Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

UNIVERSITÉ DE LA GRANDE RÉGION
UNIVERSITÄT DER GROSSREGION

Four Essays in Economic Theory: Incentives and Interaction

Vom Fachbereich Wirtschaftswissenschaften
der Rheinland-Pfälzischen Technischen Universität
Kaiserslautern-Landau genehmigte

Dissertation

vorgelegt von

Tom Rauber, M.Sc.

Tag der mündlichen Prüfung: *19. Juli 2024*
Dekan: *Prof. Dr. Jan Wenzelburger*

Vorsitzender: *Prof. Dr. Florian Sahling*
Berichterstatter: *Prof. Dr. Philipp Weinschenk*
Prof. Dr. Çağıl Koçyiğit

D 386
(2024)

Acknowledgements

Pursuing a Ph.D. is a journey, a journey full of excitement, joy, thirst for knowledge, enlightenment, but also frustration, failures, and sometimes even tears. I am tremendously grateful to my family, friends, colleagues, and academic teachers who accompanied me on my journey and whose constant and multifaceted support enabled me to reach its final destination.

First and foremost, I am thankful to my first advisor, *Professor Philipp Weinschenk*, who not only encouraged me to embark on this voyage but also provided me with the best possible means of travel – a DFG-funded position – any Ph.D. student could wish for. After the departure of my journey, whenever I felt lost, his continuous, persistent, and reliable support in any academic and personal matter guided my way. At the same time, I am infinitely grateful for the granted freedom as well as his trust in my person and character that allowed me to conduct research stays at the Universities of Luxembourg and Cambridge, which fostered my personal and academic maturation.

Whilst my voyage brought me to Luxembourg, I experienced an extremely productive period in my Ph.D. due to the vibrant work environment at the Centre for Logistics and Supply Chain Management. Therefore, I owe an immense debt of gratitude to *Professor Çağıl Koçyiğit*, who not only invited me to Luxembourg but also immediately agreed to work with me as my second advisor. Thank you for showing me the perfect role model of an early-stage professor.

Before it came to an end, my journey also took me to the British Isles, where, in the inspiring halls of the time-honored University of Cambridge, I found valuable inspiration for my research. Therefore, I am extraordinarily thankful to *Professor Flavio Toxvaerd*, who invited me to this unique temple of science and with whom I had fruitful discussions about my work.

As the focal point of my journey, however, the University of Kaiserslautern-Landau provided me with an academic home by offering invaluable support and an excellent research atmosphere. In particular, I would like to thank *Professor Daniel Heyen* not only for always encouraging and helping me on my journey but also for sharing his rich experience and honest advice about academia. To *Professor Jan*

Wenzelburger, I am thankful for the insightful conversations we had on my "macro project". My co-author on this project, *Paul Ritschel*, deserves special thanks for the excellent cooperation on this research endeavor that led to my first publication. It was a great honor and delightful pleasure to spend my journey with *Paul* as a colleague, co-author, dear friend, and traveling companion. For introducing me to fascinating topics in environmental economics, I am grateful to my co-author, *Dr. Fabian Naumann*, with whom I really enjoyed working with.

Further special thanks are due to *Professor Florian Sahling* for chairing my doctoral committee and guiding me through the organizational procedures and to *Professor Axel Gautier* for assessing my thesis for the University of the Greater Region. My current and former colleagues, *Joshua Bißbort*, *Frederik Holtel*, *Dr. Simon Koch*, *Dr. Helena Krebs*, *Professor Mario Liebensteiner*, *Jan Münning*, *Dr. Miruna Sarbu*, and *Franziska Schmidt*, who have generously supported me along my way with their help, personal experience, and research expertise, should also not go unmentioned.

Even though my doctoral journey was purely academic, it would not have been successful without the immense support I received from my private environment. My parents, *Gabi* and *Armin Rauber*, to whom this thesis is dedicated, were always by my side and helped me along the journey in every way imaginable, primarily through their wisdom that navigated me through the darkest hours of my Ph.D.; no son could wish for more. I am also incredibly thankful to my girlfriend, *Thelma Siboli*, in whose eternal love and deep care I always find peace and serenity, while her continuous encouragement and emotional support gave me the strength to continue on my way to the very end; I could not wish for any other girl to spend the rest of my life with. Lastly, I would like to thank my old friends, *Bence Arany, MBA*, *Robin Bauer*, *Dr. Mary Lucia Darst*, *Dr. Simon Föll*, *cand. Dr. Tim Kreuz*, *Ann Louie Li, MBA*, *Stephan Makiefka*, and *Enis Paloji*, for believing in me and for teaching me the essential values that directed my journey to its successful completion.

Tom Rauber
Kaiserslautern, April 2024

Abstract

Incentives are an indispensable tool in economics, as they allow for aligning interests and improving organizational performance. By exploring numerous novel aspects of incentives in different economic subfields, this cumulative dissertation expands our theoretical understanding of incentives.

First, dynamic interaction is shown to undermine the effectiveness of incentives, even to the extent that their effect is completely reversed. Higher-powered incentives then leave all interacting parties worse off, while at the same time reducing the likelihood that the interaction will be brought to a successful conclusion. This finding arises not only with exogenously given but also endogenously designed incentives. Second, a simple mechanism is proposed to circumvent adverse free-riding incentives that inevitably result from the public good nature of abatement in the context of global greenhouse gas emissions. Enabling countries to establish a joint cap-and-trade system and allowing them to determine its design endogenously through negotiations may overcome free-riding incentives and implement efficient emissions levels since countries can implicitly share efficiency gains under this regime. Third, linking the importance of saving incentives to the capital dependence of the production sector reveals a simple condition determining whether a monopolistic or competitive banking sector induces the greatest level of long-run growth and welfare in an economy. For labor-intensive production, higher-powered saving incentives provided by competitive banks drive capital accumulation and enhance welfare, while, for capital-intensive production, it is the institutional investments by a monopolistic bank that yield a more favorable outcome.

Overall, as its inherent incentive perspective extends contract theory, environmental economics, and macroeconomics, this dissertation emphasizes the pivotal role that incentives play in various forms of economic interaction.

Table of Contents

List of Figures

List of Tables

Introduction

Background. Human interaction constitutes the central building block in any society. Individuals constantly interact not only with their families, friends, colleagues, superiors, and subordinates but also with various institutions and organizations. While its outcome may be crucial from an individual or societal perspective, interaction is often strategic in nature, i.e., its actual outcome for one agent depends on the decisions made by others. Examples range from arms races (see, e.g., Baliga and Sjöström, 2004; Jackson and Morelli, 2009; Powell, 1993) and voting (see, e.g., Arrow, 1950; Gibbard, 1973; Satterthwaite, 1975) in political science over economic interaction in the form of bidding in auctions (see, e.g., Milgrom and Weber, 1982; Myerson, 1981; Vickrey, 1961) and negotiations (see, e.g., Binmore et al., 1986; Muthoo, 1999; Rubinstein, 1982) to the formation of social networks (see, e.g., Jackson, 2005; Jackson and Wolinsky, 1996; Myerson, 1991).

Recognizing its importance and omnipresence, the formalization and analysis of strategic interaction became an integral part of modern economic theory. In particular, after pioneering contributions on oligopoly pricing by Cournot (1838), Bertrand (1883), and Edgeworth (1925), the conceptual work by Neumann and Morgenstern (1944) as well as Nash (1950), extended and refined by Selten (1965, 1975) and Harsanyi (1967–1968), built the foundation of game theory as a pivotal subfield in economics. Since the 1980s, game theory has shaped a new era in economic research, as its methodology enables precise predictions regarding the outcome of strategic interaction between a small number of agents (Rubinstein, 1990). Accordingly, a flourishing field of research has emerged, which provides the tools required to analyze a wide variety of different situations of strategic interaction (Samuelson, 2016).

Rigorously applying this tool kit gave rise to the theory of incentives as a subfield in microeconomics, which greatly revolutionized economic thinking (Dixit & Besley, 1997).[1] According to Cambridge Dictionary's definition, which has also been adopted in academia, incentives are broadly defined as *"something, often*

1. Some economists even compare the methodological effects of incentive theory on microeconomics with those that Einstein's theory of relativity had on physics (Dixit & Besley, 1997).

money or a prize, offered to make someone behave in a particular way" (Cambridge Business English Dictionary, 2023, para. 1). Setting up such incentives becomes relevant whenever strategic interaction between the parties is overshadowed by both *(i)* conflicting objectives and *(ii)* asymmetric information (Laffont & Martimort, 2002). Since many forms of economic interaction indeed reflect these characteristics, a sound understanding of the effect and design of incentives is crucial for numerous applications (cf. literature syntheses by Bolton and Dewatripont, 2005; Laffont and Martimort, 2002; Salanié, 2005). Prendergast (1999, p. 7) therefore emphasizes the central role that incentives play by describing them as *"the essence of economics."*

Consequently, much research has been devoted ever since to exploring how incentives shape agents' behavior and determine organizational performance. The principal-agent framework introduced by Ross (1973) and Jensen and Meckling (1976) became the workhorse model in this endeavor since it allows for studying various incentive problems. The literature on the theory of incentives resulting from this canonical model is vast and multifaceted: it explores incentives that are performance-based (see, e.g., Fudenberg et al., 1990; Herweg et al., 2010; Holmström, 1979), team-based (see, e.g., Che and Yoo, 2001; Holmström, 1982; Kandel and Lazear, 1992), tournament-based (see, e.g., Casas-Arce and Martinez-Jerez, 2009; Lazear and Rosen, 1981; Nalebuff and Stiglitz, 1983), and career-based (see, e.g., Dewatripont et al., 1999; Lazear, 1979, 1981).[2] Insights from these and many other contributions to the highly active research on incentives have influenced modern microeconomics like no others and have awarded the theory of incentives two Nobel Memorial Prizes in Economic Sciences (Smirnov & Wait, 2017). As founding fathers of the theory, James Mirrlees and William Vickrey received their Nobel Prize in 1996 *"for their fundamental contributions to the economic theory of incentives under asymmetric information"* (Royal Swedish Academy of Sciences, 1996, para. 1), while it was awarded to Oliver Hart and Bengt Holmström in 2016 for their seminal work on contracts as incentive mechanisms.

However, research on incentives is far from exhausted, not to mention completed, with many applications left to discover (Bonner & Sprinkle, 2002; Martin et al., 2019; Mirrlees, 1997). Gibbons (1998, p. 130) underlines the field's enormous potential by even suspecting that *"much of the best economics on this subject is still to come."* In light of this substantial need for further research, the cumulative dissertation at hand is dedicated to researching incentives in order to expand our economic understanding of their effectiveness and design. Despite considering different economic contexts, all four articles in this thesis address incentives, making contributions to our comprehension of this essential topic. Before the synopsis be-

2. For excellent overviews and discussions of the literature on incentives, see Gibbons (2005), Lazear (2018), and Prendergast (1999).

low provides a more detailed summary of all four articles, their individual contributions to the theory of incentives are highlighted on a more general level.

Article 1 studies an incentive problem in an organizational context. It considers interaction between members of a team of agents to identify two manifestations of counterproductive incentives resulting from dynamic interaction: higher-powered financial incentives may harm the agents and reduce the probability that they succeed with their assigned task. The follow-up Article 2 adopts a design perspective by including a principal in the model who dynamically interacts with the agents. It examines so far unexplored detrimental effects of incentive mechanisms: a more favorable economic environment may induce incentives that leave all interacting parties worse off, while also deteriorating the probability that the agents succeed with their joint task. Article 3 focuses on incentives in an environmental context. It analyzes the interaction between two countries, proposing a simple mechanism to overcome free-riding incentives that otherwise prevent countries' greenhouse gas emissions from being reduced to the globally efficient level. Article 4 addresses saving incentives in a banking context by investigating strategic interaction between consumers and profit-maximizing banks. It contributes to the ongoing debate on the optimal degree of competition among banks (cf. Coccorese, 2017) and solves an empirical puzzle by tying the importance of saving incentives to the capital dependence of the production sector.

Synopsis. The articles in Chapter 1 and Chapter 2 of this dissertation belong to the field of contract theory and thus cover the very heart of the theory of incentives. Loosely speaking, both investigate perverse effects of incentives that result from dynamic interaction. Chapter 1 presents the article ***"Dynamic Interaction and (In)effectiveness of Financial Incentives"***, which is joint work with Philipp Weinschenk. It is currently under review at the *RAND Journal of Economics* and a working paper version circulates as Rauber and Weinschenk (2024). The article considers a dynamic incentive problem with an exogenous incentive scheme: rational agents work on a joint project for a finite number of periods. If the agents succeed with the project, each of them receives a time-dependent reward, and the game ends; otherwise, the game moves on to the next period until the project deadline is reached. Project success in a certain period materializes with a probability determined by the efforts that the agents exert in the respective period.

In this setting, we explore a novel mechanism undermining the effectiveness of financial incentives, which originates naturally from dynamic interaction. Although higher rewards in the current period unambiguously motivate the agents to exert more effort in that period, they may also discourage agents' effort provision in previous periods. This dynamic discouragement effect is the driving force that can engender interesting and intriguing consequences for *(i)* the total probability of project success ("success rate") and *(ii)* the agents' expected payoffs. First, the effort reduction in prior periods may counteract the positive effect of current efforts insofar as the overall success rate reduces. We refer to this situation as a *success reversal* since the effect of incentives on success is reversed, i.e., higher incentives for success render success less likely. A success reversal arises due to two mechanisms: either effort is shifted into an "unproductive period" – one that is unlikely to be reached –, or a weak increase in current efforts triggers a strong discouragement effect in previous periods. Second, if there are multiple agents, a *payoff reversal* can occur, i.e., higher rewards may harm the agents by reducing their expected payoffs. The intuition is as follows. Due to the team externality – each agent benefits from the efforts exerted by teammates – the discouragement effect may harm the agents' expected payoff to the extent that the direct positive effect of higher rewards is overcompensated. We also show that although both reversals result from the discouragement effect, neither one is necessary or sufficient for the existence of the other.

Chapter 2 is entitled ***"Detrimental Incentive Mechanisms in Dynamic Principal-Agent Relationships"***. An extended version of this article, which has greatly benefited from the cooperation with my supervisor, Philipp Weinschenk, is currently being reviewed by the *Journal of the European Economic Association* and available as Rauber and Weinschenk (2023). The paper argues that the above discussed adverse effects of incentives may also arise as equilibrium phenomena. To do so, a principal who designs the agents' incentives is added to the model of Chapter 1. The principal acts as a project owner, obtaining a time-dependent revenue upon successful project completion. Since agents' efforts are non-contractible, strategic interaction between the principal and the agents is subject to moral-hazard. Accordingly, the principal endogenously designs incentives for the team of agents by promising them part of the revenue if they succeed with the project. The following question then arises naturally in this setting: what is the impact of a more favorable economic environment in the sense of higher project revenues? One would expect that higher revenues clearly benefit the principal. Since the principal would use part of the increased revenues to provide stronger incentives, one is further tempted to assume that higher revenues are inevitably beneficial for the agents and improve the likelihood of project completion.

However, if the principal deploys incentive mechanisms in the form of spot contracts or long-term contracts under limited commitment – both are shown to result in an identical incentive design –, none of these conjectures is necessarily true. The driving force behind these counterintuitive findings is again a negative intertemporal effect on the effort provision in previous periods. Higher revenues in the current period induce an incentive design that leads to lower efforts in the prior period. Consequently, a *success reversal*, interpreted here in the sense that a more lucrative economic environment worsens the total probability for project success, can emerge through precisely the two mechanisms described above: higher revenues result in a contract design that either *(i)* substitutes efforts to an "unproductive" period or *(ii)* induces slightly higher efforts in the current period at the expense of drastically decreased efforts in previous periods. For a *payoff reversal*, here in the sense that agents suffer from the contract design induced by a more profitable project, an additional mechanism that originates from the principal's optimizing behavior arises. As a higher current revenue makes project success in that period more attractive relative to success in the previous period, it can be optimal for the principal to reduce incentives in the prior period in order to increase the probability of reaching the relatively more lucrative period. The reduction in the agents' incentives can be strong enough to decrease their overall expected payoff. Due to this additional mechanism resulting from endogenous rewards, payoff reversal may occur not only in teams but also in single-agent settings. Even more surprisingly, the principal may also experience lower expected profits if the project becomes more lucrative, a situation referred to as *profit reversal*. The intuition is as follows. Higher incentives also have a negative intertemporal effect since they render setting incentives in earlier periods more costly. The principal however cannot credibly account for the negative intertemporal effect of higher incentives when designing the incentive mechanisms either because of their sequential structure or lacking commitment. Although profit reversals vanish if the principal can fully commit to long-term contracts, success and payoff reversals are shown to persist under full commitment since the contract design solely focuses on maximizing the principal's expected profit.

Chapter 3, ***"Designing Emissions Trading Schemes: Negotiations on the Amount & Allocation of Permits"***, is joint work with Fabian Naumann. It belongs to the subfield of environmental economics and represents the most applied article in this dissertation. The article is currently being peer-reviewed by the *Journal of Public Economics*, while a working paper version is already accessible online as Rauber and Naumann (2023). We explore and verify Weitzman's (2014, p. 34) conjecture that a *"system based on negotiating aggregate emissions [...] could, in principle, embody [a] countervailing force against the global warming externality."* To do so, our paper sets up a simple two-country model of greenhouse gas emissions. Both countries benefit from lower emissions, regardless of where the reduction has taken place, i.e., there is a positive externality, whereas reduction costs are only incurred by the country that has implemented the abatement measures. This setup gives rise to a substantial free-riding incentive: each country is anxious to leave costly abatement activities to the other. As a result, global greenhouse gas emissions exceed the efficient level that is optimal from a societal perspective.

We then explore whether a simple mechanism can overcome free-riding incentives: enabling the countries to establish a joint cap-and-trade system and allowing them to design this scheme endogenously through negotiations. Indeed, this procedure is shown to implement the efficient level of emissions and to maximize overall welfare if the countries are sufficiently symmetric. This holds true even if the countries can strategically opt out of the negotiations. The intuition is as follows. By equating marginal costs, the cap-and-trade system itself ensures that *any* reduction in emissions is achieved in the most cost-efficient way. The countries then negotiate an emissions cap for the cap-and-trade system that implements the efficient level of emissions and thus maximizes welfare. As the countries also determine the initial allocation of certificates, they agree on an allocation of permits in which the highest welfare level is shared equally among them. Hence, countries use the allocation of certificates as a means for implicit side payments, which in turn induces cooperative behavior. Since a country cannot be awarded more than the entire share of certificates, the means for providing side payments in the cap-and-trade system is limited. In particular, we find that if the countries are too heterogeneous, then side payments may be insufficient to make both countries agree on the efficient cap.

