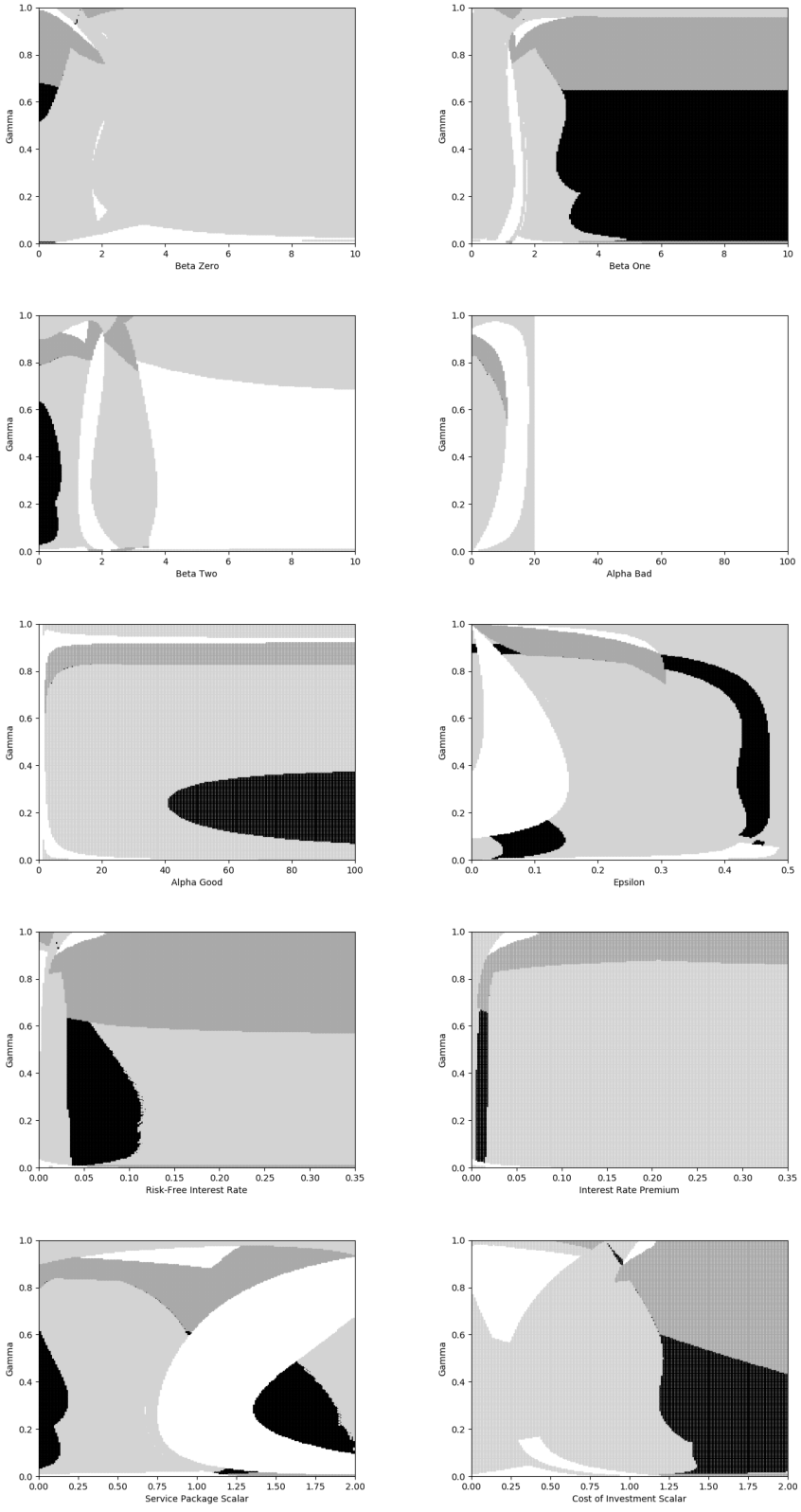


FIGURE 10. The overconfidence dynamics with a focus on γ

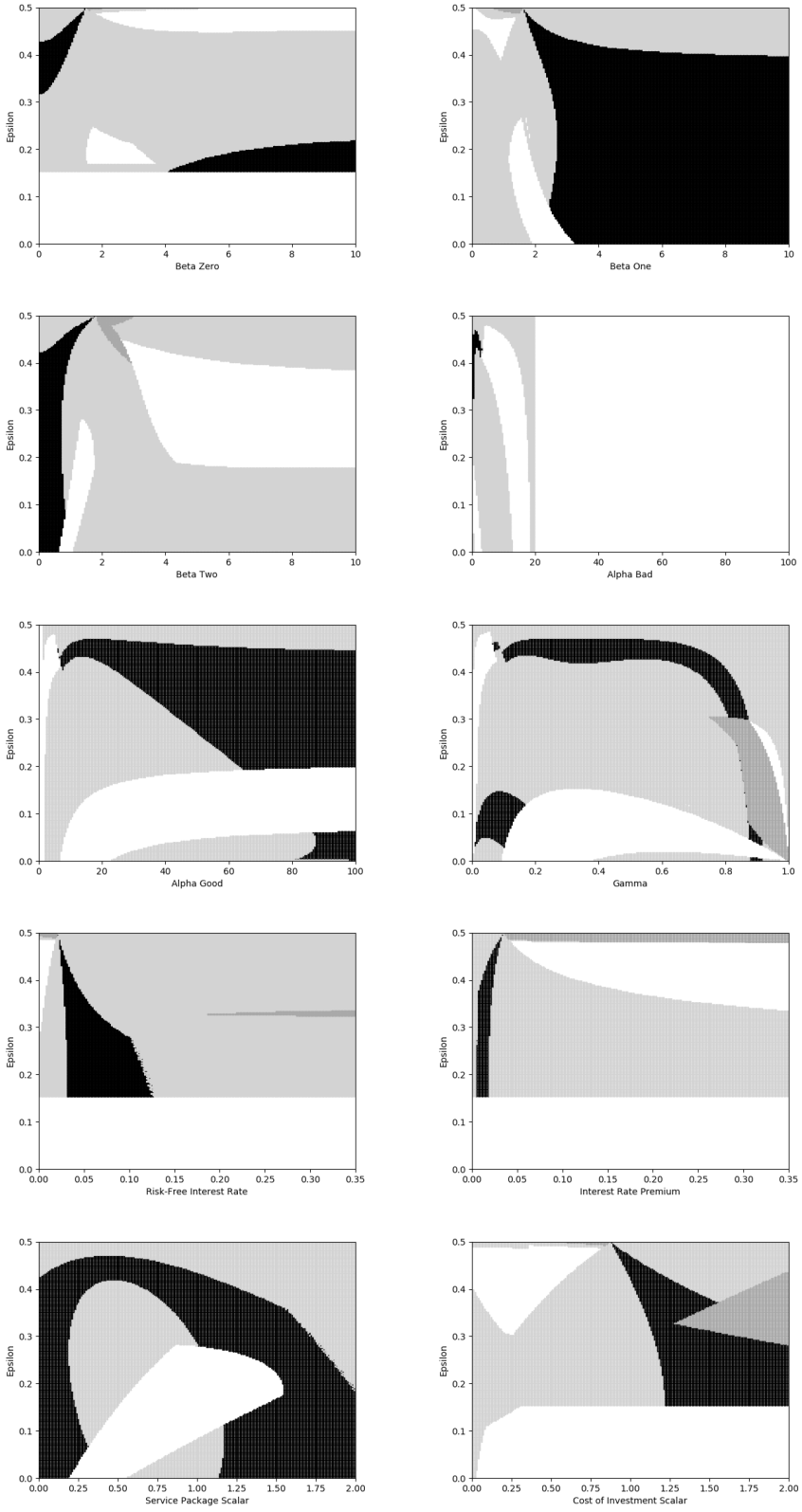


FIGURE 11. The overconfidence dynamics with a focus on ϵ

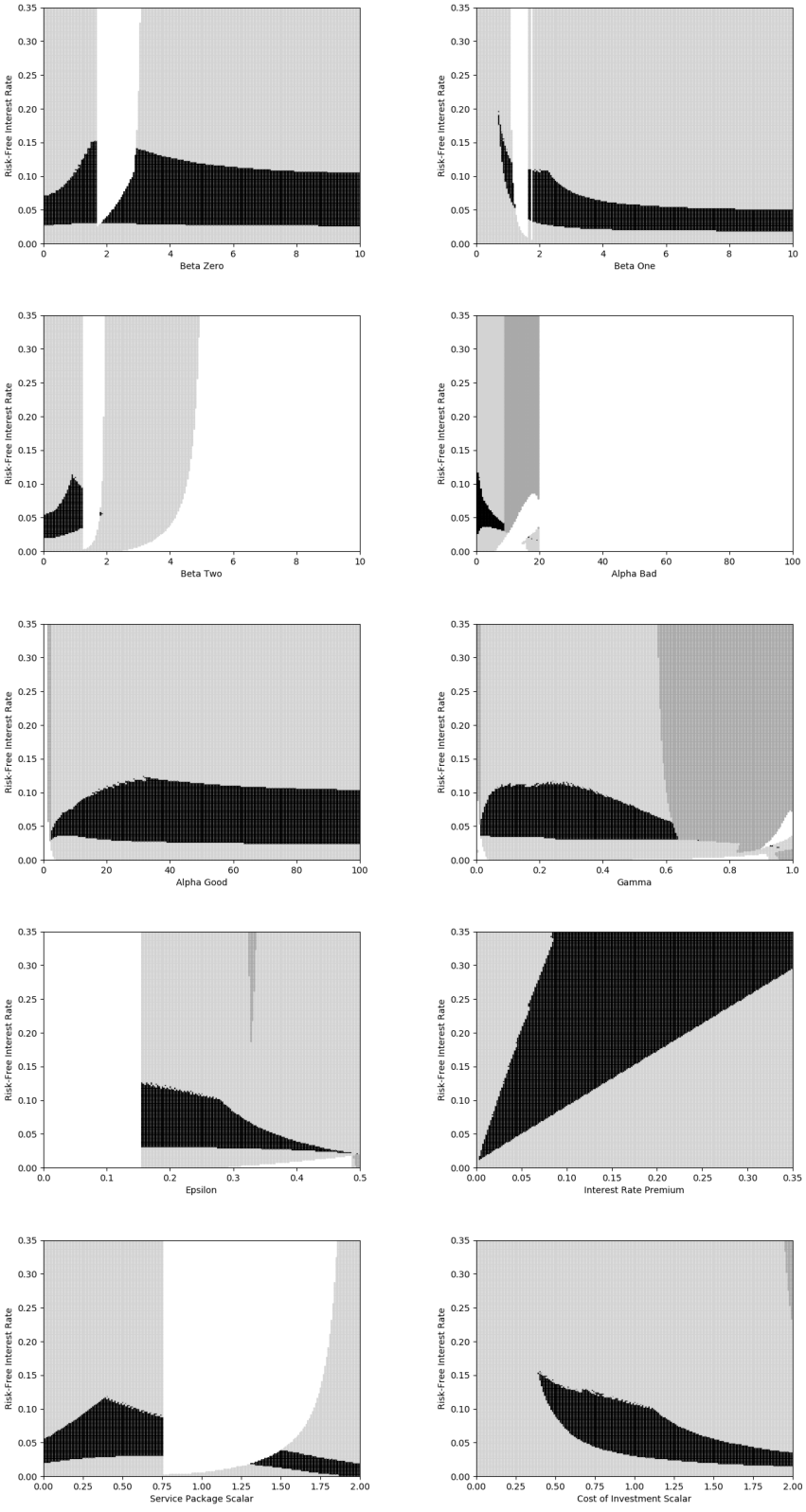


FIGURE 12. The overconfidence dynamics with a focus on $\bar{r}$

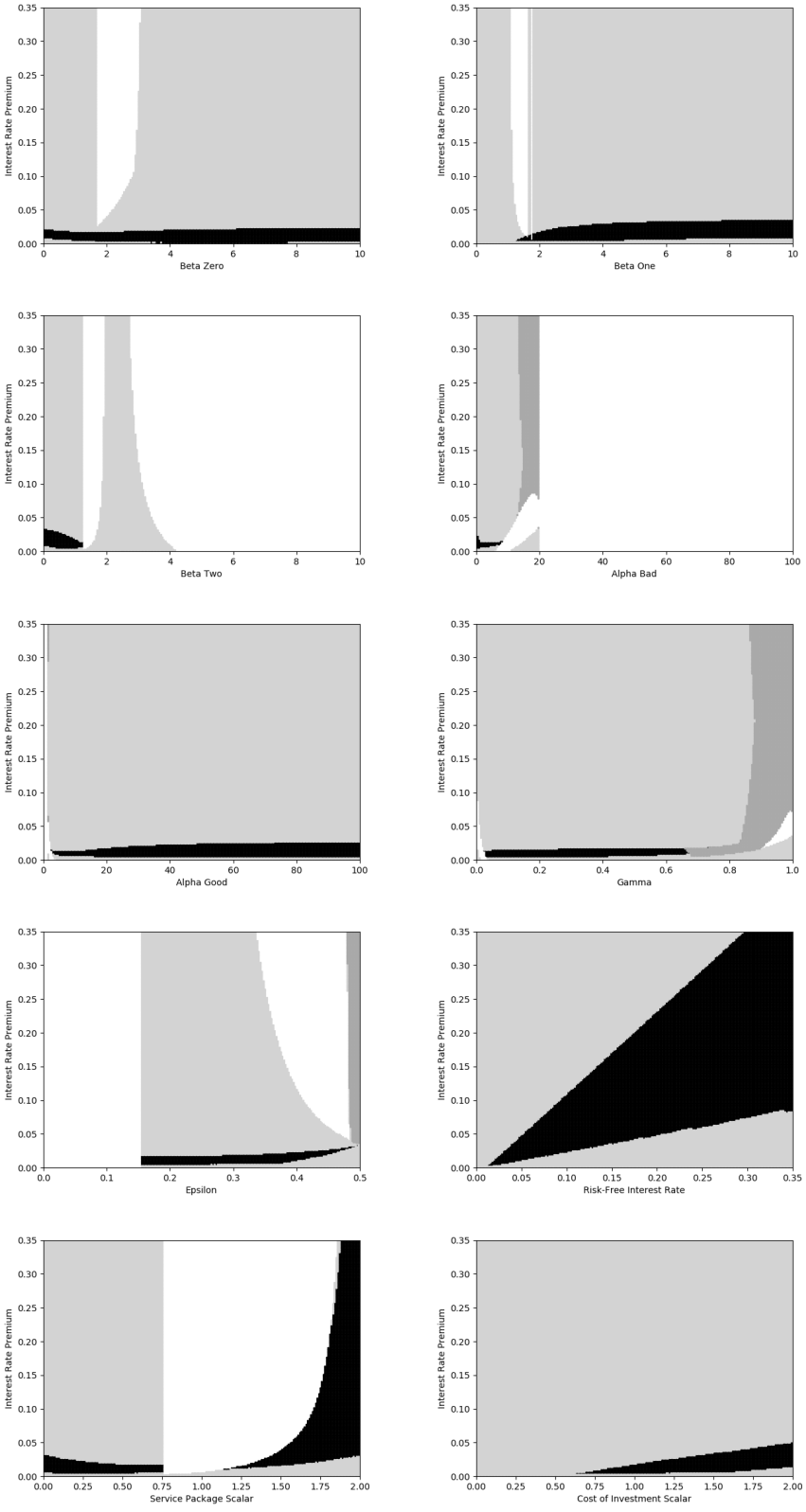


FIGURE 13. The overconfidence dynamics with a focus on $\bar{p}$

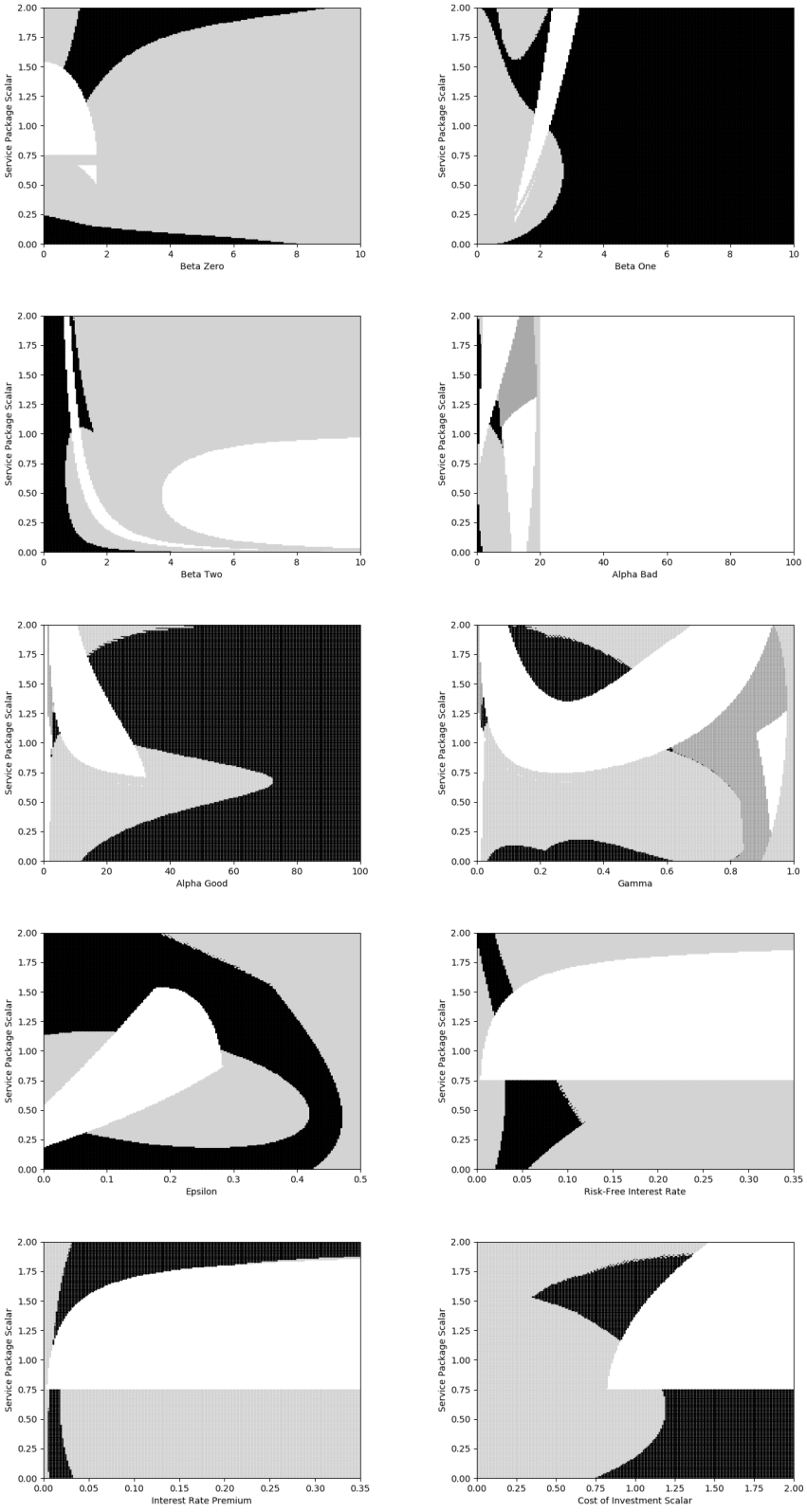


FIGURE 14. The overconfidence dynamics with a focus on $\theta_{\bar{e}}$

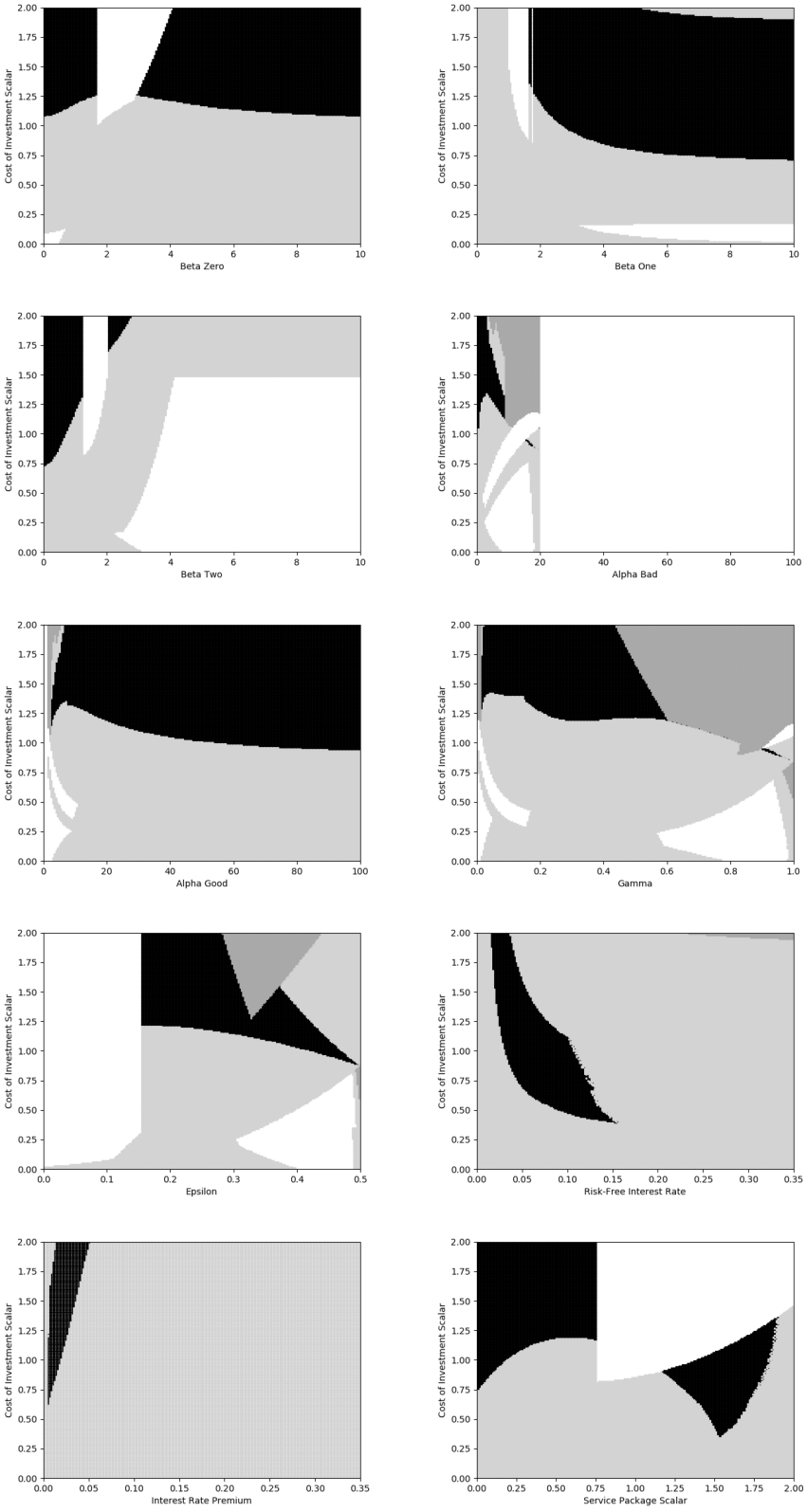


FIGURE 15. The overconfidence dynamics with a focus on θ_I

Bibliography