The article in Chapter 4, ***"Banking Competition and Capital Dependence of the Production Sector: Growth and Welfare Implications"***, is a result of joint research with Paul Ritschel. A slightly shorter version of this article, which omits all considerations regarding financial stability, is published as Rauber and Ritschel (2024) in the *International Review of Economics & Finance*. As the most technical article in this dissertation, it belongs to the subfield of microfounded macroeconomics and addresses Goldsmith's (1969, p. 408) famous question *"Does finance make a difference...?"* in a banking context. More precisely, we embed a banking sector in a standard overlapping generations growth model a la Diamond (1965) in order to study the role of banking competition for economic growth and welfare. Our paper then compares three versions of the model: *(i)* absence of financial intermediation, i.e., households invest directly in the production sector, *(ii)* a perfectly competitive banking sector that intermediates households' investments in the production sector, and *(iii)* a monopolistic bank that acts as financial intermediary.

This comparison yields our most striking theoretical contribution, namely a simple condition determining whether monopolistic or competitive banking is favorable in the long run, both from a growth and welfare perspective. Having a monopoly bank is beneficial if the economy's production sector is heavily dependent on capital and the bank's dividend payments are sufficiently restricted. By contrast, if production is labor-intensive, then competition in the banking sector maximizes growth and welfare. Our finding can explain seemingly contradictory empirical findings regarding the correlation between banking competition and economic growth (see, Beck et al., 2004; Cetorelli and Gambera, 2001; Deidda and Fattouh, 2005; Hoxha, 2013; Maudos and Fernandez de Guevara, 2006).

Intuitively, capital accumulation in our model is fed from two channels: private savings and institutional investments in the form of bank equity. More interbank competition affects both channels. While it increases saving incentives – households are granted higher deposit rates for their savings –, it also reduces the intermediation margins claimed by banks and thus their equity. Saving incentives are decisive for capital accumulation in a labor-intensive production environment since households receive a relatively large share of the economy's total income as wage income. A competitive banking sector thus induces greater economic growth than its monopolistic counterpart due to its higher-powered saving incentives. In a capital-intensive production environment, on the contrary, relatively low wage incomes render saving incentives less important. Instead, abundant institutional investments provided by a monopolistic bank supply the production with the required capital and thus yield the greatest level of economic growth. Our result that a banking monopoly can also be welfare-improving is explained in terms of the theory of the second best by Lipsey and Lancaster (1956). Capital accumulation is generally inefficient, as each generation only focuses on maximizing its own (short-run) utility. Introducing a banking monopoly as an additional market imperfection may then improve overall welfare. In particular, for capital-intensive production, private savings are insufficient to implement efficient capital accumulation. Institutional investments from a monopolistic bank then improve welfare by shifting the accumulation path towards a higher level of long-run output.

Our analysis reveals three further insights. First, the presence of banks propels long-run economic growth independently of the degree of interbank competition. Since deposit contracts offered by banks constitute a vehicle for risk-sharing, they incentivize households to release additional funds into the production sector. Second, banks cannot induce persistent business cycles and complex dynamics in an otherwise calm economy. That is, the introduction of a banking sector does not adversely affect qualitative growth patterns in the long run. Third, our model lends support to the competition-fragility hypothesis, stating that a competitive banking system is more vulnerable to banking crises. Not only are competitive banks less able to absorb losses due to their lower stock of equity, but erroneous expectations may also entail more severe adverse consequences under this banking regime due to lower intermediation margins.

References

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, *58*(4), 328–346. [1]

Baliga, S., & Sjöström, T. (2004). Arms races and negotiations. *Review of Economic Studies*, *71*(2), 351–369. [1]

Beck, T., Demirgüç-Kunt, A., & Maksimovic, V. (2004). Bank competition and access to finance: International evidence. *Journal of Money, Credit and Banking*, *36*(3), 627–648. [6]

Bertrand, J. (1883). Théorie mathématique de la richesse sociale. *Journal des Savants*, (1883), 499–508. [1]

Binmore, K., Rubinstein, A., & Wolinsky, A. (1986). The nash bargaining solution in economic modelling. *Rand Journal of Economics*, *17*(2), 176–188. [1]

Bolton, P., & Dewatripont, M. (2005). *Contract theory*. MIT Press. [2]

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society*, *27*(4-5), 303–345. [2]

Cambridge Business English Dictionary. (2023). *Economic incentive*. Retrieved December 7, 2023, from https://dictionary.cambridge.org/dictionary/english/economic-incentive. [2]

Casas-Arce, P., & Martinez-Jerez, F. A. (2009). Relative performance compensation, contests, and dynamic incentives. *Management Science*, *55*(8), 1306–1320. [2]

Cetorelli, N., & Gambera, M. (2001). Banking market structure, financial dependence and growth: International evidence from industry data. *Journal of Finance*, *56*(2), 617–648. [6]

Che, Y.-K., & Yoo, S.-W. (2001). Optimal incentives for teams. *American Economic Review*, *91*(3), 525–541. [2]

Coccorese, P. (2017). Banking competition and economic growth. In J. A. Bikker & L. Spierdijk (Eds.), *Handbook of Competition in Banking and Finance* (pp. 230–263). Edward Elgar Publishing. [3]

Cournot, A. A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses*. L. Hachette. [1]

Deidda, L., & Fattouh, B. (2005). Concentration in the banking industry and economic growth. *Macroeconomic Dynamics*, *9*(2), 198–219. [6]

Dewatripont, M., Jewitt, I., & Tirole, J. (1999). The economics of career concerns, part i: Comparing information structures. *Review of Economic Studies*, *66*(1), 183–198. [2]

Diamond, P. A. (1965). National debt in a neoclassical growth model. *American Economic Review*, *55*(5), 1126–1150. [6]

Dixit, A., & Besley, T. (1997). James mirrlees' contributions to the theory of information and incentives. *Scandinavian Journal of Economics*, *99*(2), 207–235. [1]

Edgeworth, F. Y. (1925). The pure theory of monopoly. In F. Y. Edgeworth (Ed.), *Papers relating to political economy* (pp. 111–142, Vol. 1). MacMillian. [1]

Fudenberg, D., Holmstrom, B., & Milgrom, P. (1990). Short-term contracts and long-term agency relationships. *Journal of Economic Theory*, *51*(1), 1–31. [2]

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, *41*(4), 587–601. [1]

Gibbons, R. (1998). Incentives in organizations. *Journal of Economic Perspectives*, *12*(4), 115–132. [2]

Gibbons, R. (2005). Incentives between firms (and within). *Management Science*, *51*(1), 2–17. [2]

Goldsmith, R. (1969). *Financial structure and development*. Yale University Press. [6]

Harsanyi, J. C. (1967–1968). Games with incomplete information played by bayesian players (parts i, ii, iii). *Management Science*, *14*, 159–182, 320–334, 486–502. [1]

Herweg, F., Müller, D., & Weinschenk, P. (2010). Binary payment schemes: Moral hazard and loss aversion. *American Economic Review*, *100*(5), 2451–2477. [2]

Holmström, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, *10*(1), 74–91. [2]

Holmström, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, *13*(2), 324–340. [2]

Hoxha, I. (2013). The market structure of the banking sector and financially dependent manufacturing sectors. *International Review of Economics & Finance*, *27*, 432–444. [6]

Jackson, M. O. (2005). A survey of network formation models: Stability and efficiency. In G. Demange & M. Wooders (Eds.), *Group formation in economics: Networks, clubs, and coalitions* (pp. 11–57). Cambridge University Press. [1]

Jackson, M. O., & Morelli, M. (2009). Strategic militarization, deterrence and wars. *Quarterly Journal of Political Science*, *4*(4), 279–313. [1]

Jackson, M. O., & Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, *71*(1), 44–74. [1]

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305–360. [2]

Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*, *100*(4), 801–817. [2]

Laffont, J.-J., & Martimort, D. (2002). *The theoy of incentives: The principal agent problem*. Princeton University Press. [2]

Lazear, E. P. (1979). Why is there mandatory retirement? *Journal of Political Economy*, *87*(6), 1261–1284. [2]

Lazear, E. P. (1981). Agency, earnings profiles, productivity, and hours restrictions. *American Economic Review*, *71*(4), 606–620. [2]

Lazear, E. P. (2018). Compensation and incentives in the workplace. *Journal of Economic Perspectives*, *32*(3), 195–214. [2]

Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, *89*(5), 841–864. [2]

Lipsey, R. G., & Lancaster, K. (1956). The general theory of second best. *Review of Economic Studies*, *24*(1), 11–32. [7]

Martin, G. P., Wiseman, R. M., & Gomez-Mejia, L. R. (2019). The interactive effect of monitoring and incentive alignment on agency costs. *Journal of Management*, *45*(2), 701–727. [2]

Maudos, J., & Fernandez de Guevara, J. (2006). Banking competition, financial dependence and economic growth. *MPRA Paper No. 15254*. [6]

Milgrom, P. R., & Weber, R. J. (1982). A theory of auctions and competitive bidding. *Econometrica*, *50*(5), 1089–1122. [1]

Mirrlees, J. A. (1997). Information and incentives: The economics of ccarrots and sticks. *Economic Journal*, *107*(444), 1311–1329. [2]

Muthoo, A. (1999). *Bargaining theory with applications*. Cambridge University Press. [1]

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, *6*(1), 58–73. [1]

Myerson, R. B. (1991). *Game theory: Analysis of conflict*. Harvard University Press. [1]

Nalebuff, B. J., & Stiglitz, J. E. (1983). Prizes and incentives: Towards a general theory of compensation and competition. *Bell Journal of Economics*, 21–43. [2]

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, *36*(1), 48–49. [1]

Neumann, J. v., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press. [1]

Powell, R. (1993). Guns, butter, and anarchy. *American Political Science Review*, *87*(1), 115–132. [1]

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, *37*(1), 7–63. [2]

Rauber, T., & Naumann, F. (2023). Designing emissions trading schemes: Negotiations on the amount & allocation of permits. *Working Paper. Available at SSRN 4636471*. [5]

Rauber, T., & Ritschel, P. (2024). Banking competition and capital dependence of the production sector: Growth and welfare implications. *International Review of Economics & Finance*, *89*, 676–698. [6]

Rauber, T., & Weinschenk, P. (2023). Detrimental incentive mechanisms in dynamic principal-agent relationships. *Working Paper. Available at SSRN 4515328*. [4]

Rauber, T., & Weinschenk, P. (2024). Dynamic interaction and (in)effectiveness of financial incentives. *Working Paper. Available at SSRN 3987829*. [3]

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *American Economic Review*, *63*(2), 134–139. [2]

Royal Swedish Academy of Sciences. (1996). *Press release*. Retrieved December 7, 2023, from https://www.nobelprize.org/prizes/economic-sciences/1996/press-release. [2]

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, *50*(1), 97–109. [1]

Rubinstein, A. (1990). Introduction. In A. Rubinstein (Ed.), *Game theory in economics* (pp. x–xviii). Edward Elgar Publishing. [1]

Salanié, B. (2005). *The economics of contracts: A primer*. MIT Press. [2]

Samuelson, L. (2016). Game theory in economics and beyond. *Journal of Economic Perspectives*, *30*(4), 107–130. [1]

Satterthwaite, M. A. (1975). Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, *10*(2), 187–217. [1]

Selten, R. (1965). Spieltheoretische behandlung eines oligopolmodells mit nachfrageträgheit. *Zeitschrift für die gesamte Staatswissenschaft*, *121*(2). [1]

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4(1), 25–55. [1]

Smirnov, V., & Wait, A. (2017). Contracts, incentives and organizations: Hart and holmström nobel laureates. *Review of Political Economy*, *29*(4), 493–522. [2]

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, *16*(1), 8–37. [1]

Weitzman, M. L. (2014). Can negotiating a uniform carbon price help to internalize the global warming externality? *Journal of the Association of Environmental and Resource Economists*, *1*(1/2), 29–49. [5]

Chapter 1

Dynamic Interaction and (In)effectiveness of Financial Incentives*

Joint with Philipp Weinschenk

Abstract

This article studies the effectiveness of financial incentives in a simple model in which a team of rational agents works on a joint project. We show that project success may become less likely and agents can be worse off if they face higher rewards. These success and payoff reversals arise naturally in dynamic settings. The underlying mechanism is that dynamic interaction undermines the effectiveness of incentives: while stronger incentives raise agents' efforts within a period, they also discourage efforts in preceding periods. Our findings contribute to a better theoretical understanding of the prevailing empirical patterns of project delays and failures.

Keywords: Financial Incentives, Dynamic Interaction, Team Production, Discouragement Effect, Success Reversal, Payoff Reversal.

JEL Classification: C73, D82, J30, M52.

---

*Acknowledgments. We are grateful to Heski Bar-Isaac, Matthew Elliott, Bård Harstad, Fabian Herweg, Daniel Heyen, Simon Koch, Çağıl Koçyiğit, Benny Mantin, Daniel Müller, Paul Ritschel, Miruna Sarbu, Flavio Toxvaerd, and Jan Wenzelburger for valuable feedback and ideas. This paper has also benefited from the audiences at the 2023 North American Summer Meeting of the Econometric Society in Los Angeles and the IAAEU Workshop on Labour Economics in Trier, as well as research seminars at the Universities of Cambridge, Kaiserslautern-Landau, and Luxembourg. Finally, for generous financial support, we owe a great debt of gratitude to Deutsche Forschungsgemeinschaft (Grant No. 431144273).

1.1 Introduction

Incentives play a pivotal role in economics.[1] It is crucial to understand how they shape agents' behavior and determine organizational performance. We argue that our current understanding of financial incentives is incomplete. In particular, we show that dynamic interaction – a situation frequently encountered in practice – undermines the effectiveness of financial incentives, even to the extent that their effect is completely reversed. Financial incentives then become counterproductive in two ways, namely in terms of *(i)* incentivizing agents to succeed in their project and *(ii)* improving agents' well-being. These phenomena are illustrated in the following simple example.

Example 1 (Introductory Example).
Consider two agents $i = 1, 2$ who can work on a project for two periods $t = 1, 2$. In every period, their efforts are perfect substitutes and the project's success probability is $p(e_{t,1}, e_{t,2}) = e_{t,1} + e_{t,2}$, while individual effort costs are $c(e_{t,i}) = e_{t,i}^2$. Given the reward profile $\boldsymbol{z}_1 = (z_1, z_2)$, where z_t is the reward that both agents receive for success in t, Table 1.1 shows that although financial incentives are unambiguously stronger with the reward profile $\boldsymbol{z}_1^{\mathrm{B}}$ than with $\boldsymbol{z}_1^{\mathrm{A}}$, these rewards (i) make project success less likely and (ii) leave agents worse off.

Table 1.1. Equilibrium success rate and expected payoffs.

reward profile	(overall) success rate	(overall) expected payoffs
$\boldsymbol{z}_1^{\mathrm{A}} = (0.95,\ 0.1)$	95%	0.67
$\boldsymbol{z}_1^{\mathrm{B}} = (0.98,\ 0.7)$	88%	0.65

To explore the mechanism which renders financial incentives counterproductive, we build a simple yet flexible model that captures static and dynamic interaction, as well as single-agent and team problems. We suppose that a set of rational agents (male) work on a joint project. Each agent individually decides how much effort to invest. Agents are incentivized to provide efforts by time-dependent rewards that they obtain for successful project completion. Rewards are exogenous from the agents' perspective, i.e., they are either directly determined by the economic environment or designed by an optimizing party. Settings in which rewards are determined by the economic environment range from the exploration of natural

1. Nobel laureate Aumann (2006, p. 17075) even states: *"Economics is all about incentives"*.

resources over the introduction of new technologies to R&D activities (Bolton & Harris, 1999). For rewards designed by a principal (female), we can infer from Rauber and Weinschenk (2023) that *any* reward profile in our setup can be rationalized as an outcome of an optimization process if the designing party has limited commitment power or determines the rewards sequentially. Hence, both forms of counterproductive incentives that we discover are equilibrium phenomena.

Our analysis starts by considering the static version of the model. When agents work on the project only for a single period, financial incentives work as one would expect: stronger incentives motivate agents to invest more effort, which increases the project's success rate, i.e., the overall probability that it is completed, and agents' expected payoffs. While this *motivation effect* – the positive relationship between rewards and effort provision within a period – persists in the dynamic version of the model, another intertemporal mechanism arises that erodes the effectiveness of incentives: higher rewards render success in the preceding period relatively less attractive and thus discourage agents from investing effort in that period. Since this adverse *discouragement effect*[2] can outweigh the motivation effect, the impact of financial incentives may completely reverse. In this case, dynamic interaction gives rise to two intriguing phenomena.

First, we refer to a *success reversal* as a situation in which stronger incentives for project success make success less likely. That is, higher rewards lower the success rate. We distinguish two mechanisms causing success reversals: either a weak motivating effect is opposed by a strong discouragement effect, or the interplay of both effects shifts efforts from a "productive" to an "unproductive" period for project success. Second, although higher incentives increase state-dependent payoffs, they may lower agents' expected payoffs from working on the project. Stronger financial incentives are then detrimental to the agents' well-being, a phenomenon we call *payoff reversal*. Intuitively, as agents also benefit from efforts exerted by their teammates, stronger incentives may lower teammates' efforts due to the discouragement effect to such an extent that the positive effect of higher incentives is overcompensated. Although both phenomena originate from the discouragement effect, we further demonstrate that success and payoff reversals can but do not necessarily need to occur together.

It is noteworthy that dynamic interaction is the only prerequisite necessary to jeopardize the effectiveness of incentives. Since dynamic interaction is common in practice, our paper identifies a highly plausible – yet so far unnoticed – reason why

2. Although the discouragement effect is simple and, in this sense, fundamental, it has not yet been formally investigated in the literature.

incentives may be ineffective. We thereby provide a theoretical explanation for project failures and delays that are regularly reported in the literature and popular press (see, for example, Eizakshiri et al., 2015; Jenkins et al., 1984; Park, 2021). Even if high-powered incentives are provided, agents may strategically withhold efforts, which delays or completely prevents project completion. This finding also helps to understand why performance-based compensation schemes are not extensively used in practice.

Related Literature. By considering rational agents, we pursue a non-behavioral approach to reveal a novel dynamic mechanism as a root cause for counterproductive incentives.[3] In a model of team production, where players choose their efforts sequentially but only once, Winter (2009) also finds a manifestation of success reversals.[4] Intuitively, with sufficiently high rewards, exerting effort becomes a dominant strategy for late movers, allowing early movers to free-ride. Klor et al. (2014) verify the predictions of Winter's model in two lab experiments. Our analysis, however, identifies an utterly different mechanism for success reversals based on dynamic interaction. Moreover, we obtain the result that agents themselves can suffer from stronger financial incentives. We are not aware of any other paper that discovers this surprising insight.

Other related publications are in the literature on dynamic moral hazard in teams. While their models share some features with ours, they focus either on the principal and her optimal incentive design or on other aspects, such as deadlines and team size, rather than on how financial incentives impact project success and agents' well-being. According to their treatment of efforts over time, we can divide these related papers into two groups.