1. *General agreement on tariffs and trade*, 1947, Source: `http://www.wto.org`.
2. *World trade report 2011*, 2011, Source: `http://www.wto.org`.
3. Philippe Aghion, Pol Antràs, and Elhanan Helpman, *Negotiating free trade*, Journal of International Economics **73** (2007), no. 1, 1–30.
4. Shiri Alon and Gabi Gayer, *Utilitarian preferences with multiple priors*, Econometrica **84** (2016), no. 3, 1181–1201.
5. F.J. Anscombe and R.J. Aumann, *A definition of subjective probability*, Annals of Mathematical Statistics **34** (1963), 199–205.
6. Robert D. Arnott and Peter L. Bernstein, *What risk premium is 'normal'?*, Financial Analysts Journal **58** (2002), no. 2, 64–85.
7. Robert Aumann and Roger Myerson, *Endogenous formation of links between players and coalitions: an application of the shapley value*, The Shapley Value (1988), 175–191.
8. Kyle Bagwell, Chad P. Bown, and Robert W. Staiger, *Is the wto passé?*, Journal of Economic Literature **54** (2016), no. 4, 1125–1231.
9. Kyle Bagwell and Robert W. Staiger, *Regionalism and multilateral tariff cooperation*, Working Paper 5921, National Bureau of Economic Research, February 1997.
10. Kyle Bagwell and Robert W Staiger, *The design of trade agreements*, Handbook of Commercial Policy **1** (2016), 435–529.
11. Richard E. Baldwin and Thorvaldur Gylfason, *A domino theory of regionalism*, Expanding Membership of the European Union, Cambridge University Press, 1995, pp. 25–53.