As in our paper, other models consider agents' efforts as *strategic substitutes* over time, leading to a reciprocal relation between current and future efforts. Bonatti and Hörner (2011) show how agents tend to procrastinate, that is, they spend inefficiently little effort too late, and how deadlines can be used to mitigate this problem. In a related setup from the literature on experimentation with a two-staged project, Moroni (2022) demonstrates that the principal can prevent agents from procrastinating by designing contracts that include asymmetries in terms of

3. The literature emphasizes several behavioral economic explanations for the ineffectiveness of incentives, such as the fact that higher incentives erode the norm of reciprocity (Fehr et al., 1997), crowd out people's intrinsic motivation (Frey & Oberholzer-Gee, 1997), and change the perception of contracts (Gneezy & Rustichini, 2000a, 2000b).

4. Winter (2009) coined the term "incentive reversal" to refer to a situation where higher rewards incentivize agents to exert less effort. This immediately implies a success reversal in his model since the probability of success increases in agents' efforts.

compensation and assignment of agents.[5] Weinschenk (2016) also analyzes procrastination in teams, identifying discriminatory contracts as a remedy for procrastination, while deadlines are shown to be never beneficial. Weinschenk (2021) compares time-consistent and time-inconsistent agents with the result that (naive) time-inconsistent agents are better off when it comes to project completion. Mason and Välimäki (2015) extended by Altan (2019) derive the optimal wage profile in a dynamic moral-hazard model of project completion with a single agent. While the principal optimally implements a wage scheme that declines over time under full commitment, the highest expected payoff in a weakly renegotiation-proof equilibrium is given by a constant wage over time.

Other dynamic models of a team working together to complete a project consider efforts as *strategic complements* over time by adding up efforts of all agents over all periods, which is interpreted as project progress. Georgiadis (2015) shows that, in this setting, the optimal team size is increasing in the length of the project, whereas the principal gradually reduces the team size when progress is made. In a similar (but deterministic) model, Georgiadis et al. (2014) derive the project size defined by a principal with limited commitment power: the principal extends the project size as it progresses and delegates decision rights to the agents if her commitment power is rather low.

Outline. The remainder of this paper is organized as follows. The next section introduces our model. Subsequently, we examine agents' equilibrium efforts before analyzing the effect of higher incentives in the static version of our model in Section 1.4. Our main analysis regarding the dynamic version of the model is conducted in Section 1.5. The final section concludes by discussing the implications of our findings. All proofs are provided in Appendix 1.A.1.

5. Bergemann and Hege (2005), Green and Taylor (2016), Halac et al. (2016), and Hörner and Samuelson (2013) are some examples of other articles from the literature on experimentation that share some features with our model.

1.2 Model

Consider a set of risk-neutral rational agents $\mathcal{N} := \{1, \dots, n\}$ working on a joint project. Agents choose their individual efforts simultaneously in every period $t = 1, \dots, T$. The probability that agents succeed in period t is $p(e_{t,1}, \dots, e_{t,n})$, where $e_{t,i} \in \mathbb{R}_+$ denotes the effort exerted by agent $i \in \mathcal{N}$ in period t.

Assumption 1 (Success Function)**.**
The success function $p : \mathbb{R}^n_+ \to [0, 1)$ is thrice continuously differentiable, symmetric, strictly increasing, weakly concave, and satisfies $p(0, \dots, 0) = 0$.[6]

Investing effort is costly for agent i. These effort costs are captured by a cost function $c(e_{t,i})$.

Assumption 2 (Cost Function)**.**
The cost function $c : \mathbb{R}_+ \to \mathbb{R}_+$ is thrice continuously differentiable, strictly increasing, strictly convex, and satisfies $c(0) = c'(0) = 0$.

If the agents succeed in t, each agent receives a reward $z_t \in \mathbb{R}_{++}$ and the game ends. The rewards from period t onwards define a reward profile $\boldsymbol{z}_t = \left(z_t, \dots, z_T\right)$. The game also ends if agents do not succeed until period T, where $0 < T < \infty$. We use the subgame perfect Nash equilibrium as a solution concept.

Some observations are immediate from the above model description. First, we can interpret the reward z_t as financial incentive in t since it incentivizes agents to invest effort in that period. Second, the model contains two implicit symmetry assumptions: agents are symmetric, and the success function is the same in every period.[7] Third, the model captures static ($T = 1$) and dynamic ($T > 1$) interaction, as well as single-agent ($n = 1$) and team ($n > 1$) problems.

6. Note that these assumptions allow agents' efforts to be either substitutes or complements. While our analysis does not hinge on the assumption that $p(0, \dots, 0) = 0$, it allows us to avoid tedious case distinctions.

7. In particular, to make the static and dynamic version of the model comparable, we do not examine spillover effects between different periods. While spillover effects are logically absent in the static model, they could be included in the dynamic model, which would, however, not provide any additional insights.

1.3 Equilibrium Efforts

Given the reward profile $\boldsymbol{z}_t$ and the efforts exerted by the other agents $j \in \mathcal{N} \setminus i$, the problem of agent i in period t is to maximize his expected payoff in that period over his effort choice $e_{t,i}$. Using Bellman's (1957) Principle of Optimality, agent i's decision problem reads:

$$\begin{aligned} \max_{e_{t,i} \in \mathbb{R}_+} \; v_{t,i}(e_{t,1}, \dots, e_{t,n}) \; &= \; p(e_{t,1}, \dots, e_{t,n})\, z_t \\ &\quad + \big(1 - p(e_{t,1}, \dots, e_{t,n})\big)\, \delta v^*_{t+1,i} - c(e_{t,i}), \end{aligned} \tag{1.1}$$

where $\delta \in (0, 1]$ denotes the discount factor and $v^*_{t+1,i}$ the optimal value of agent i's expected payoff in period $t+1$, i.e., his continuation payoff. Lemma 1 now examines equilibrium efforts, i.e., the efforts that arise if all agents choose their effort provision as a solution to Problem (1.1).

Lemma 1 (Equilibrium Efforts).

(i) *There exists a unique subgame perfect Nash equilibrium, which is symmetric in the sense that* $e^*_{t,i} = e^*_{t,j} = e^*_t$ *and* $v^*_{t,i} = v^*_{t,j} = v^*_t$ *for all* $i, j \in \mathcal{N}$ *and* $t \in \{1, \dots, T\}$.

(ii) *Agent i's equilibrium effort in period t is*

$$e^*_t \; = \; e(z_t, v^*_{t+1}) \; := \; \begin{cases} 0 & \text{if } z_t \; \le \; \delta v^*_{t+1} \\ e^{\text{FOC}}(z_t, v^*_{t+1}) & \text{if } z_t \; > \; \delta v^*_{t+1}, \end{cases}$$

where the map $e^{\text{FOC}}(z_t, v^*_{t+1})$ *is uniquely determined by*

$$\frac{\partial p(e_t, \dots, e_t)}{\partial e_{t,i}} \big(z_t - \delta v^*_{t+1}\big) - c'(e_t) \; = \; 0. \tag{FOC$_t$}$$

Lemma 1 *(i)* states that the game has a unique subgame perfect Nash equilibrium, in which all agents choose the same effort level within a period and thus expect identical payoffs. Part *(ii)* is straightforward. If the net reward $z_t - \delta v^*_{t+1}$ (i.e., the difference between the reward for project success and the discounted continuation payoff) is non-positive, then agents are at least weakly better off when having no success in period t. They thus minimize the probability of succeeding by investing zero effort. For positive net rewards, however, each agent chooses an effort level that balances the marginal benefit of effort (i.e., a higher probability of receiving the reward z_t instead of δv^*_{t+1}) with the marginal effort costs.

To facilitate the exposition, we exploit the symmetry of equilibrium efforts and define $P(e_t) := p\left(e_t, \ldots, e_t\right)$.[8] The following lemma establishes all relevant properties of the equilibrium efforts that will be used throughout the paper.

Lemma 2 (Properties of Equilibrium Efforts).

For each period $t \in \{1, \ldots, T\}$ and agent $i \in \mathcal{N}$, the equilibrium effort satisfies[9]

$$\text{(i)} \qquad e(z_t, v^*_{t+1}) \begin{cases} = 0 & \text{if } z_t \leq \delta v^*_{t+1} \\ > 0 & \text{if } z_t > \delta v^*_{t+1}, \end{cases} \quad \textit{and} \quad \lim_{z_t \to \infty} e(z_t, v^*_{t+1}) = \infty,$$

$$\text{(ii)} \qquad \frac{\partial e(z_t, v^*_{t+1})}{\partial z_t} = -\frac{1}{\delta} \frac{\partial e(z_t, v^*_{t+1})}{\partial v^*_{t+1}} \begin{cases} = 0 & \text{if } z_t \leq \delta v^*_{t+1} \\ > 0 & \text{if } z_t > \delta v^*_{t+1}. \end{cases}$$

Lemma 2 adopts an intraperiod perspective, i.e., it explores how equilibrium efforts in period t are affected by rewards and continuation payoffs in that period. Part *(i)* states that equilibrium efforts are positive if and only if net rewards are positive. In this case only, there is a positive probability that the agents succeed with the project in period t. Equilibrium efforts converge to infinity and thus induce success for sure if rewards converge to infinity. Part *(ii)* establishes that higher rewards have no effect for non-positive net rewards, while they strictly increase equilibrium efforts for positive net rewards. Intuitively, stronger incentives motivate agents to invest more effort in order to raise the probability of success. We call this the *motivation effect* of incentives. A higher continuation payoff has the opposite effect. Since reaching the subsequent period becomes more attractive, agents reduce their efforts in period t in order to increase the probability that the game continues. As we will discover in the dynamic version of the model, this is part of the mechanism that undermines the effectiveness of incentives.

Figure 1.1 portrays equilibrium efforts in period t as a map $z_t \mapsto e(z_t, v^*_{t+1})$. In Appendix 1.A.2, we show that the curvature of this map is determined by the interplay of the derivatives of the success and cost function. In particular, this curvature will be relevant for the effect of higher rewards on the success rate. $P''' \leq 0$ together with $c''' \geq 0$ ensure a concave shape as depicted in Figure 1.1. The figure also shows the motivation effect resulting from an increase in rewards from z^A_t to z^B_t.

8. Observe that this definition immediately implies that $\partial p(e_t, \ldots, e_t)/\partial e_{t,i} = P'(e_t)/n$.

9. Note that for $z_t - \delta v^*_{t+1} \to 0$ the equilibrium effort has a kink, where the derivative for $z_t - \delta v^*_{t+1} \to 0^-$ is zero, but the derivative for $z_t - \delta v^*_{t+1} \to 0^+$ is positive. For technical reasons, we concentrate on derivative for $z_t - \delta v^*_{t+1} \to 0^-$ throughout the paper.

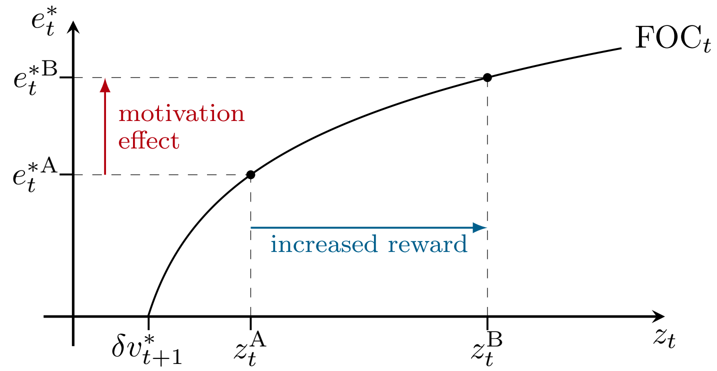


Figure 1.1. Equilibrium efforts and motivation effect in period t.

In light of Lemma 1 and 2, the following remark introduces notational simplifications that facilitate the exposition throughout the remainder of the paper.

Remark 1 (Notation).

(i) We henceforth write $e(\boldsymbol{z}_t)$ instead of $e(z_t, v_{t+1}^*)$ because, formally, v_{t+1}^* depends on $\boldsymbol{z}_{t+1}$ only.

(ii) Using the Bellman Equation (1.1) allows us to define agent i's equilibrium expected payoff in period t as $v_t^* = v_t(\boldsymbol{z}_t) := v_{t,i}(e(\boldsymbol{z}_t), \ldots, e(\boldsymbol{z}_t))$.

This puts us in the position to define the *success rate*, i.e., the equilibrium probability that the agents succeed with the project throughout the game:

$$S^* = S(\boldsymbol{z}_1) := 1 - \prod_{t=1}^{T} \Big(1 - P(e(\boldsymbol{z}_t))\Big).$$

1.4 Analysis of the Static Model

As a benchmark, we first study the static version of the model where $T = 1$.

Success Rate. The success rate in the static model is $S(z_1) = P(e(z_1))$. Its derivative reads

$$\frac{\partial S(z_1)}{\partial z_1} = P'(e_1(z_1)) \frac{\partial e_1(z_1)}{\partial z_1}. \tag{1.2}$$

As agents only work on the project for one period, their continuation payoff is zero, $v_2^* = 0$. Hence, Lemma 2 together with (1.2) implies that $\partial S(z_1)/\partial z_1 > 0$. Financial incentives thus work in terms of project success: higher rewards motivate agents to invest more effort, which makes success more likely.

Expected Payoffs. From (1.1) and the Envelope Theorem, it follows that

$$\frac{\partial v_1(z_1)}{\partial z_1} = P(e(z_1)) + z_1 \frac{n-1}{n} P'(e(z_1)) \frac{\partial e(z_1)}{\partial z_1}. \tag{1.3}$$

In view of Lemma 2, expected payoffs thus satisfy $\partial v_1(z_1)/\partial z_1 > 0$. The intuition is as follows. Consider the case with a single agent first. Since the agent may keep his effort and thereby his effort costs constant, a higher reward is necessarily beneficial (cf. first term on the r.h.s. of (1.3)). If there is a team of agents, each agent additionally benefits from higher efforts exerted by his teammates (cf. second term on the r.h.s. of (1.3)). In the static model, financial incentives are therefore effective also in terms of improving agents' well-being since agents are better off with higher-powered incentives.

1.5 Analysis of the Dynamic Model

The dynamic model where $T > 1$ is considered next. We first examine the mechanism that undermines the effectiveness of incentives and explore how it reverses the impact of incentives on the success rate and the expected payoffs. Subsequently, we provide numerical examples illustrating these reversals.

1.5.1 General Insights

Our analysis proceeds by investigating how higher rewards affect expected payoffs in the respective period and how these expected payoffs, in turn, affect equilibrium efforts in other periods.

Expected Payoffs in the Current Period. From (1.1) and the Envelope Theorem, it follows that

$$\frac{\partial v_t(\boldsymbol{z}_t)}{\partial z_t} = P(e(\boldsymbol{z}_t)) + \left(z_t - \delta v^*_{t+1}\right) \frac{n-1}{n} P'(e(\boldsymbol{z}_t)) \frac{\partial e(\boldsymbol{z}_t)}{\partial z_t}. \tag{1.4}$$

In light of Lemma 2, (1.4) reveals that higher rewards have the following effect:

$$\frac{\partial v_t(\boldsymbol{z}_t)}{\partial z_t} \begin{cases} = 0 & \text{if } z_t \leq \delta v^*_{t+1} \\ > 0 & \text{if } z_t > \delta v^*_{t+1}. \end{cases} \tag{1.5}$$

Hence, higher rewards in period t raise the expected payoff in that period, except for the case where net rewards are non-positive. In this case, agents invest zero effort and thus succeed with probability zero, such that marginally increasing their rewards has no effect on expected payoffs.

Effort Provision in Other Periods. While an increase in the reward z_t does not affect the effort provision in periods later than t, there can be an effect on earlier periods. Intuitively, once a period $\tau > t$ is reached, the reward in t can no longer be obtained and is thus irrelevant for the agents' effort choice. For a period $\tau < t$, on the contrary, the reward in t is still obtainable and might consequently affect equilibrium efforts in τ via the continuation payoff. Indeed, as Proposition 1 states, higher rewards cause an interesting intertemporal effect.

Proposition 1 (Discouragement Effect)**.**
Consider the dynamic model where $T \geq t > 1$. If $z_{t-1} > \delta v_t^*$ and $z_t > \delta v_{t+1}^*$, then a higher reward z_t reduces equilibrium efforts in period $t-1$,

$$\frac{\partial e(\boldsymbol{z}_{t-1})}{\partial z_t} \begin{cases} = 0 & \text{if } z_{t-1} \leq \delta v_t^* \text{ or } z_t \leq \delta v_{t+1}^* \\ < 0 & \text{if } z_{t-1} > \delta v_t^* \text{ and } z_t > \delta v_{t+1}^*. \end{cases} \tag{1.6}$$

It is not surprising that higher rewards may leave equilibrium efforts in the preceding period unaffected. We can read off the first line in (1.6) that this may happen for two reasons. First, as net rewards in the preceding period are non-positive for $z_{t-1} \leq \delta v_t^*$, agents already invest zero effort in that period, rendering a further reduction in efforts impossible. Second, for $z_t \leq \delta v_{t+1}^*$, agents exert zero effort in period t and fail for sure in that period, such that higher rewards have no effect on the continuation payoffs.

More interestingly, as the second line in (1.6) reveals, higher rewards can also reduce equilibrium efforts in the preceding period. This *discouragement effect* counteracts the beneficial impact of higher incentives in period t. It occurs since higher rewards in period t increase agents' continuation payoffs if $z_t > \delta v_{t+1}^*$, which, in turn, reduces net rewards in the preceding period. Accordingly, investing effort in the preceding period becomes less attractive and agents reduce their effort provision whenever this is possible, i.e., whenever equilibrium efforts are positive. Figure 1.2 illustrates the discouragement effect generated by an inframarginal increase in rewards from z_t^A to z_t^B.

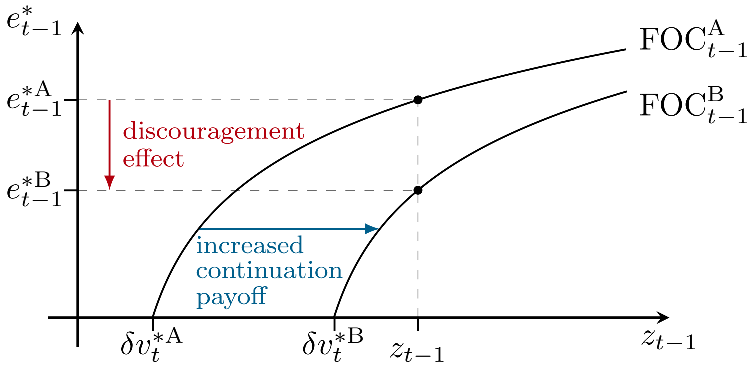


Figure 1.2. Equilibrium efforts and discouragement effect in period $t-1$.