12. European Central Bank, *The definition of price stability*, `https://web.archive.org/web/*/https://www.ecb.europa.eu/mopo/strategy/pricestab/html/index.en.html`, Accessed: 2019-06-18.
13. B. Douglas Bernheim, Bezalel Peleg, and Michael Whinston, *Coalition-proof nash equilibria i. concepts*, Journal of Economic Theory **42** (1987), no. 1, 1–12.
14. Jagdish Bhagwati, *Regionalism and multilateralism: an overview*, New dimensions in regional integration **22** (1993), 51.
15. Colin Camerer and Dan Lovallo, *Overconfidence and excess entry: An experimental approach*, American Economic Review **89** (1999), no. 1, 306–318.
16. Michael Suk-Young Chwe, *Farsighted coalitional stability*, Journal of Economic Theory **63** (1994), no. 2, 299 – 325.
17. Marta Coelho, David de Meza, and Diane Reyniers, *Irrational exuberance, entrepreneurial finance and public policy*, International Tax and Public Finance **11** (2004), no. 4, 391–417.
18. Arnold C. Cooper, Carolyn Y. Woo, and William C. Dunkelberg, *Entrepreneurs' perceived chances for success*, Journal of Business Venturing **3** (1988), no. 2, 97 – 108.
19. Hervé Crès, Itzhak Gilboa, and Nicolas Vieille, *Aggregation of multiple prior opinions*, Journal of Economic Theory **146** (2011), no. 6, 2563–2582.
20. Aswath Damodaran, *Return on capital (roc), return on invested capital (roic) and return on equity (roe): Measurement and implications*, SSRN Electronic Journal (2007).
21. Eric Danan, Thibault Gajdos, Brian Hill, and Jean-Marc Tallon, *Robust social decisions*, American Economic Review **106** (2016), no. 9, 2407–25.
22. Jean-Etienne de Bettignies and James A. Brander, *Financing entrepreneurship: Bank finance versus venture capital*, Journal of Business Venturing **22** (2007), no. 6, 808 – 832.
23. Werner F. M. De Bondt and Richard H. Thaler, *Financial decision-making in markets and firms: A behavioral perspective*, Working Paper 4777, National Bureau of Economic Research, June 1994.