Examining (1.6) allows us to state the following corollary.

Corollary 1 (Relevance of the Discouragement Effect).
Consider the dynamic model where $T > \tau \geq 1$. If the reward profile $\mathbf{z}_\tau$ is either non-increasing over time or the increase is sufficiently small, then higher rewards in any period $t \in \{\tau+1, \dots, T\}$ cause a discouragement effect.

Many economic scenarios are indeed characterized by non-increasing rewards. While the economic environment naturally defines such reward structures in situations ranging from infrastructure projects (cf. Ekici and Retharekar, 2013) to cryptocurrency mining (cf. Arenas et al., 2020), it can also be optimal for a principal to design non-increasing rewards in various forms of principal-agent relationships (see, e.g., Altan, 2019; Mason and Välimäki, 2015; Mayer, 2022). The discouragement effect is thus a relevant force in many economic settings.

Comparing Proposition 1 and Lemma 2 reveals that the discouragement effect only emerges in conjunction with the motivation effect. Both effects combined can be interpreted as an intertemporal substitution effect of efforts since higher rewards in period t increase equilibrium efforts in that period but decrease equilibrium efforts in the preceding period. Thus, there is a shift of efforts from period $t-1$ to period t. As we will explore next, this interplay between the motivation and discouragement effect has surprising implications for the project's success rate and the agents' expected payoffs.

Success Rate. For simplicity, the effect of higher rewards on the success rate is demonstrated for two periods in the main text.[10] If $T = 2$, then the success rate

10. For the analysis of multiple periods, i.e., the case where $T > 2$, we refer to Appendix 1.A.3.

can be rewritten to

$$S(\boldsymbol{z}_1) \;=\; P(e(\boldsymbol{z}_1)) + \big(1 - P(e(\boldsymbol{z}_1))\big) P(e(\boldsymbol{z}_2)). \tag{1.7}$$

Observe that the success rate is increasing in the agents' efforts. The question is thus how higher rewards change equilibrium efforts and how these, in turn, affect the success rate. We first examine the effect of higher rewards in period 1. Formally, differentiating (1.7) and using Lemma 2 leads to

$$\frac{\partial S(\boldsymbol{z}_1)}{\partial z_1} \;=\; \big(1 - P(e(\boldsymbol{z}_2))\big) P'(e_1(\boldsymbol{z}_1)) \frac{\partial e_1(\boldsymbol{z}_1)}{\partial z_1} \begin{cases} = 0 & \text{if } z_1 \;\leq\; \delta v_2^* \\ > 0 & \text{if } z_1 \;>\; \delta v_2^*. \end{cases}$$

Due to the motivation effect, higher rewards in the first period at least weakly raise equilibrium efforts in that period, while they leave second-period efforts unaffected. Therefore, higher first-period rewards at least weakly increase the success rate. By contrast, an increase in the second-period rewards causes two countervailing effects:

$$\begin{aligned} \frac{\partial S(\boldsymbol{z}_1)}{\partial z_2} =& \overbrace{\big(1 - P(e(\boldsymbol{z}_1))\big) P'(e(\boldsymbol{z}_2)) \frac{\partial e(\boldsymbol{z}_2)}{\partial z_2}}^{\geq 0} \\ &+ \underbrace{\big(1 - P(e(\boldsymbol{z}_2))\big) P'(e(\boldsymbol{z}_1))) \frac{\partial e(\boldsymbol{z}_1)}{\partial z_2}}_{\leq 0}. \end{aligned} \tag{1.8}$$

On the one hand, it motivates agents to increase their equilibrium efforts in the second period, which increases the success rate (cf. first term on the r.h.s. of (1.8)). On the other, it discourages agents' equilibrium efforts in the first period, which decreases the success rate (cf. second term on the r.h.s. of (1.8)). Both effects are strict for positive net rewards. If the discouragement effect dominates the motivation effect, then the success rate is decreasing in the second-period reward. That is, higher rewards for project success make success less likely. We refer to this scenario as a success reversal.

Definition 1 (Success Reversal).
Let the reward profile $\boldsymbol{z}_1 \in \mathbb{R}^{\mathrm{T}}_{++}$ be given and $T > 1$. A success reversal occurs if $\partial S(\boldsymbol{z}_1)/\partial z_t < 0$ holds for some $t \in \{2, \ldots, T\}$.

To explore why success reversals occur, we rearrange (1.8) to

$$\frac{\partial S(\boldsymbol{z}_1)}{\partial z_2} < 0 \iff \underbrace{\frac{\Big(1-P(e(\boldsymbol{z}_2))\Big)P'(e(\boldsymbol{z}_1))}{\Big(1-P(e(\boldsymbol{z}_1))\Big)P'(e(\boldsymbol{z}_2))}}_{\text{MRS}} \cdot \underbrace{\frac{-\frac{\partial e(\boldsymbol{z}_1)}{\partial z_2}}{\frac{\partial e(\boldsymbol{z}_2)}{\partial z_2}}}_{\text{MRE}} > 1. \tag{1.9}$$

The first fraction in (1.9) is the *marginal rate of intertemporal substitution* (MRS) of first-period efforts for second-period efforts. It measures how efforts in period one can be substituted with efforts in period two to keep the success rate constant. The second fraction in (1.9) is referred to as the *marginal rate of incentivized efforts* (MRE), which captures the ratio of changes in equilibrium efforts in the first and second period caused by a marginal increase in the second-period reward. Theorem 1 states sufficient conditions for a success reversal to arise.

Theorem 1 (Success Reversal)**.**
Consider the dynamic model where $T = 2$ and let z_2 be given.

(i) If P'' is sufficiently close to zero on a sufficiently large interval $[\underline{e}, \overline{e}]$ and $c''(\overline{e})$ is sufficiently small, then a success reversal occurs for z_1 in the left local neighborhood of the reward that implements $\overline{e}$.
(ii) If P'' is strictly negative but finite and $P'(0)/c''(0)$ is sufficiently large, then a success reversal occurs for z_1 in the right local neighborhood of δv_2^*.

As Condition (1.9) can be fulfilled due to large values of the MRS or the MRE, two different mechanisms for success reversals need to be distinguished. First, Theorem 1 *(i)* demonstrates that if the curvature of P is rather low, then a success reversal arises since the MRS obtains large values. This is the case for reward profiles that implement rather high success probabilities in the first period. Intuitively, a success reversal occurs because the intertemporal substitution of efforts shifts efforts from a "productive" period for project success to an "unproductive" one, i.e., from a period that is reached with certainty to one that is unlikely to be reached. We therefore refer to this situation as a success reversal due to *substitution of efforts*. Second, Theorem 1 *(ii)* reveals that a success reversal also arises due to large values of the MRE. Since the properties specified for the success and cost function ensure that marginal efforts are high when net rewards are low, this form of success reversal occurs for reward profiles that yield sufficiently low net rewards in the first period. Equilibrium efforts then react rather sensitively in the first period but comparably insensitively in the second period towards changes in the second-period reward. A strong discouragement effect then outweighs a weak motivation effect. Accordingly, we call this a success reversal due to *sensitivity of efforts*.

Corollary 2 demonstrates that a success reversal due to substitution of efforts can occur even with a standard CES function success function and any cost function whose second derivative is bounded.

Corollary 2 (CES Success Function).
Consider the dynamic model where $T = 2$. Suppose that the success probability is captured by a CES function

$$p\left(e_{t,1},\ldots,e_{t,n}\right) = \beta\left[\sum_{i=1}^{n} e_{t,i}^{\rho}\right]^{\alpha/\rho},$$

with $0 < \alpha \leq 1$, $\beta > 0$, $\rho \leq 1$, and domain $e_{t,i} \in [0, \bar{e})$, where $\bar{e} := n^{-1/\rho}/\beta^{1/\alpha}$. Let z_2 be given such that equilibrium efforts in period 2 are interior, i.e., $e_2^* \in (0, \bar{e})$. If $\lim_{e_{t,i}\to\bar{e}} c''(e_{t,i})$ is finite, then a success reversal occurs for z_1 in the left local neighborhood of the reward that implements $\bar{e}$.

It is noteworthy that the afore-described mechanisms do not hinge on the number of agents, i.e., success reversals can arise in single-agent and team problems. Regarding the number of periods, we show in Appendix 1.A.3 that whenever there are success reversals in the two-period model, we can construct reward profiles in the model with $T > 2$ that also cause success reversals.

Expected Payoffs in Other Periods. While higher rewards in period t do not affect expected payoffs in later periods $\tau > t$ (by the same arguments stated for the effort provision), they may impact expected payoffs in earlier periods $\tau < t$.

The case of a *single* agent is analyzed first. In this case, higher expected payoffs in period t cause higher expected payoffs in all previous periods. To see this, consider period $t-1$ first. Since there is always *(i)* a strictly positive probability that the game continues and *(ii)* the possibility for the agent to maintain his effort choice and thus his effort costs, a higher continuation payoff necessarily increases the equilibrium expected payoff in $t-1$. Repeating this argument shows that higher expected payoffs in period t cause higher expected payoffs in all preceding periods $\tau \in \{1,\ldots,t-1\}$. Together with Equation (1.5), this implies that

$$\frac{\partial v_\tau(\boldsymbol{z}_\tau)}{\partial z_t} \begin{cases} = 0 & \text{if } z_t \leq \delta v_{t+1}^* \\ > 0 & \text{if } z_t > \delta v_{t+1}^*, \end{cases}$$

for all $\tau \in \{1, \ldots, t\}$. The positive impact of higher rewards holds, in particular, for the expected payoff in the first period, which represents the agent's expected payoff from the entire game (i.e., from working on the project). Thus, in the case of a single agent, higher rewards are at least weakly beneficial for the agent. Interestingly, as we will see next, this is not necessarily true with *multiple* agents.

Proposition 2 (Expected Payoff in the Preceding Period).
Consider the dynamic model where $T \geq t > 1$ and $n > 1$. If $z_{t-1} > \delta v_t^*$ and $z_t > \delta v_{t+1}^*$, then a higher reward z_t has an ambiguous effect on the equilibrium expected payoff in $t-1$,

$$\frac{\partial v_{t-1}(\boldsymbol{z}_{t-1})}{\partial z_t} \propto \overbrace{1 - P(e(\boldsymbol{z}_{t-1}))}^{>0} \\ + \underbrace{\left(z_{t-1} - \delta v_t^*\right) \frac{n-1}{n} P'(e(\boldsymbol{z}_{t-1})) \frac{\partial e(z_{t-1}, v(\boldsymbol{z}_t))}{\partial v_t^*}}_{<0}. \tag{1.10}$$

The intuition behind Proposition 2 is as follows. Whenever equilibrium efforts are positive in periods $t-1$ and t, higher rewards in t engender two countervailing effects. On the one hand, if the team fails in period $t-1$, it benefits from increased rewards in the next period, i.e., higher rewards have a *beneficial direct* effect on expected payoffs. This effect is captured by the first line on the r.h.s. of (1.10). On the other hand, agents benefit from efforts invested by their teammates, i.e., there is a positive team externality. All teammates, however, lower their equilibrium efforts due to the discouragement effect if the reward in the subsequent period increases. Hence, higher rewards also cause a *detrimental indirect* effect. This is captured by the second line on the r.h.s. of (1.10). We next define a payoff reversal.

Definition 2 (Payoff Reversal).
Let the reward profile $\boldsymbol{z}_1 \in \mathbb{R}^{\mathrm{T}}_{++}$ be given and $T > 1$. A payoff reversal occurs if $\partial v_{t-1}(\boldsymbol{z}_{t-1})/\partial z_t < 0$ holds for some $t \in \{2, \ldots, T\}$.

Theorem 2 now states sufficient conditions for payoff reversals to arise.

Theorem 2 (Payoff Reversal).
Consider the dynamic model where $T \geq t > 1$ and $n > 1$. Let $\boldsymbol{z}_t$ be given such that $z_t > \delta v_{t+1}^*$. If P'' is sufficiently close to zero on a sufficiently large interval $[\underline{e}, \bar{e}]$ and $c''(\bar{e})$ is sufficiently small, then a payoff reversal occurs for z_{t-1} in the left local neighborhood of the reward that implements $\bar{e}$.

Theorem 2 reveals that if the curvature of P is rather low, then a payoff reversal arises for those reward profiles that implement relatively high success probabilities in period $t-1$. Intuitively, since the likelihood of failure is rather low in the preceding period, it is unlikely that agents actually reach period t and benefit from higher rewards. In this case, the adverse impact of lower equilibrium efforts dominates such that agents suffer from higher rewards.

An analog version of Corollary 2 applies to payoff reversals: with a CES success function and any cost function whose second derivative is bounded, payoff reversals arise for reward profiles that implement equilibrium efforts in period $t-1$ sufficiently close to $\bar{e}$.

Comparison of Reward Profiles. Relating to the terminology of decision theory, we introduce the following definition.

Definition 3 (Dominant Reward Profile).
Compare two reward profiles, $\mathbf{z}_1^{\mathrm{A}} \in \mathbb{R}_{++}^{\mathrm{T}}$ and $\mathbf{z}_1^{\mathrm{B}} \in \mathbb{R}_{++}^{\mathrm{T}}$. If $z_t^{\mathrm{A}} < z_t^{\mathrm{B}}$ holds in all periods $t \in \{1,\dots,T\}$, then $\mathbf{z}_1^{\mathrm{B}}$ dominates $\mathbf{z}_1^{\mathrm{A}}$.

When comparing two reward profiles in the static model, it is straightforward that the dominant reward profile induces a higher success rate and generates higher expected payoffs. In the dynamic model, however, results are more surprising. The existence of success and payoff reversals (cf. Theorem 1 and 2) implies the following corollary.

Corollary 3 (Dominant Versus Dominated Reward Profiles).
Consider the dynamic model $T > 1$ and compare two reward profiles, $\mathbf{z}_1^{\mathrm{A}}$ and $\mathbf{z}_1^{\mathrm{B}}$, where $\mathbf{z}_1^{\mathrm{B}}$ dominates $\mathbf{z}_1^{\mathrm{A}}$. Then the success rate need not be higher with $\mathbf{z}_1^{\mathrm{B}}$. Moreover, if $n > 1$, then the equilibrium expected payoffs resulting from $\mathbf{z}_1^{\mathrm{B}}$ need not be higher in any period $t \in \{1,\dots,T-1\}$.

Corollary 3 states that even though a reward profile provides higher rewards in every single period, it may cause a lower success rate and/or generate lower expected payoffs. Financial incentives may thus be counterproductive, i.e., unambiguously stronger incentives render success less likely and/or leave agents worse off. Generally speaking, stronger financial incentives become counterproductive in such situations since they induce a less favorable intertemporal distribution of equilibrium efforts, which overcompensates their beneficial impact.

1.5.2 Numerical Examples

This section shows the extend to which success and payoff reversals arise. Consider the simplest form of the dynamic model where $T = 2$. Let $n = 4$, $\delta = 0.95$, effort costs be characterized by a power function of degree three[11] with scaling factor $\kappa > 0$, and the success function be CES with $\alpha = 1$, $\beta > 0$, and $\rho \leq 1$,

$$c\left(e_{t,i}\right) = \kappa\left(e_{t,i}\right)^3, \qquad \text{and} \qquad p\left(e_{t,1},\ldots,e_{t,n}\right) = \beta\left[\sum_{i=1}^{n} e_{t,i}^{\rho}\right]^{1/\rho}. \tag{1.11}$$

Define $\bar{z}$ as the reward that implements $e_2^* = n^{-1/\rho}/\beta$, which constitutes the upper bound for the rewards in both periods.[12]

Success Reversals. Calculating and inserting equilibrium efforts in the success function shows that the success rate only depends on relative rewards $z_t/\bar{z}$. Figure 1.3 depicts iso success rate lines.

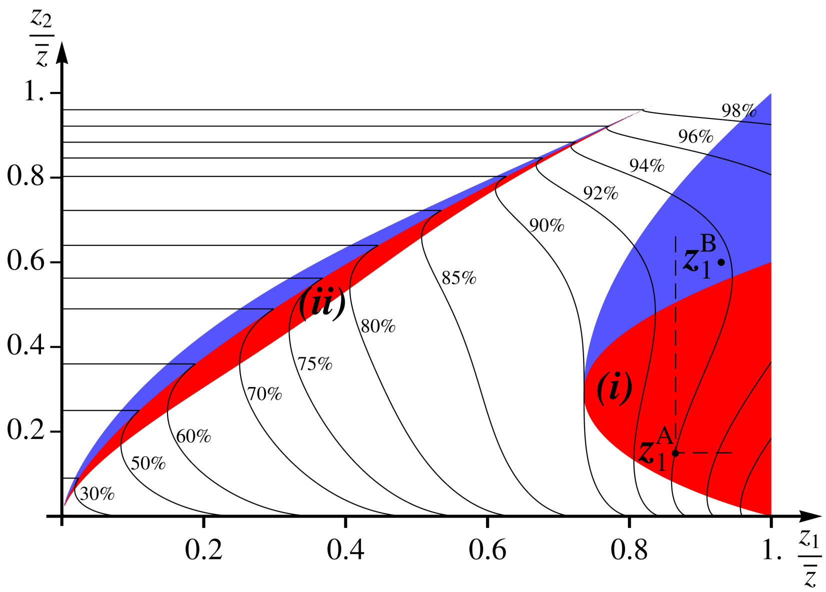


Figure 1.3. Iso success rate lines and success reversals.
($n = 4$, $\alpha = 1$, and $\delta = 0.95$)

11. Payoff reversals and success reversals due to substitution of efforts also arise for quadratic effort costs. Degree three, however, allows us to study all forms of reversals in a single example.

12. $\bar{z}$ is the reward that implements success for sure in period 2 and captures the model parameters κ, ρ, and β. Note that for all $t \in \{1,2\}$ focusing on rewards $z_t < \bar{z}$ ensures that *(i)* equilibrium efforts e_t^* are unique and *(ii)* $p\left(e_t^*,\ldots,e_t^*\right)$ satisfies Assumption 1.

Success reversals occur for those combinations of rewards for which the iso success rate lines have a positive slope (cf. *Red Area (i)* and *(ii)* in Figure 1.3). Considering both areas together demonstrates that success reversals may arise for any positive reward in the first period, i.e., there are always some second-period rewards for which the success rate decreases. Success reversals due to substitution of efforts arise for reward profiles in *Red Area (i)*, whereas success reversals due to sensitivity of efforts emerge in *Red Area (ii)*. To illustrate Corollary 3, compare $\boldsymbol{z}_1^A = \left(0.86\,\bar{z},\ 0.15\,\bar{z}\right)$ and $\boldsymbol{z}_1^B = \left(0.93\,\bar{z},\ 0.6\,\bar{z}\right)$. While $\boldsymbol{z}_1^B$ clearly dominates $\boldsymbol{z}_1^A$, Figure 1.3 shows that $\boldsymbol{z}_1^A$ incentivizes a higher success rate. By the same logic, for any given reward profile in the *red* and *blue* areas, there exists a dominated reward profile that induces a higher success rate. We can thus conclude that *red* and *blue* areas mark those reward profiles that may be counterproductive for success.