24. Leonidas Enrique de la Rosa, *Overconfidence and moral hazard*, Games and Economic Behavior **73** (2011), no. 2, 429 – 451.
25. Peter A. Diamond, *Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment*, Journal of Political Economy **75** (1967), no. 5, 765–766.
26. Juergen Eichberger and Ruediger Pethig, *Constitutional choice of rules*, European Journal of Political Economy **10** (1994), no. 2, 311 – 337.
27. Giovanni Facchini, Peri Silva, and Gerald Willmann, *The customs union issue: Why do we observe so few of them?*, Journal of International Economics **90** (2013), no. 1, 136 – 147.
28. Philippa Foot, *The problem of abortion and the doctrine of double effect*, Oxford Review **5** (1967), 5–15.
29. Taiji Furusawa and Hideo Konishi, *Free trade networks*, Journal of International Economics **72** (2007), no. 2, 310–335.
30. Thibault Gajdos and Feriel Kandil, *The ignorant observer*, Social Choice and Welafre **31** (2008), 193–232.
31. Sanjeev Goyal and Sumit Joshi, *Bilateralism and free trade*, International Economic Review **47** (2006), no. 3, 749–778.
32. Simon Grant, Atsushi Kajii, Ben Polak, and Zvi Safra, *Generalized utilitarianism and harsanyi's impartial observer theorem*, Econometrica **78** (2010), no. 6, 1939–1971.
33. Gene M Grossman, *The purpose of trade agreements*, Handbook of Commercial Policy **1** (2016), 379–434.
34. John Harsanyi, *An equilibrium-point interpretation of stable sets and a proposed alternative definition*, Management Science **20** (1974), no. 11, 1472–1495.
35. John C. Harsanyi, *Cardinal utility in welfare economics and in the theory of risk-taking*, Journal of Political Economy **61** (1953), no. 5, 434–435.
36. ———, *Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility*, Journal of Political Economy **63** (1955), no. 4, 309–321.
37. ———, *Rational behaviour and bargaining equilibrium in games and social situations*, Campbirdge University Press, 1977.
38. P. Jean-Jacques Herings, Ana Mauleon, and Vincent Vannetelbosch, *Farsightedly stable networks*, Games and Economic Behavior **67** (2009), no. 2, 526–541.
39. P.J.J. Herings, A. Mauleon, and V. Vannetelbosch, *Stability of networks under level-k farsightedness*, Research Memorandum 030, Maastricht University, Graduate School of Business and Economics (GSBE), January 2014.
40. Matthew O Jackson and Asher Wolinsky, *A strategic model of social and economic networks*, Journal of economic theory **71** (1996), no. 1, 44–74.
41. Alexander Keck and Andreas Lendle, *New evidence on preference utilization*, WTO Staff Working Papers ERSD-2012-12, World Trade Organization (WTO), Economic Research and Statistics Division, September 2012.
42. Hiau Looi Kee, Alessandro Nicita, and Marcelo Olarreaga, *Estimating trade restrictiveness indices*, The Economic Journal **119** (2009), no. 534, 172–199.
43. Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji, *A smooth model of decision making under ambiguity*, Econometrica **73** (2005), no. 6, 1849–1892.
44. James Lake, *Free trade agreements as dynamic farsighted networks*, Economic Inquiry **55** (2017), no. 1, 31–50.
45. Augustin Landier and David Thesmar, *Financial contracting with optimistic entrepreneurs: Theory and evidence*, October 2003.
46. Erik Lie and Heidi J. Lie, *Multiples used to estimate corporate value*, Financial Analysts Journal **58** (2002), no. 2, 44–54.
47. R. G. Lipsey and Kelvin Lancaster, *The general theory of second best*, The Review of Economic Studies **24** (1956), no. 1, 11–32.
48. Giovanni Maggi, *International trade agreements*, Handbook of International Economics, vol. 4, 2014, pp. 317–390.
49. Ulrike Malmendier and Geoffrey Tate, *Ceo overconfidence and corporate investment*, The Journal of Finance **60**, no. 6, 2661–2700.
50. Michael Manove and Atilano Jorge Padilla, *Banking (Conservatively) With Optimists*, CEPR Discussion Papers 1918, C.E.P.R. Discussion Papers, June 1998.