Payoff Reversals. Since expected payoffs do not solely depend on relative rewards, let $\kappa = \rho = 1$ and $\beta = 0.5$. Figure 1.4 portrays iso expected payoff lines for first-period expected payoffs.

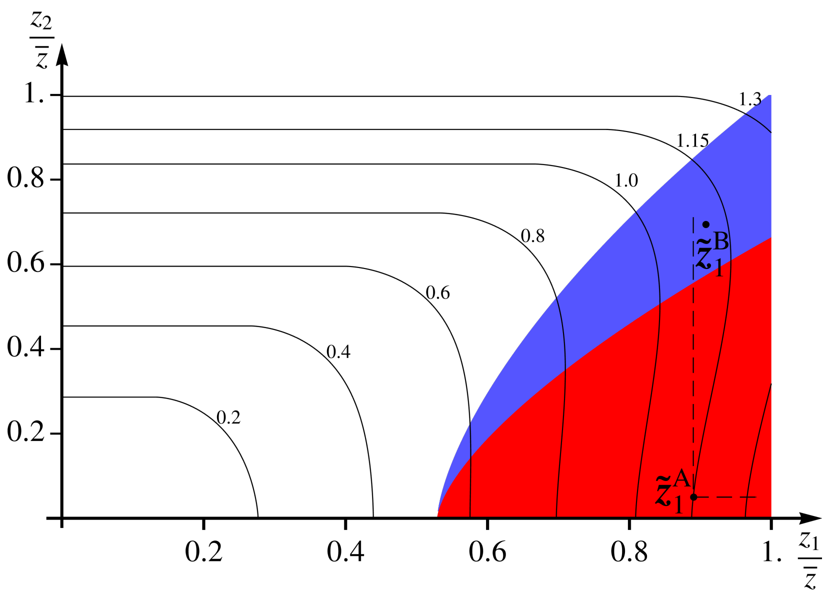


Figure 1.4. Iso payoff lines and payoff reversals.
($n = 4$, $\alpha = 1$, $\beta = 0.5$, $\delta = 0.95$, and $\kappa = \rho = 1$)

The *red* area marks reward profiles that cause a payoff reversal. Analogously to before, reward profiles that may be counterproductive, i.e., those for which a dominated reward profile with a higher expected payoff in period 1 exists, are represented by the *red* and *blue* area. For example, although $\tilde{\boldsymbol{z}}_1^B = \left(0.91\,\bar{z},\ 0.7\,\bar{z}\right)$ specifies higher rewards in both periods than $\tilde{\boldsymbol{z}}_1^A = \left(0.89\,\bar{z},\ 0.05\,\bar{z}\right)$, Figure 1.4 reveals that agents' expected payoffs from working on the project are higher with $\tilde{\boldsymbol{z}}_1^A$.

Comparison of Success and Payoff Reversals. Since both reversals originate from the discouragement effect, a natural question is whether a success reversal implies a payoff reversal or vice versa. Comparing the red areas in Figures 1.3 and 1.4 reveals that success and payoff reversals can occur together but also separately. This finding allows us to state the following proposition.

Proposition 3 (Comparison of Reversals)**.**
Success reversals are neither necessary nor sufficient for payoff reversals and vice versa.

Figure 1.5 illustrates Proposition 3 by exploring which combinations of first-period reward and discount factor can give rise to reversals, i.e., for which combinations the success rate and/or first-period expected payoffs are locally decreasing in the second-period reward.

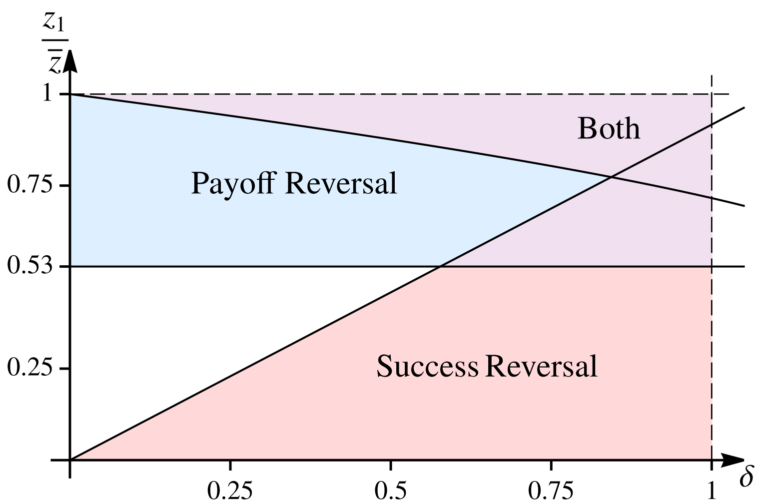


Figure 1.5. Existence of success and payoff reversals.
($n = 4$, $\alpha = 1$, $\beta = 0.5$, and $\kappa = \rho = 1$)

1.6 Conclusion

This paper explores the effectiveness of financial incentives. Table 1.2 summarizes our results by presenting the impact of an inframarginal increase in rewards on the project's success rate and the agents' expected payoffs in period $\tau \in \{1, \ldots, t\}$.

Table 1.2. Effect of an inframarginal increase Δz_t in the reward in period t.

	single agent	multiple agents
static model	$\Delta S^* > 0,\ \ \Delta v^*_\tau > 0$	$\Delta S^* > 0,\ \ \Delta v^*_\tau > 0$
dynamic model	$\Delta S^* \gtreqless 0,\ \ \Delta v^*_\tau \geq 0$	$\Delta S^* \gtreqless 0,\ \ \Delta v^*_\tau \gtreqless 0$

Since dynamic interaction is shown to undermine the effectiveness of incentives, stronger incentives may be *(i)* counterproductive for project success and *(ii)* detrimental to agents (i.e., lower their expected payoffs from working on the project). Although both phenomena result from the intertemporal discouragement effect of higher incentives, they can occur together but also separately.

Our paper provides a novel, non-behavioral explanation for the ineffectiveness of incentives. The results can be useful in shedding light on various real-world phenomena. First, in practice, many projects are either delayed or not completed at all (cf. Fabricius and Büttgen, 2015; Park, 2021; Van Genuchten, 1991). Long delays are typical in the construction of public infrastructure as well as in defense and high-tech projects.[13] Well-known examples include the Sydney Opera House and Boston's Third Harbor Tunnel Project, which were completed ten and eight years later, respectively, than initially planned. Our paper offers an explanation for such delays and failures since it shows that agents may strategically invest little effort during project phases, even if strong incentives are provided. Second, it has been empirically shown that only a fraction of employees work in jobs with financial incentives and that these incentives tend to be small relative to total income (see, e.g., Baktash et al., 2022; Bell and Van Reenen, 2014; Hong et al., 2019; Lemieux et al., 2009). Our paper helps to understand these findings as it reveals that, with dynamic interaction, incentives can be counterproductive due to their adverse intertemporal effect.

13. Park (2021), for instance, finds that among 113 major infrastructure projects in the US and the UK between 1999 and 2018 about 75% were delayed.

Appendix 1.A

Appendix 1.A.1 provides all proofs of our main results. Appendix 1.A.2 examines the curvature of equilibrium efforts, while Appendix 1.A.3 explores the effect of higher incentives on the success rate in the case of multiple periods $T > 2$.

1.A.1 Proofs of the Main Results

Proof of Lemma 1. The proof works in two steps. We first show existence and uniqueness $\hat{e}_{t,i}$, which denotes agent i's best response, i.e., his optimal effort choice in period t given the efforts of the other agents. Subsequently, we argue that efforts and expected payoffs are identical across agents in the subgame perfect Nash equilibrium and that they are either zero or determined by (FOC$_t$).

Step 1. Taking into account the non-negativity condition $e_{t,i} \geq 0$, the Lagrangian of the maximization problem (1.1) writes as

$$\begin{aligned} \mathscr{L}(e_{t,i}) &= v_{t,i}(e_{t,1}, \ldots, e_{t,n}) + \lambda_{t,i} e_{t,i} \\ &= p(e_{t,1}, \ldots, e_{t,n})\, z_t + \big(1 - p(e_{t,1}, \ldots, e_{t,n})\big)\, \delta v^*_{t+1,i} - c(e_{t,i}) + \lambda_{t,i} e_{t,i}. \end{aligned} \tag{1.A.1}$$

The Karush-Kuhn-Tucker conditions are necessary and, by concavity, also sufficient for a maximum:

$$\begin{aligned} \frac{\partial \mathscr{L}}{\partial e_{t,i}} &= \frac{\partial p(e_{t,1}, \ldots, e_{t,n})}{\partial e_{t,i}} \big(z_t - \delta v^*_{t+1,i}\big) - c'(e_{t,i}) + \lambda_{t,i} \;=\; 0, \\ & e_{t,i} \;\geq\; 0, \; \lambda_{t,i} \;\geq\; 0, \quad \text{and} \quad \lambda_{t,i} e_{t,i} \;=\; 0 \quad \text{for all } t, i. \end{aligned} \tag{1.A.2}$$

We have to distinguish between the case where $\lambda^*_{t,i} > 0$ and the case where $\lambda^*_{t,i} = 0$.

First, if $\lambda^*_{t,i} > 0$, then agent i's effort choice in period t is uniquely determined by $\hat{e}_{t,i} = 0$ due to the complementary slackness condition. This case is relevant if $z_t - \delta v^*_{t+1,i} < 0$, since otherwise (1.A.2) cannot be satisfied.

Second, if $\lambda^*_{t,i} = 0$, then the derivative in (1.A.2) simplifies to

$$\frac{\partial p(e_{t,1}, \ldots, e_{t,n})}{\partial e_{t,i}} \big(z_t - \delta v^*_{t+1,i}\big) - c'(e_{t,i}) \;=\; 0, \tag{1.A.3}$$

such that the effort $\hat{e}_{t,i}$ is determined by the solution to the first-order condition. This case is relevant if the net reward is non-negative, $z_t - \delta v^*_{t+1,i} \geq 0$. Now, two subcases need to be distinguished.

If $z_t - \delta v^*_{t+1,i} = 0$, then it follows from Assumption 2 that $\hat{e}_{t,i} = 0$ is the unique solution to (1.A.3).

If $z_t - \delta v^*_{t+1,i} > 0$, then existence and uniqueness of $\hat{e}_{t,i}$ can be seen as follows. Considering the limits of the l.h.s. of (1.A.3), we get

$$\lim_{e_{t,i}\to 0} \frac{\partial p(e_{t,1},\dots,e_{t,n})}{\partial e_{t,i}} \left(z_t - \delta v^*_{t+1,i}\right) - c'(e_{t,i}) \ > \ 0,$$
$$\lim_{e_{t,i}\to \infty} \frac{\partial p(e_{t,1},\dots,e_{t,n})}{\partial e_{t,i}} \left(z_t - \delta v^*_{t+1,i}\right) - c'(e_{t,i}) \ < \ 0,$$

where the last limit obtains since $\partial p(e_{t,1},\dots,e_{t,n})/\partial e_{t,i} \to 0$ because p is strictly increasing, weakly concave and bounded by Assumption 1. Hence, the Intermediate Value Theorem implies the existence of a solution $\hat{e}_{t,i} > 0$. Differentiating the l.h.s. of (1.A.3) yields

$$\frac{\partial^2 p(e_{t,1},\dots,e_{t,n})}{\partial e^2_{t,i}} \left(z_t - \delta v^*_{t+1,i}\right) - c''(e_{t,i}),$$

which is strictly negative by Assumption 1 and 2. Hence, $\hat{e}_{t,i}$ is unique.

Step 2. We derive the subgame perfect Nash equilibrium by backward induction. Consider **period T** first. As the game ends after T at the latest, the continuation payoff in this period is zero for all agents, $v^*_{T+1,i} = v^*_{T+1,j} = v^*_{T+1} = 0$ for all $i,j \in \mathcal{N}$. Since it thus holds that $z_T - \delta v^*_{T+1,i} > 0$ for all $i \in \mathcal{N}$, every agent chooses his best response $\hat{e}_{T,i}$ according to (1.A.3). Symmetry of p now implies that efforts must be identical across agents in the Nash equilibrium, $e^*_{T,i} = e^*_{T,j} = e^*_T$ for all $i,j \in \mathcal{N}$, since otherwise (1.A.3) cannot be satisfied for all $i \in \mathcal{N}$. The equilibrium effort e^*_T is thus determined by the unique solution to

$$\frac{\partial p(e_T,\dots,e_T)}{\partial e_{T,i}} \left(z_T - \delta v^*_{T+1}\right) - c'(e_T) \ = \ 0,$$

which is exactly (FOC$_t$) in the case where $t = T$. Identical equilibrium efforts imply identical effort costs across agents, such that we can read off (1.1) that the equilibrium expected payoff $v^*_{T,i}$ is also unique and identical across agents, $v^*_{T,i} = v^*_{T,j} = v^*_T$ for all $i,j \in \mathcal{N}$.

Consider **period $T-1$** next. Now, two cases can occur. If $z_{T-1} - \delta v^*_T < 0$, then $z_{T-1} - \delta v^*_{T,i} < 0$ holds for all $i \in \mathcal{N}$, such that choosing zero effort is the best response for every agent. Hence, in equilibrium, it holds that $e^*_{T-1,i} = e^*_{T-1,j} = e^*_{T-1} = 0$ for all $i,j \in \mathcal{N}$. By contrast, if $z_{T-1} - \delta v^*_T \geq 0$, then $z_{T-1} - \delta v^*_{T,i} \geq 0$ holds for all

$i \in \mathcal{N}$, such that every agent chooses his best response $\hat{e}_{T-1,i}$ according to (1.A.3). Symmetry of p again implies that $e^*_{T-1,i} = e^*_{T-1,j} = e^*_{T-1}$ holds for all $i,j \in \mathcal{N}$, where e^*_{T-1} is determined by the unique solution to

$$\frac{\partial p(e_{T-1},\ldots,e_{T-1})}{\partial e_{T-1,i}}\left(z_{T-1} - \delta v^*_T\right) - c'(e_{T-1}) \;=\; 0,$$

which is exactly (FOC$_t$) in the case where $t = T-1$. Identical equilibrium efforts together with (1.1) then imply that the equilibrium expected discounted payoff $v^*_{T-1,i}$ is also unique and identical across agents, $v^*_{T-1,i} = v^*_{T-1,j} = v^*_{T-1}$ for all $i,j \in \mathcal{N}$.

Repeating the previous arguments until period 1 completes the proof of Lemma 1.[14] □

Proof of Lemma 2.

Part (i). The fact that $e(z_t, v^*_{t+1})$ satisfies

$$e(z_t, v^*_{t+1}) \begin{cases} = 0 & \text{if } z_t \;\le\; \delta v^*_{t+1} \\ > 0 & \text{if } z_t \;>\; \delta v^*_{t+1}, \end{cases}$$

follows directly from the proof of Lemma 1. To see that equilibrium efforts converge to infinity for $z_t \to \infty$, note first that $z_t \to \infty$ implies that $z_t > \delta v^*_{t+1}$. Hence, equilibrium efforts are determined by the solution to (FOC$_t$), which can be rewritten as

$$\left(z_t - \delta v^*_{t+1}\right) \;=\; \frac{nc'(e_t)}{P'(e_t)}. \tag{1.A.4}$$

For $z_t \to \infty$, the l.h.s. of (1.A.4) converges to infinity. For (1.A.4) to hold, the r.h.s. must also converge to infinity. Since r.h.s. is continuous and strictly increasing in e_t by Assumption 1 and 2, this requires that $e^*_t \to \infty$. Hence, we have that $e(z_t, v^*_{t+1}) \to \infty$ for $z_t \to \infty$.

Part (ii). If $z_t \le \delta v^*_{t+1}$, then $e(z_t, v^*_{t+1}) = 0$, which directly implies that

$$\frac{\partial e(z_t, v^*_{t+1})}{\partial z_t} \;=\; -\frac{1}{\delta}\frac{\partial e(z_t, v^*_{t+1})}{\partial v^*_{t+1}} \;=\; 0.$$

14. Note that we used the fact that $e^*_t = 0$ if $z_t - \delta v^*_{t+1,i} = 0$ in the definition of $e(z_t, v^*_{t+1})$.

By contrast, if $z_t > \delta v^*_{t+1}$, then differentiation of (FOC_t) yields

$$\frac{\partial e(z_t, v^*_{t+1})}{\partial z_t} = \frac{\overbrace{-P'(e(z_t, v^*_{t+1}))}^{<0}}{\underbrace{P''(e(z_t, v^*_{t+1}))\left(z_t - \delta v^*_{t+1}\right)}_{\leq 0} \underbrace{- nc''(e(z_t, v^*_{t+1}))}_{<0}} > 0, \tag{1.A.5}$$

and

$$\frac{\partial e(z_t, v^*_{t+1})}{\partial v^*_{t+1}} = \frac{\overbrace{\delta P'(e(z_t, v^*_{t+1}))}^{>0}}{\underbrace{P''(e(z_t, v^*_{t+1}))\left(z_t - \delta v^*_{t+1}\right)}_{\leq 0} \underbrace{- nc''(e(z_t, v^*_{t+1}))}_{<0}} < 0. \tag{1.A.6}$$

Comparing (1.A.5) and (1.A.6), reveals that

$$\frac{\partial e(z_t, v^*_{t+1})}{\partial z_t} = -\frac{1}{\delta}\frac{\partial e(z_t, v^*_{t+1})}{\partial v^*_{t+1}} > 0,$$

for $z_t > \delta v^*_{t+1}$. □

Proof of Proposition 1. By the chain rule, it holds for the derivative of the equilibrium efforts in period $t-1$ with respect to z_t that

$$\frac{\partial e(\boldsymbol{z}_{t-1})}{\partial z_t} = \frac{\partial e(z_{t-1}, v_t(\boldsymbol{z}_t))}{\partial v^*_t}\frac{\partial v_t(\boldsymbol{z}_t)}{\partial z_t}. \tag{1.A.7}$$

Since the first factor in (1.A.7) is strictly negative for $z_{t-1} > \delta v^*_t$ and zero otherwise by Lemma 2, while the second factor in (1.A.7) is strictly positive for $z_t > \delta v^*_{t+1}$ and zero otherwise by (1.5), it follows that

$$\frac{\partial e(\boldsymbol{z}_{t-1})}{\partial z_t} = \begin{cases} = 0 & \text{if } z_{t-1} \leq \delta v^*_t \text{ or } z_t \leq \delta v^*_{t+1} \\ < 0 & \text{if } z_{t-1} > \delta v^*_t \text{ and } z_t > \delta v^*_{t+1}. \end{cases}$$

□

Proof of Corollary 1. Consider $T > \tau \geq 1$. From Proposition 1, it follows that higher rewards in any periods $t \in \{\tau+1, \dots, T\}$ cause a discouragement effect if $z_t > \delta v^*_{t+1}$ holds for all $t \in \{\tau, \dots, T\}$. Hence, it remains to show that this condition is indeed satisfied if the reward profile $\boldsymbol{z}_\tau$ is either *(i)* non-increasing over time or *(ii)* the increase is sufficiently small.

To this end, suppose first that $\boldsymbol{z}_\tau$ is non-increasing over time. Consider period T. Since $z_T > 0$ and $v^*_{T+1} = 0$, it holds that $z_T > \delta v^*_{T+1}$. By Lemma 2, agents' equilibrium efforts are interior, $e^*_T > 0$. Since $e^*_T > 0$, agents' effort costs are strictly positive in period T, $c(e^*_T) > 0$. Hence, (1.1) implies that $v^*_T < \max\{z_T, v^*_{T+1}\} = z_T$. Consider period $T-1$ next. Since $\boldsymbol{z}_\tau$ is non-increasing over time, we have that $z_{T-1} \geq z_T > v^*_T$. By Lemma 2, agents' equilibrium efforts are strictly positive again, $e^*_{T-1} > 0$, which then implies $v^*_{T-1} < \max\{z_{T-1}, v^*_T\} = z_{T-1}$.

Repeating the previous arguments until period τ proves that $z_t > \delta v^*_{t+1}$ holds for all $t \in \{\tau, \dots, T\}$ if $\boldsymbol{z}_\tau$ is non-increasing over time. Finally, by continuity, $z_t > \delta v^*_{t+1}$ also holds for all $t \in \{\tau, \dots, T\}$ if $\boldsymbol{z}_\tau$ is increasing over time, but the increase is sufficiently small. □

Proof of Theorem 1. Let $T = 2$ and z_2 be given.

Part (i). Denote $\bar{z}$ the reward that implements $\bar{e}$ in period 1.[15] We first show that $z_1 \to \bar{z}^-$ implies MRS $\to \infty$ if P'' is zero on a sufficiently large interval $[\underline{e}, \bar{e}]$. To do so, note that if P'' is zero for $e \in [\underline{e}, \bar{e}]$, then evaluating the MRS at $\bar{z}$ and using that $P'(\bar{e}) = P'(\underline{e})$ yields

$$\frac{\Big(1 - P(e(\boldsymbol{z}_2))\Big) P'(\underline{e})}{\Big(1 - P(\bar{e})\Big) P'(e(\boldsymbol{z}_2))}. \tag{1.A.8}$$

If the interval $[\underline{e}, \bar{e}]$ is sufficiently large, then $P(\bar{e}) = P(\underline{e}) + P'(\underline{e})(\bar{e} - \underline{e}) \to 1$. Since all other terms in (1.A.8) are strictly positive and finite, this directly implies MRS $\to \infty$. Now, it remains to show that the MRE is sufficiently large for $\bar{z}$. To this end, note that evaluating (1.A.6) at $\bar{z}$ yields

$$\frac{\partial e(\bar{z}, v^*_2)}{\partial v^*_2} = \frac{\delta P'(\bar{e})}{-nc''(\bar{e})} = \frac{\delta P'(\underline{e})}{-nc''(\bar{e})}, \tag{1.A.9}$$

which converges to $-\infty$ if $c''(\bar{e}) \to 0$. Since z_2 is strictly positive, (1.5) shows that $\partial v_2(\boldsymbol{z}_2)/\partial z_2$ is strictly positive. This implies together with (1.A.7) and (1.A.9) that

15. Note from Lemma 2 that since $e(v^*_{t+1}, v^*_{t+1}) = 0$ and $\lim_{z_t \to \infty} e(z_t, v^*_{t+1}) = \infty$ indeed *any* effort level can be implemented.

$\partial e(\boldsymbol{z}_1)/\partial z_2 \to -\infty$ for $\bar{z}$ and $c''(\bar{e}) \to 0$. Since $z_2 > 0$, we find that $\partial e(\boldsymbol{z}_2)/\partial z_2$ is strictly positive and finite by (1.A.5). It thus holds that MRE $\to \infty$ for $\bar{z}$ and $c''(\bar{e}) \to 0$. Hence, by continuity, we can infer that if P'' is sufficiently close to zero on a sufficiently large interval $e \in [\underline{e}, \bar{e}]$ and $c''(\bar{e})$ is sufficiently small, then Condition (1.9) is satisfied for z_1 in the left local neighborhood of $\bar{z}$.

Part (ii). We show that (1.9) holds since $z_1 \to \delta v_2^{*+}$ implies MRE $\to \infty$ if P'' is strictly negative but finite and $P'(0)/c''(0) \to \infty$. To do so, note first that $z_1 \to \delta v_2^{*+}$ implies $e(z_1, v_2^*) \to 0^+$, by (FOC$_t$) and Assumption 1 and 2. Using (1.A.6), the marginal equilibrium effort is thus

$$\lim_{z_1 \to \delta v_2^{*+}} \frac{\partial e(z_1, v_2^*)}{\partial v_2^*} = \frac{\delta P'(0)}{P''(0)\,0 - nc''(0)} = -\frac{\delta P'(0)}{nc''(0)}. \tag{1.A.10}$$

Hence, (1.A.10) reveals that if P'' is finite and $P'(0)/c''(0) \to \infty$, then $z_1 \to \delta v_2^{*+}$ implies $\partial e(z_1, v_2^*)/\partial v_2^* \to -\infty$. Since z_2 is strictly positive, (1.5) shows that $\partial v_2(\boldsymbol{z}_2)/\partial z_2$ is strictly positive. We can thus conclude from (1.A.7) that $z_1 \to \delta v_2^{*+}$ implies $\partial e(\boldsymbol{z}_1)/\partial z_2 \to -\infty$. Since $z_2 > 0$ and P'' strictly negative, we find that $\partial e(\boldsymbol{z}_2)/\partial z_2$ is strictly positive and finite by (1.A.5). Accordingly, if $z_1 \to \delta v_2^{*+}$, then MRE $\to \infty$. To see that the MRS, on the other hand, is strictly positive in the limit for $z_1 \to \delta v_2^{*+}$, simply note that

$$\lim_{z_1 \to \delta v_2^{*+}} \frac{\big(1 - P(e(\boldsymbol{z}_2))\big) P'(e(\boldsymbol{z}_1))}{\big(1 - P(e(\boldsymbol{z}_1))\big) P'(e(\boldsymbol{z}_2))} = \frac{\big(1 - P(e(\boldsymbol{z}_2))\big) P'(0)}{P'(e(\boldsymbol{z}_2))} > 0,$$

because z_2 is finite. Hence, by continuity, we can infer that if P'' is finite and strictly negative while $P'(0)/c''(0)$ is sufficiently large, then Condition (1.9) is satisfied for z_1 in the right local neighborhood of δv_2^*. □

Proof of Corollary 2. Let $T = 2$ and z_2 be given such that $e_2^* \in (0, \bar{e})$. Denote $\bar{z}$ the (finite) reward that implements $\bar{e}$ in period 1. The proof then directly follows from (1.9). If $z_1 \to \bar{z}^-$, then $\big(1 - P(e(\boldsymbol{z}_1))\big) \to 0$, which implies that MRS $\to \infty$ since all other terms are either independent of z_1 or strictly positive and finite. Moreover, from (1.A.7) together with (1.A.6) and (1.5), it follows that MRE is strictly positive and finite for $z_1 \to \bar{z}^-$ if $\lim_{e_{t,i} \to \bar{e}} c''(e_{t,i})$ is finite. Hence, Condition (1.9) is satisfied for $z_1 \to \bar{z}^-$ since the l.h.s. converges to infinity. We can thus conclude that (1.9) holds for z_1 in the left local neighborhood of $\bar{z}$. □

Proof of Proposition 2. Using Equation (1.1), Lemma 1, and the Envelope Theorem, we obtain

$$\begin{aligned}\frac{\partial v_{t-1}(\mathbf{z}_{t-1})}{\partial z_t} &= \left(1-P(e(\mathbf{z}_{t-1}))\right)\frac{\partial v_t(\mathbf{z}_t)}{\partial z_t} \\ &\qquad + \left(z_{t-1}-\delta v_t^*\right)\tfrac{n-1}{n}P'(e(\mathbf{z}_{t-1}))\frac{\partial e(\mathbf{z}_{t-1})}{\partial z_t}.\end{aligned} \tag{1.A.11}$$

If $z_t \leq \delta v_{t+1}^*$, then $\partial v_{t-1}(\mathbf{z}_{t-1})/\partial z_t = 0$ by (1.A.11) together with (1.5) and (1.6). By contrast, if $z_t > \delta v_{t+1}^*$, then (1.5) implies that $\partial v_t(\mathbf{z}_t)/\partial z_t > 0$. Using (1.A.7), we can then rewrite (1.A.11) as

$$\frac{\partial v_{t-1}(\mathbf{z}_{t-1})}{\partial z_t} \propto \underbrace{1-P(e(\mathbf{z}_{t-1}))}_{>0}+\underbrace{\left(z_{t-1}-\delta v_t^*\right)\tfrac{n-1}{n}P'(e(\mathbf{z}_{t-1}))\frac{\partial e(z_{t-1},v(\mathbf{z}_t))}{\partial v_t^*}}_{\leq 0}. \tag{1.A.12}$$

The first term on the r.h.s. of (1.A.12) is always positive by Assumption 1, while the second term is strictly negative for $z_{t-1} > \delta v_t^*$ and zero otherwise by Assumption 1 and Lemma 2. Hence, two countervailing effects arise if $z_{t-1} > \delta v_t^*$ and $z_t > \delta v_{t+1}^*$ holds. □

Proof of Theorem 2. Let $T \geq t > 1$, $n > 1$ and $\mathbf{z}_t$ be given such that $z_t > \delta v_{t+1}^*$. Denote $\bar{z}$ the reward that implements $\bar{e}$ in period $t-1$.[16] From Proposition 2, we can infer that $\partial v_{t-1}(\mathbf{z}_{t-1})/\partial z_t < 0$ if and only if

$$1-P(e(\mathbf{z}_{t-1})) < -\left(z_{t-1}-\delta v_t^*\right)\tfrac{n-1}{n}P'(e(\mathbf{z}_{t-1}))\frac{\partial e(z_{t-1},v(\mathbf{z}_t))}{\partial v_t^*}.$$

Inserting (FOC_t) and (1.A.6) yields

$$1-P(e(\mathbf{z}_{t-1})) < \frac{-\delta(n-1)c'(e(\mathbf{z}_{t-1}))P'(e(\mathbf{z}_{t-1}))}{P''(e(\mathbf{z}_{t-1}))\left(z_{t-1}-\delta v_t^*\right)-nc''(e(\mathbf{z}_{t-1}))}. \tag{1.A.13}$$

If P'' is zero for $e \in [\underline{e}, \bar{e}]$, then evaluating (1.A.13) for $\bar{z}$ and using that $P'(\underline{e}) = P'(\bar{e})$ yields

$$1-P(\bar{e}) < \frac{\delta(n-1)}{n}\frac{c'(\bar{e}))P'(\underline{e})}{c''(\bar{e})}. \tag{1.A.14}$$

If the interval $[\underline{e}, \bar{e}]$ is sufficiently large, then $P(\bar{e}) = P(\underline{e}) + P'(\underline{e})(\bar{e}-\underline{e}) \to 1$ such that the l.h.s. of (1.A.14) converges to zero. If, in addition, $c''(\bar{e}) \to 0$, then the

16. Note from Lemma 2 that since $e(v_{t+1}^*, v_{t+1}^*) = 0$ and $\lim_{z_t\to\infty} e(z_t, v_{t+1}^*) = \infty$ indeed *any* effort level can be implemented.

r.h.s. of (1.A.14) converges to infinity. Hence, by continuity, we can infer that if P'' is sufficiently close to zero on a sufficiently large interval $[\underline{e}, \overline{e}]$ and $c''(\overline{e})$ is sufficiently small, then Condition (1.9) is satisfied for z_1 in the left local neighborhood of $\overline{z}$. □

Proof of Corollary 3. The corollary follows directly from the existence of success and payoff reversals in combination with continuity of equilibrium efforts and expected payoffs: if increases in rewards are sufficiently small in periods that have a positive impact on the success rate/expected payoff, then they can be overcompensated by sufficiently large increases in periods with a negative impact. □

Proof of Proposition 3. The proposition follows directly from comparing Figures 1.3 and 1.4. □

1.A.2 Curvature of Equilibrium Efforts

For $z_t > \delta v^*_{t+1}$, the equilibrium efforts are determined by (FOC$_t$), and their derivative with respect to z_t is (1.A.5). Computing the derivative of (1.A.5), we get that

$$\begin{aligned}\frac{\partial^2 e(z_t, v^*_{t+1})}{\partial z_t^2} &= -\frac{P''(\cdot)\frac{\partial e(z_t,v^*_{t+1})}{\partial z_t}\left[P''(\cdot)\left(z_t-\delta v^*_{t+1}\right)-nc''(e(z_t,v^*_{t+1}))\right]}{\left[P''(\cdot)\left(z_t-\delta v^*_{t+1}\right)-nc''(e(z_t,v^*_{t+1}))\right]^2} \\ &+\frac{P'(\cdot)\left(P''(\cdot)+\frac{\partial e(z_t,v^*_{t+1})}{\partial z_t}\left[P'''(\cdot)\left(z_t-\delta v^*_{t+1}\right)-nc'''(e(z_t,v^*_{t+1}))\right]\right)}{\left[P''(\cdot)\left(z_t-\delta v^*_{t+1}\right)-nc''(e(z_t,v^*_{t+1}))\right]^2}.\end{aligned} \tag{1.A.15}$$

Inserting Equation (1.A.5) in (1.A.15) and further simplifying yields

$$\begin{aligned}\frac{\partial^2 e(z_t, v^*_{t+1})}{\partial z_t^2} &= \frac{2\overbrace{P'(\cdot)}^{>0}\overbrace{P''(\cdot)}^{\leq 0}}{\underbrace{\left[P''(\cdot)\left(z_t-\delta v^*_{t+1}\right)-nc''(e(z_t,v^*_{t+1}))\right]^2}_{>0}} \\ &+\frac{\overbrace{P'(\cdot)}^{>0}\overbrace{\frac{\partial e(z_t,v^*_{t+1})}{\partial z_t}}^{>0}\overbrace{\left[P'''(\cdot)\left(z_t-\delta v^*_{t+1}\right)-nc'''(e(z_t,v^*_{t+1}))\right]}^{=:\text{Factor 1}}\Big)}{\underbrace{\left[P''(\cdot)\left(z_t-\delta v^*_{t+1}\right)-nc''(e(z_t,v^*_{t+1}))\right]^2}_{>0}}.\end{aligned} \tag{1.A.16}$$

Consider the case where $P'' = 0$ first. In this case, the second derivative is negative if $c''' > 0$, zero if $c''' = 0$, and positive if $c''' < 0$.

Now, consider the case where $P'' < 0$. In this case, the sign of the second derivative is determined by the interplay of the derivatives of p and c. If Factor 1 is negative, or positive but sufficiently small, the second derivative is negative such that $e(z_t, v_{t+1}^*)$ is concave in z_t for $z_t > \delta v_{t+1}^*$. Note that this is always satisfied for $c''' \geq 0$ and $P''' \leq 0$. If Factor 1 is positive and sufficiently large, the second derivative is positive, implying the convexity of $e(z_t, v_{t+1}^*)$. This is the case whenever P'' and P''' are close to zero, and $c''' < 0$ is sufficiently small.

1.A.3 Success Rates with Multiple Periods

We briefly discuss the case of more than two periods. With $T > 2$, define the conditional probability

$$P^{\text{After}} := \text{prob}\,(\text{success in } t > 2 \mid \text{no success in } t = 1, 2)$$

and rewrite the success rate as

$$S(\mathbf{z}_1) = \Big[P(e(\mathbf{z}_1)) + \big(1 - P(e(\mathbf{z}_1))\big)P(e(\mathbf{z}_2))\Big]\big(1 - P^{\text{After}}\big) + P^{\text{After}}.$$

Since z_2 only affects equilibrium efforts in the first two periods, P^{After} is independent of z_2. Therefore, it holds that

$$\frac{\partial S(\mathbf{z}_1)}{\partial z_2} \propto \big(1 - P(e(\mathbf{z}_1))\big)P'(e(\mathbf{z}_2))\frac{\partial e(\mathbf{z}_2)}{\partial z_2} + \big(1 - P(e(\mathbf{z}_2))\big)P'(e(\mathbf{z}_2)))\frac{\partial e(\mathbf{z}_1)}{\partial z_2}. \tag{1.A.17}$$

Thus, whether $\partial S(\mathbf{z}_1)/\partial z_2 \gtreqless 0$ in the case of three or more periods is determined by the same condition as in the case of two periods, cf. (1.8) and (1.A.17).

This observation has a striking implication: From any two-period reward profile that causes a success reversal, we can construct reward profiles in the model with $T > 2$ that also induce success reversals. The idea is as follows. Suppose the reward profile $\tilde{\mathbf{z}}_1 = (\tilde{z}_1, \tilde{z}_2)$ generates a success reversal, then the resulting equilibrium efforts satisfy Condition (1.A.17). We construct a new reward profile $\hat{\mathbf{z}}_1 = (\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_T)$, where we can select arbitrary rewards $\hat{\mathbf{z}}_3$. The equilibrium efforts in the first two periods remain unchanged if we choose $\hat{z}_1$ and $\hat{z}_2$, such that the new reward profile provides the same net rewards in these periods as the initial profile. This requires setting $\hat{z}_2 = \tilde{z}_2 + \delta\hat{v}_3^*$ and $\hat{z}_1 = \tilde{z}_1 + \delta\hat{v}_2^* - \delta\tilde{v}_2^*$. Since the conditions for a success reversal are otherwise identical (see arguments above), there must also be a success reversal for the new reward profile $\hat{\mathbf{z}}_1$.