51. Paul Missios, Kamal Saggi, and Halis Murat Yildiz, *External trade diversion, exclusion incentives and the nature of preferential trade agreements*, Journal of International Economics **99** (2016), 105–119.
52. Philippe Mongin, *Consistent bayesian aggregation*, Journal of Economic Theory **66** (1995), 313–351.
53. ———, *The impartial observer theorem of social ethics*, Economics and Philosophy **17** (2001), 147–179.
54. Tobias J. Moskowitz and Annette Vissing-Jorgensen, *The returns to entrepreneurial investment: A private equity premium puzzle?*, American Economic Review **92** (2002), no. 4, 745–778.
55. Ryo Ichi Nagahisa and Sang Chul Suh, *A characterization of the walras rule*, Social Choice and Welfare **12** (1995), no. 4, 335–352.
56. Leandro Nascimento, *The ex ante aggregation of opinions under uncertainty*, Theoretical Economics **7** (2012), 535–570.
57. U.S. Department of the Treasury, *Daily treasury yield curve rates*, `https://web.archive.org/web/*/https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield`, Accessed: 2019-06-18.
58. Xiangyu Qu, *Separate aggregation of beliefs and values under ambiguity*, Economic Theory **63** (2017), no. 2, 503–519.
59. H. Raiffa, *Decision analysis - introductory lectures on choices under uncertainty*, Addison Wesley, Reading, MA, 1970, traduction française: Analyse de la décision : introduction aux choix en avenir incertain, Dunod, 1973.
60. John Rawls, *A theory of justice*, Belknap, 1971.
61. Debraj Ray and Rajiv Vohra, *The farsighted stable set*, Econometrica **83** (2015), no. 3, 977–1011.
62. Marco Da Rin, Thomas F. Hellmann, and Manju Puri, *A survey of venture capital research*, Working Paper 17523, National Bureau of Economic Research, October 2011.
63. Kamal Saggi, Alan Woodland, and Halis Murat Yildiz, *On the relationship between preferential and multilateral trade liberalization: the case of customs unions*, American Economic Journal: Microeconomics **5** (2013), no. 1, 63–99.
64. Kamal Saggi and Halis Murat Yildiz, *Bilateralism, multilateralism, and the quest for global free trade*, Journal of International Economics **81** (2010), no. 1, 26–37.
65. Michael Sandel, *Justice: What's the right thing to do?*, Farrar, Straus and Giroux, 2010.
66. Stefano Scarpetta, Philip Hemmings, Thierry Tressel, and Jaejoon Woo, *The role of policy and institutions for productivity and firm dynamics*, (2002), no. 329.
67. Kyoungwon Seo, *Ambiguity and second-order belief*, Econometrica **77** (2009), no. 5, 1575–1605.
68. Adam Smith, *The theory of moral sentiments*, London: A. Millar, 1759.
69. Shelley Taylor and Jonathon D. Brown, *Illusion and well-being: A social psychological perspective on mental health*, **103** (1988), 193–210.
70. ———, *Positive illusions and well-being revisited: Separating fact from fiction*, **116** (1994), 21–7; discussion 28.
71. Masako Ueda, *Banks versus venture capital: Project evaluation, screening, and expropriation*, The Journal of Finance **59** (2004), no. 2, 601–621.
72. Roope Uusitalo, *Homo entreprenaurus?*, Applied Economics **33** (2001), no. 13, 1631–1638.
73. William S. Vickrey, *Measuring marginal utility by reaction to risk*, Econometrica **13** (1945), no. 4, 319–333.
74. Laurent Vilanova, Nadege Marchand, and Walid Hichri, *Financing and advising with (over)confident entrepreneurs : an experimental investigation*, Tech. report, 2015.
75. John von Neumann and Oskar Morgenstern, *Theory of games and economic behavior*, Princeton University Press, 1944.
76. Jin Zhang, Zhiwei Cui, and Lei Zu, *The evolution of free trade networks*, Journal of Economic Dynamics and Control **38** (2014), 72–86.
77. Jin Zhang, Licun Xue, and Lei Zu, *Farsighted free trade networks*, International Journal of Game Theory **42** (2013), no. 2, 375–398.