It is straightforward that the success rate can also decrease in the rewards of later periods. The simplest example is the one where there are negative net rewards in the first $T-2$ periods and positive net rewards in the last two periods. The agents then fail for sure in the first $T-2$, such that the success rate reads

$$S(\mathbf{z}_1) \;=\; P(e(\mathbf{z}_{T-1})) + \big(1-P(e(\mathbf{z}_{T-1}))\big)P(e(\mathbf{z}_T)).$$

Since efforts are zero in the first $T-2$ periods, it holds that

$$\begin{aligned}\frac{\partial S(\mathbf{z}_1)}{\partial z_T} =& \big(1-P(e(\mathbf{z}_{T-1}))\big)P'(e(\mathbf{z}_T))\frac{\partial e(\mathbf{z}_T)}{\partial z_T} \\ &+ \; \big(1-P(e(\mathbf{z}_T))\big)P'(e(\mathbf{z}_T)))\frac{\partial e(\mathbf{z}_{T-1})}{\partial z_T}.\end{aligned} \tag{1.A.18}$$

Thus, except for the time indices, the conditions for a success reversal are identical to the case of two periods, cf. (1.8) and (1.A.18). Again, this allows us to construct reward profiles in the model with $T>2$ that also induce success reversals. Suppose the reward profile $\tilde{\mathbf{z}}_1 = \big(\tilde{z}_1, \tilde{z}_2\big)$ generates a success reversal. The reward profile $\hat{\mathbf{z}}_1 = \big(\hat{z}_1, \ldots, \hat{z}_{T-2}, \tilde{z}_1, \tilde{z}_2\big)$ then also causes a success reversal, if $\hat{z}_t < \delta\hat{v}^*_{t+1}$ holds for all $t \in \{1,\ldots,T-2\}$.[17]

17. The reward profile $\hat{\mathbf{z}}_1 = \big(\varepsilon, \ldots, \varepsilon, \tilde{z}_1, \tilde{z}_2\big)$ where ε is sufficiently small would be an example of a reward profile that gives rise to a success reversal since $\hat{z}_t < \delta\hat{v}^*_{t+1}$ is satisfied for all $t \in \{1,\ldots,T-2\}$.

References

Altan, B. (2019). Dynamic moral hazard with sequential tasks. *Economics Letters*, *183*, 108606. [16, 23]

Arenas, M., Reutter, J., Toussaint, E., Ugarte, M., Vial, F., & Vrgoč, D. (2020). Cryptocurrency mining games with economic discount and decreasing rewards. In C. Paul & M. Bläser (Eds.), *37th international symposium on theoretical aspects of computer science (stacs 2020)* (54:1–54:16). Dagstuhl Publishing. [23]

Aumann, R. J. (2006). War and peace. *Proceedings of the National Academy of Sciences*, *103*(46), 17075–17078. [13]

Baktash, M. B., Heywood, J. S., & Jirjahn, U. (2022). Worker stress and performance pay: German survey evidence. *Journal of Economic Behavior & Organization*, *201*, 276–291. [32]

Bell, B., & Van Reenen, J. (2014). Bankers and their bonuses. *Economic Journal*, *124*(574), F1–F21. [32]

Bellman, R. (1957). *Dynamic programming*. Princeton University Press. [18]

Bergemann, D., & Hege, U. (2005). The financing of innovation: Learning and stopping. *RAND Journal of Economics*, *36*(4), 719–752. [16]

Bolton, P., & Harris, C. (1999). Strategic experimentation. *Econometrica*, *67*(2), 349–374. [14]

Bonatti, A., & Hörner, J. (2011). Collaborating. *American Economic Review*, *101*(2), 632–663. [15]

Eizakshiri, F., Chan, P. W., & Emsley, M. W. (2015). Where is intentionality in studying project delays? *International Journal of Managing Projects in Business*, *8*(2), 349–367. [15]

Ekici, A., & Retharekar, A. (2013). Multiple agents maximum collection problem with time dependent rewards. *Computers & Industrial Engineering*, *64*(4), 1009–1018. [23]

Fabricius, G., & Büttgen, M. (2015). Project managers' overconfidence: How is risk reflected in anticipated project success? *Business Research*, *8*, 239–263. [32]

Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*, *65*(4), 833–860. [15]

Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review*, *87*(4), 746–755. [15]

Georgiadis, G. (2015). Projects and team dynamics. *Review of Economic Studies*, *82*(1), 187–218. [16]

Georgiadis, G., Lippman, S. A., & Tang, C. S. (2014). Project design with limited commitment and teams. *RAND Journal of Economics*, *45*(3), 598–623. [16]

Gneezy, U., & Rustichini, A. (2000a). A fine is a price. *Journal of Legal Studies*, *29*(1), 1–17. [15]

Gneezy, U., & Rustichini, A. (2000b). Pay enough or don't pay at all. *Quarterly Journal of Economics*, *115*(3), 791–810. [15]

Green, B., & Taylor, C. (2016). Breakthroughs, deadlines, and self-reported progress: Contracting for multistage projects. *American Economic Review*, *106*(12), 3660–3699. [16]

Halac, M., Kartik, N., & Liu, Q. (2016). Optimal contracts for experimentation. *Review of Economic Studies*, *83*(3), 1040–1091. [16]

Hong, B., Kueng, L., & Yang, M.-J. (2019). Complementarity of performance pay and task allocation. *Management Science*, *65*(11), 5152–5170. [32]

Hörner, J., & Samuelson, L. (2013). Incentives for experimenting agents. *RAND Journal of Economics*, *44*(4), 632–663. [16]

Jenkins, A. M., Naumann, J. D., & Wetherbe, J. C. (1984). Empirical investigation of systems development practices and results. *Information & Management*, *7*(2), 73–82. [15]

Klor, E. F., Kube, S., Winter, E., & Zultan, R. (2014). Can higher rewards lead to less effort? incentive reversal in teams. *Journal of Economic Behavior & Organization*, *97*, 72–83. [15]

Lemieux, T., MacLeod, W. B., & Parent, D. (2009). Performance pay and wage inequality. *Quarterly Journal of Economics*, *124*(1), 1–49. [32]

Mason, R., & Välimäki, J. (2015). Getting it done: Dynamic incentives to complete a project. *Journal of the European Economic Association*, *13*(1), 62–97. [16, 23]

Mayer, S. (2022). Financing breakthroughs under failure risk. *Journal of Financial Economics*, *144*(3), 807–848. [23]

Moroni, S. (2022). Experimentation in organizations. *Theoretical Economics*, *17*(3), 1403–1450. [15]

Park, J. E. (2021). Schedule delays of major projects: What should we do about it? *Transport Reviews*, *41*(6), 814–832. [15, 32]

Rauber, T., & Weinschenk, P. (2023). Detrimental incentive mechanisms in dynamic principal-agent relationships. *Working Paper. Available at SSRN 4515328*. [14]

Van Genuchten, M. (1991). Why is software late? an empirical study of reasons for delay in software development. *IEEE Transactions on Software Engineering*, *17*(6), 582–590. [32]

Weinschenk, P. (2016). Procrastination in teams and contract design. *Games and Economic Behavior*, *98*, 264–283. [16]

Weinschenk, P. (2021). On the benefits of time-inconsistent preferences. *Journal of Economic Behavior & Organization*, *182*, 185–195. [16]

Winter, E. (2009). Incentive reversal. *American Economic Journal: Microeconomics*, *1*(2), 133–47. [15]

Chapter 2

Detrimental Incentive Mechanisms in Dynamic Principal-Agent Relationships*

Abstract

This article explores a dynamic moral-hazard setting in which a principal hires a team of agents for a project. As the project generates revenue upon completion, the principal incentivizes agents' efforts by designing bonuses for success. If bonuses are provided through spot or renegotiation-proof long-term contracts, an increase in project revenue may adversely affect *(i)* agents' expected payoffs, *(ii)* the principal's expected profit, and *(iii)* the likelihood of project success. Even for long-term contracts under full commitment, the detrimental effects on agents and project success can persist, while, for the principal, these incentive mechanisms render higher revenues necessarily beneficial.

Keywords: Principal-Agent Models, Dynamic Moral Hazard, Teams, Incentives, Managerial Bonuses, Renegotiation-Proofness.

JEL Classification: C73, D82, D86, J33, M52.

*Acknowledgments. We are grateful to Heski Bar-Isaac, Matthew Elliott, Bård Harstad, Fabian Herweg, Daniel Heyen, Çağıl Koçyiğit, Daniel Müller, Paul Ritschel, Flavio Toxvaerd, and Philipp Weinschenk for valuable feedback and ideas. This paper has also benefited from the audience at the 2024 Asia Meeting of the Econometric Society in New Delhi, as well as research seminars at the Universities of Cambridge, Hohenheim, Kaiserslautern-Landau, and Luxembourg. Finally, for generous financial support, we owe a great debt of gratitude to Deutsche Forschungsgemeinschaft (Grant No. 431144273).

2.1 Introduction

Since its introduction by Ross (1973) and Jensen and Meckling (1976), the principal-agent paradigm has been widely used as a tool to analyze the effectiveness and design of incentives. Much work has been devoted ever since to exploring how a principal optimally sets incentives for an agent under asymmetric information (cf. literature syntheses by Bolton and Dewatripont, 2005; Laffont and Martimort, 2002; Salanié, 2005). The resulting literature characterizes incentive mechanisms in the form of optimal contracts for many conceivable situations of interaction between a principal and an agent. In particular, starting with seminal contributions by Mirrlees (1976) and Holmström (1979), the principal-agent relationship overshadowed by *moral hazard* has received special attention in economic research.[1] Optimal contracts were, for instance, derived for moral hazard in conjunction with multiple tasks (e.g., Holmström and Milgrom, 1991; Laux, 2001; Mylovanov and Schmitz, 2008), multiple agents (e.g., Halac et al., 2021; Holmström, 1982; Winter, 2004), and dynamic interaction (e.g., Piskorski and Westerfield, 2016; Rogerson, 1985; Sannikov, 2008).

Our paper contributes to this literature by considering a fairly general model of dynamic moral hazard that allows for multiple agents to investigate *(i)* how the economic environment affects equilibrium incentive mechanisms under different contract classes (spot contracting or long-term contracting with limited or full commitment) and *(ii)* how these, in turn, affect the contracting parties and project success. We reveal several novel and surprising effects: inter alia, it is shown that higher revenues upon project completion may generate an incentive mechanism that leaves the agents and the principal worse off, while at the same time reducing the likelihood of success. We gain these insights in a parsimonious moral-hazard model, where a principal (female) hires a team of agents (male) to work on a project. The agents have a finite number of periods to succeed with the project, and the success probability in a period is determined by agents' efforts in the respective period. Since the completed project generates a time-dependent revenue for the principal, she incentivizes her agents to provide efforts by offering time-dependent bonuses upon project success.

In studying incentive mechanisms in this model, we first explore how the principal designs spot contracts in equilibrium. Future spot contracts can be correctly foreseen by all parties due to rational expectations. Accordingly, the principal antic-

1. See Georgiadis (2022) for an excellent overview of the theoretical and empirical literature on moral hazard.

ipates the profit she can expect from the rest of the game when designing bonuses, just as agents anticipate their future payoffs when they decide on their effort provision. We later show that the equilibrium sequence of spot contracts obtained in this setting coincides with the equilibrium renegotiation-proof long-term contract. Therefore, our results gained for spot contracts carry over entirely to a second highly relevant class of incentive mechanisms.

Analyzing how more favorable economic environments, i.e., higher revenues upon project success, change the design of spot contracts reveals several interesting effects. From an intraperiod perspective, higher revenues cause the principal to increase agents' bonuses in order to incentivize higher efforts. However, from an interperiod perspective, higher bonuses discourage agents' effort provision in the previous period. The principal correctly foresees this *discouragement effect* and may react in two possible ways. First, the principal may increase the bonus payment in the preceding period in order to counteract the discouragement effect. We establish conditions when this *reincentive effect* arises and show that it is never optimal for the principal to offset or even overcompensate the discouragement effect completely. Instead, the principal only increases earlier bonuses up to a level that equilibrium efforts in the previous period remain lower than initially. Second, the principal may decrease the bonuses in the preceding period and thereby amplify the discouragement effect. We specify intuitive conditions for this *disincentive effect* to occur. These mechanisms, originating from agents' optimizing behavior and the principal's equilibrium contract design, can then engender three surprising consequences for the contracting parties and the project itself.

(i) From an agent's perspective, a higher revenue may entail an equilibrium contract that harms his payoff that he can expect from working on the project. There are two root causes for this *payoff reversal*. First, although higher revenues lead to higher bonuses in the same period – a positive effect for the expected payoff –, the negative effect of lower equilibrium bonuses in the previous period due to the disincentive effect can outweigh. Second, an additional negative effect arises from the team externality in the case of multiple agents. Since teammates reduce their equilibrium efforts in the previous period, succeeding in that period becomes less likely, which can ultimately harm the agent's expected payoff. With multiple agents, expected payoffs may thus decline even if higher revenues cause the principal to pay higher bonuses in all periods.

(ii) Even more surprisingly, the incentive mechanisms that the principal designs with higher revenues can reduce her own expected profit. The intuition behind this *profit reversal* is as follows. Higher revenues are per se – i.e., for a given contract – beneficial for the principal's expected profit. However, with higher revenues, the equilibrium contract may change and incentivize higher efforts in less lucrative periods, while implementing lower efforts in more profitable periods.

(iii) Focusing on the project itself, the incentive mechanism associated with higher revenues may induce an intertemporal distribution of efforts that deteriorates the chances for successful project completion. This is a phenomenon we call *success reversal*.

The second result is probably the most remarkable one. From an abstract perspective, it shows that a player can actually suffer from higher payments if she has to interact with her own future self. This holds even though her expectations are perfectly rational and her preferences are time-consistent.[2]

Moreover, we investigate whether these reversals also arise in the scenario with full commitment (where the principal can commit to *any* long-term contract). It is straightforward that profit reversals cannot occur with full commitment. The reason is that the principal is then able to maintain a contract if revenues increase, such that she necessarily benefits from higher revenues. By contrast, we show that payoff and success reversal can still emerge with full commitment.

These insights gained in our simple and intuitive model therefore serve as a theoretical foundation that might help to explain why lucrative projects tend to be delayed or fail completely. Agents may strategically withhold efforts in lucrative environments harming the project and its stakeholders when their incentives are designed endogenously and they engage in dynamic interaction. Since such settings are frequently encountered in practice, our results offer an explanation of why project delays and failures are such a prevalent phenomenon even in highly profitable projects. Indeed, there is ample empirical and anecdotal evidence in both academic research and the popular press reporting this phenomenon; see the conclusion for a detailed discussion. Our analysis hints towards the crucial importance of the intertemporal distribution of a project's revenues for its performance, which has been widely overlooked so far.

2. It is well-known that, with erroneous beliefs or time-inconsistent preferences, a player might act against her own long-run interest if she is offered options that are tempting in some aspects, but detrimental to her long-run interests. For instance, longer deadlines can cause time-inconsistent players to procrastinate, which might ultimately be harmful.

Related Literature. Incentive mechanisms in the form of spot contracts and long-term contracts are frequently encountered in the literature (e.g., Chiappori et al., 1994; Fudenberg et al., 1990; Holmström, 1983). The latter form is subdivided in *(i)* long-term contracts with full commitment and *(ii)* renegotiation-proof long-term contracts where the principal cannot exclude future renegotiations due to her limited commitment power.

Pioneering contributions in the dynamic principal-agent literature examine the relationship between those three contract classes. When comparing the outcomes of long-term contracts under full and limited commitment, early papers by Chiappori et al. (1994), Dewatripont (1989), and Laffont and Tirole (1990) reveal that the lack of commitment constitutes a severe source of friction in dynamic principal-agent relationships. That is, limited commitment leads to incentive mechanisms that are less favorable for the principal since, in the words of Rey and Salanie (1990, p. 516), "*the incentive problems created by private information are generally best overcome through ex ante commitment to ex post inefficiencies*." For the principal in our model, it is beneficial to commit to potentially insufficient bonuses in late periods to obtain "cheap incentives" early in the game, i.e., to incentivize substantial efforts in early periods with relatively low bonus payments. Concerning spot and renegotiation-proof long-term contracts, Hart and Tirole (1988) were among the first to find that both contracts can, but do not necessarily have to, result in the same outcome. Intuitively, as emphasized by Bolton (1990), renegotiation-proof long-term contracts deviate from the sequence of spot contracts whenever they allow parties to improve efficiency through intertemporal transfers. In our model, such transfers would take the form of "cheap incentives" in early periods. Since renegotiation-proof long-term contracts are shown to be unable to implement these "cheap incentives", they do not enable efficiency-enhancing transfers, which is why they coincide with the sequence of spot contracts in equilibrium.

Among the related articles in the dynamic moral-hazard literature, many investigate the design of incentive mechanisms for project completion in *single-agent* settings. With one agent and multi-staged projects, Toxvaerd (2006) shows that the renegotiation-proof long-term contract performs better in terms of the expected project completion time than spot contracts. In addition, he finds that increasingly high-powered incentives should be provided as progress is made. Mason and Välimäki (2015), extended by Altan (2019), on the contrary, find optimal incentives for project completion to be decreasing over time if they are provided via spot contracts or long-term contracts with full commitment. More recent single-agent contributions by Green and Taylor (2016), Halac et al. (2016), Mayer (2022), and Varas (2018) investigate the design of long-term contracts under full commitment in moral-hazard models with additional manifestations of asymmetric information. For a principal, who can neither observe the agent's effort nor his project-specific skills, Halac et al. (2016) characterize the optimal bonus and penalty contracts. Varas (2018) investigates a setting with hidden effort and imperfect observability of the project's quality. He finds the optimal long-term contract to be two-staged, with a first phase of dynamic incentives followed by a second phase where incentives are stationary. Similarly, the optimal incentive mechanism is shown to consist of two stages in Mayer (2022) if the agent is able to hide project failure. Solely in the second stage does the principal provide high-powered incentives for success and disclosure of failure. An optimal long-term contract that is two-staged is also specified by Green and Taylor (2016) for the case where the agent can, in addition to his effort, privately observe project progress.

In contrast to the above papers, our model additionally accommodates team problems by allowing *multiple agents*. The literature on the design of incentive mechanisms for a project under dynamic moral hazard is rather sparse in this context (Chandrasekher, 2015). Che and Yoo (2001) find that the optimal incentive mechanism for limited commitment is either characterized by high-powered individual incentives or by low-powered team-based incentives. Chandrasekher (2015) shows for multiple outputs (a series of projects) how the first-best outcomes for all contracting parties are approximated by spot contracts in conjunction with an auditing mechanism. In a model where efforts accumulate over time, Georgiadis (2015) reveals that the optimal symmetric contract rewards agents only for completing the project, not for reaching intermediate milestones. He further demonstrates that such contracts specify higher rewards the longer the project's length is. A related contribution from the literature on strategic experimentation is Moroni (2022).[3] She finds the optimal long-term contract under full commitment for a two-stage project with unknown feasibility to be asymmetric regarding agents' remuneration and task allocation.

Our basic setup also shares some features with those in Weinschenk (2016, 2021), while our main findings are closest to those reported in dynamic moral-hazard models by Ohlendorf and Schmitz (2012), as well as Rauber and Weinschenk (2024). Although the latter describe some mechanisms that parallel those leading to success and payoff reversals in our model, they focus on exogenous incentives. We, however, take a design perspective and examine how contracts are endogenously designed by a principal. Thereby, we also identify adverse effects for the

3. Other, to some extent, related papers from this strand of literature are, for instance, Bergemann and Hege (2005), Bonatti and Hörner (2011), and Hörner and Samuelson (2013).

designing party in the form of profit reversals, which are logically absent in their model. Ohlendorf and Schmitz (2012), on the other hand, discover that a principal with limited commitment may prefer a project with lower returns. Nonetheless, their mechanism is utterly different from ours and somewhat limited to the two-period structure of their model: if the returns of the project are too high, then, due to her limited commitment power, the principal loses her credibility to use the threat of project termination in the second period to incentivize the agent in the first period.

Outline. The remainder of this article is organized as follows. The next section lays out the basic principal-agent framework, including all relevant assumptions. In Section 2.3, we examine the moral-hazard component of the model to derive agents' equilibrium effort provision and establish the equilibrium bonus scheme for spot contracts. Afterwards, we conduct the core analysis and present our central results concerning payoff, profit, and success reversals. Section 2.5 then directs attention to long-term contracts by analyzing the extent to which our results carry over to these incentive mechanisms. We show, inter alia, that the equilibrium renegotiation-proof long-term contract is identical to the sequence of equilibrium spot contracts. The final section concludes our analysis and highlights essential implications. All proofs are provided in Appendix 2.A.

2.2 Model

A risk-neutral principal hires a set of risk-neutral agents $\mathcal{N} := \{1, \ldots, n\}$ to work on her project. The agents have $T \in \mathbb{N}_+$ time periods to complete the project. The probability that the project succeeds in period $t \in \{1, \ldots, T\}$ is determined by a success function $p(e_{t,1}, \ldots, e_{t,n})$, where $e_{t,i} \in \mathbb{R}_+$ denotes the effort exerted in period t by agent i.

Assumption 3 (Success Function).
The success function $p : \mathbb{R}^n_+ \to [0, 1)$ is thrice continuously differentiable, symmetric, strictly increasing, weakly concave, and satisfies[4]

$$p(0, \ldots, 0) = 0, \quad \textit{and} \quad \frac{\partial^3 p(e_{t,1}, \ldots, e_{t,n})}{\partial e_{t,i}^3} \leq 0.$$

4. Note that these assumptions allow agents' efforts to be either substitutes or complements.

Agents' efforts are non-contractible and chosen simultaneously in every period. Investing effort causes costs for agent i in period t that are captured by a cost function $c(e_{t,i})$, which satisfies the following assumption.

Assumption 4 (Cost Function)**.**
The cost function $c : \mathbb{R}_+ \to \mathbb{R}_+$ *is thrice continuously differentiable, strictly increasing, strictly convex, and satisfies* $c(0) = c'(0) = 0$, *and* $c''' \geq 0$.

As will become clear later, the properties imposed on the third derivatives in Assumptions 3 and 4 guarantee that agents' equilibrium efforts are concave and equilibrium bonus payments are unique.

If the project succeeds in period t, the principal earns a revenue $r_t \in \mathbb{R}_{++}$, pays each agent a bonus b_t, and the game terminates. Agents are protected by limited liability such that $b_t \in \mathbb{R}_+$. The bonus payment b_t can be interpreted as agents' incentive in t, since it motivates them to invest effort in this period.[5] Agents' reservation utilities are set to zero such that the limited-liability constraint implies their participation. Revenues from period t onwards define a *revenue profile* $\boldsymbol{r}_t := (r_t, \ldots, r_T)$. Similarly, agents' bonuses define a *bonus scheme* $\boldsymbol{b}_t := (b_t, \ldots, b_T)$. If agents do not succeed until the end of period T, the game also terminates. The principal then receives no revenue and pays no bonus. We assume that all parties are rational (in particular, have correct expectations) and deploy the subgame-perfect Nash equilibrium as a solution concept.

It should be stressed that our model contains some implicit symmetry assumptions: all agents are homogeneous, and both functions p and c are time-invariant.[6] The model captures static ($T = 1$) and dynamic ($T > 1$) moral-hazard settings, as well as single-agent ($n = 1$) and team ($n > 1$) problems.

5. Additionally to the bonus payment, agents might also receive a base wage that is independent of success. Since agents' effort provision is independent of such a base wage, the principal would implement this base wage as low as possible, i.e., a base wage of zero. We can thus omit base wages in our analysis.

6. In our analysis, we concentrate on the effect of a more favorable economic environment in the sense of higher project revenues. We could also allow for time-dependent cost or success functions, since lower costs or higher success probabilities constitute a more favorable environment as well. Technically, this affects the parties' continuation payoffs and thus their incentives in a similar manner to higher revenues.

2.3 Effort Provision and Bonus Design

Before we can turn our attention to the principal's equilibrium design of bonus schemes, it is first necessary to analyze the efforts that agents choose in equilibrium if they face given bonuses.

2.3.1 Equilibrium Effort Choice

Given the bonus scheme $\boldsymbol{b}_1$, the decision problem of the representative agent i in period t manifests in maximizing his expected payoff over his effort choice $e_{t,i}$. Using the Principle of Optimality by Bellman (1957), the agent's maximization problem in t reads

$$\begin{aligned}\max_{e_{t,i}\in\mathbb{R}_+} v_{t,i}(e_{t,1},\dots,e_{t,n}) \;=\; & p(e_{t,1},\dots,e_{t,n})\, b_t \\ & + \big(1-p(e_{t,1},\dots,e_{t,n})\big)\,\delta v^*_{t+1,i} - c(e_{t,i}),\end{aligned} \tag{2.1}$$

where $\delta \in (0,1]$ is the time discount factor and $v^*_{t+1,i}$ denotes the expected payoff in the subsequent period $t+1$, given equilibrium effort choices from $t+1$ onward, that is, the agent's *continuation payoff*.[7] Lemma 3 characterizes the central properties of the equilibrium efforts, i.e., the efforts that arise if all agents choose their effort provision as a solution to Problem (2.1).

Lemma 3. *Let the bonus scheme $\boldsymbol{b}_1$ be given. Then the following holds.*

(i) There exists a unique subgame perfect Nash equilibrium, which is symmetric in the sense that $e^*_{t,i} = e^*_{t,j} = e^*_t$ and $v^*_{t,i} = v^*_{t,j} = v^*_t$ for all $i,j \in \mathcal{N}$ and $t \in \{1,\dots,T\}$.

(ii) Agent i's equilibrium effort in period t is

$$e^*_t \;=\; e(b_t, v^*_{t+1}) \;:=\; \begin{cases} 0 & \text{if } b_t \;\leq\; \delta v^*_{t+1} \\ e^{\text{FOC}}(b_t, v^*_{t+1}) & \text{if } b_t \;>\; \delta v^*_{t+1}, \end{cases}$$

where the map $e^{\text{FOC}}(b_t, v^*_{t+1})$ is uniquely determined by

$$\frac{\partial p(e_t,\dots,e_t)}{\partial e_{t,i}}\big(b_t - \delta v^*_{t+1}\big) - c'(e_t) \;=\; 0. \tag{FOC$_t$}$$

(iii) If $b_t > \delta v^*_{t+1}$, then $e(b_t, v^*_{t+1})$ is strictly increasing and weakly concave in b_t.

7. Note that, since no payments are made after period T, the continuation payoff satisfies $v^*_{t+1,i} = 0$ for all $t \geq T$.

Lemma 3 *(i)* reveals that, for a given bonus scheme, the game possesses a unique subgame perfect Nash equilibrium in which all agents exert equal effort levels within a period and thus expect identical payoffs. Part *(ii)* is also intuitive. If the net bonus $b_t - \delta v^*_{t+1}$ (i.e., the difference between the bonus for project success and the discounted continuation payoff) is non-positive, then agents are at least weakly better off when the game continues. They thus ensure the game's continuation by investing zero effort. For positive net bonuses, by contrast, each agent chooses an effort level that balances the marginal benefit of effort (i.e., a higher probability of receiving the bonus b_t instead of δv^*_{t+1}) with the marginal effort costs. Higher bonuses then increase equilibrium efforts, as shown in Part *(iii)*. However, there is diminishing sensitivity of efforts towards higher bonuses, i.e., equilibrium efforts are concave in the bonus of the respective period.

Considering Lemma 3, the following remark introduces some notational simplifications that facilitate the exposition throughout the remainder of the paper.

Remark 2 (Notation)**.**

(i) *Exploiting the symmetry of efforts, we define $P(e_t) := p\left(e_t, \ldots, e_t\right)$.*[8]

(ii) *We henceforth write $e(\boldsymbol{b}_t)$ instead of $e(b_t, v^*_{t+1})$ because, formally, v^*_{t+1} depends on $\boldsymbol{b}_{t+1}$ only.*

(iii) *Inserting in Bellman Equation (2.1) allows us to define agent i's equilibrium expected payoff in period t as $v^*_t = v_t(\boldsymbol{b}_t) := v_{t,i}(e(\boldsymbol{b}_t), \ldots, e(\boldsymbol{b}_t))$.*

2.3.2 Equilibrium Bonus Design

Thus far, we have examined agents' equilibrium effort choices for a *given* bonus scheme. The following question arises naturally: how are these bonuses designed in equilibrium? To answer this question, note first that once the project is successful, the principal must use part of her revenue to pay the contractual bonuses to the n agents. Consequently, the principal's profit in case of success in period t amounts to $r_t - nb_t$. When designing the bonus scheme, the principal must also take into account that, if an agent is confronted with $\boldsymbol{b}_t$ in period t, he will implement the effort level $e^*_t = e(\boldsymbol{b}_t)$ in order to maximize his expected payoff, cf. Lemma 3. In light of this *incentive constraint*, the principal's expected profit generated from offering the bonus scheme $\boldsymbol{b}_t$ in period t is then recursively defined by[9]

$$\pi_t(\boldsymbol{b}_t) \;=\; P(e(\boldsymbol{b}_t))(r_t - nb_t) + \left(1 - P(e(\boldsymbol{b}_t))\right)\delta\pi_{t+1}(\boldsymbol{b}_{t+1}). \qquad (2.2)$$

8. Note that this definition directly implies that $\partial p(e_t, \ldots, e_t)/\partial e_{t,i} = P'(e_t)/n$.
9. Similarly to the agents' continuation payoff, recognize that $\pi_{t+1}(\boldsymbol{b}_{t+1}) = 0$ for all $t \geq T$.

We formalize the principal's decision problem next. To this end, we first examine the case where the principal offers *spot contracts*, that is, each bonus payment b_t is designed in the respective period t. Since all parties have correct expectations, the equilibrium bonus scheme $\boldsymbol{b}_1^* = (b_1^*, \dots, b_T^*)$ is then sequentially determined by backward induction, starting with b_T^* in period T. The principal's problem in period t consists in maximizing her expected profit in t over her bonus payment b_t, given the project's revenue profile $\boldsymbol{r}_t$ and the equilibrium bonus payments $\boldsymbol{b}_{t+1}^*$ in all subsequent periods,

$$\max_{b_t \in \mathbb{R}_+} \pi_t(b_t, \boldsymbol{b}_{t+1}^*) \;=\; P(e(b_t, \boldsymbol{b}_{t+1}^*))(r_t - nb_t) + \big(1 - P(e(b_t, \boldsymbol{b}_{t+1}^*))\big)\delta \pi_{t+1}(\boldsymbol{b}_{t+1}^*). \quad (2.3)$$

The first-order condition for Problem (2.3) is then

$$\begin{aligned} \frac{\mathrm{d}\pi_t(b_t, \boldsymbol{b}_{t+1}^*)}{\mathrm{d}b_t} \;=\; P'(e(b_t, \boldsymbol{b}_{t+1}^*)) \frac{\partial e(b_t, \boldsymbol{b}_{t+1}^*)}{\partial b_t} \big(r_t - nb_t - \delta \pi_{t+1}(\boldsymbol{b}_{t+1}^*)\big) \\ - nP(e(b_t, \boldsymbol{b}_{t+1}^*)) \;=\; 0. \end{aligned} \quad (2.4)$$

Before we establish the existence of a solution to Condition (2.4), some considerations are in order. In light of Lemma 3, we can conclude from (2.4) that π_t is weakly decreasing for all $b_t \geq 0$ if and only if $r_t - n\,\delta v_{t+1}^* - \delta \pi_{t+1}(\boldsymbol{b}_{t+1}^*) \leq 0$ or, equivalently,

$$r_t \;\leq\; \underline{r}_t \;:=\; \delta\big[n\, v_{t+1}(\boldsymbol{b}_{t+1}^*) + \pi_{t+1}(\boldsymbol{b}_{t+1}^*)\big].$$

The principal then provides no incentives to ensure that the game proceeds to period $t+1$, since success in t is not profitable enough.[10] Note that the revenue r_t can be interpreted as the aggregate surplus that the principal and her agents divided among each other in case of success in period t because effort costs are sunk when the uncertainty in t resolves. By contrast, $\underline{r}_t$ represents the discounted future surplus they can expect in case of failure in t. These considerations then lead to the following lemma.

Lemma 4. *There exists a unique equilibrium bonus $b_t^* = b_t(\boldsymbol{r}_t)$ defined by Condition (2.4), which exhibits the following properties:*

$$b_t(\boldsymbol{r}_t) \begin{cases} = 0 & \text{if } r_t \;\leq\; \underline{r}_t \\ > \;\delta v_{t+1}(\boldsymbol{b}_{t+1}^*) & \text{if } r_t \;>\; \underline{r}_t. \end{cases}$$

10. More precisely, the principal is then indifferent between all b_t that implement zero effort, i.e., that satisfy $0 \leq b_t \leq \delta v_{t+1}^*$. We assume that, in this special case, she chooses the lowest possible bonus $b_t = 0$.

Intuitively, the principal only sets incentives in period t, i.e., implements a positive bonus, if the revenue r_t is high enough to make all parties better off than in $t+1$. Otherwise, no bonus is offered in order to ensure continuation of the game. An implication of Lemma 4 is that we can define the equilibrium bonus scheme contingent on the project's revenue profile by setting

$$\boldsymbol{b}_t^* = \boldsymbol{b}_t(\boldsymbol{r}_t) := \big(b_t(\boldsymbol{r}_t), \ldots, b_T(\boldsymbol{r}_T)\big).$$

This definition, together with Lemma 3, enables us to link the project's revenue to the effort provision in period t:

$$e(\boldsymbol{b}_t(\boldsymbol{r}_t)) \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \\ > 0 & \text{if } r_t > \underline{r}_t, \end{cases} \quad \text{and} \quad \frac{\partial e(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\partial b_t} \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \\ > 0 & \text{if } r_t > \underline{r}_t. \end{cases} \tag{2.5}$$

The implemented effort is thus strictly positive and increasing in the bonus if the revenue exceeds the threshold $\underline{r}_t$, and zero otherwise.

To complete this section, we introduce the *success rate* S_t^* measuring the equilibrium probability that the project will be successful if the game has progressed to period t. Formally, the success rate is defined in a recursive manner by[11]

$$S_t^* = S_t(\boldsymbol{b}_t^*) = S_t(\boldsymbol{b}_t(\boldsymbol{r}_t)) := P(e(\boldsymbol{b}_t(\boldsymbol{r}_t))) + \Big(1 - P(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))\Big) S_{t+1}(\boldsymbol{b}_{t+1}(\boldsymbol{r}_{t+1})). \tag{2.6}$$

The probability S_1^* then measures the likelihood of successful project completion during the entire game.

2.4 Revenue Effects

The previous section has demonstrated that the project's revenue profile is pivotal for the design of bonus schemes. These revenues capture the economic environment and will generally differ over time, $r_1 \neq r_2, \ldots, \neq r_T$. Project completion in a boom phase, for instance, can be substantially more lucrative than completion during a recession. More importantly, the profitability of the project can also change within a certain period. A change in the project's revenue will then have a threefold impact, namely, *(i)* on agents' expected payoffs, *(ii)* on the principal's expected profit, as well as *(iii)* on the project's success rate. In the dynamic version of our principal-agent model ($T > 1$), this impact will manifest in both *intraperiod* and *interperiod* effects. The former are considered first.

11. To be technically precise, note that $S_{t+1}(\boldsymbol{b}_{t+1}(\boldsymbol{r}_{t+1})) = 0$ for all $t \geq T$.

2.4.1 Intraperiod Effects

All effects triggered by higher revenues originate in the relationship between project revenues and implemented bonus payments. For this reason, we investigate how a change in the revenue r_t affects the equilibrium bonus b_t^* in the lemma below.

Lemma 5. *The equilibrium bonus b_t^* in period t satisfies*

$$\frac{\partial b_t(\boldsymbol{r}_t)}{\partial r_t} \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \\ > 0 & \text{if } r_t > \underline{r}_t. \end{cases}$$

It now follows from (2.5) and Lemma 5 that agents' effort provision in period t is weakly increasing in the revenue of that period,

$$\frac{\partial e(b_t(\boldsymbol{r}_t))}{\partial r_t} = \frac{\partial e(b_t(\boldsymbol{r}_t))}{\partial b_t} \frac{\partial b_t(\boldsymbol{r}_t)}{\partial r_t} \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \\ > 0 & \text{if } r_t > \underline{r}_t. \end{cases} \tag{2.7}$$

A higher revenue in period t causes an intraperiod *motivation effect*, i.e., it leads to weakly higher bonuses that in turn incentivize weakly higher efforts. Taking an agent's perspective, we shed light on how his expected payoff is affected by the project's revenues. Invoking the Envelope Theorem, we obtain from (2.1) that

$$\begin{aligned} \frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} &= P(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))\frac{\partial b_t(\boldsymbol{r}_t)}{\partial r_t} \\ &+ \Big(b_t(\boldsymbol{r}_t) - \delta v_{t+1}(\boldsymbol{b}_{t+1}(\boldsymbol{r}_{t+1}))\Big)\frac{n-1}{n}P'(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))\frac{\partial e(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\partial b_t}\frac{\partial b_t(\boldsymbol{r}_t)}{\partial r_t}. \end{aligned} \tag{2.8}$$

Considering (2.7) and Lemmas 3-5, we can infer from (2.8) that

$$\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \\ > 0 & \text{if } r_t > \underline{r}_t. \end{cases} \tag{2.9}$$

Hence, increased profitability in period t has a weakly positive intraperiod effect on the agent's expected payoff. This result is intuitive: with higher revenues, it is more lucrative for the principal to implement higher efforts. To do so, she sets higher bonuses, which benefits the agents.

Considering the other part of the principal-agent relationship, we now explore the effect of a higher revenue r_t on the principal's expected profit. Applying the Envelope Theorem and using (2.5), it follows that

$$\frac{\mathrm{d}\pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} = P(e(\boldsymbol{b}_t(\boldsymbol{r}_t)) \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \\ > 0 & \text{if } r_t > \underline{r}_t. \end{cases} \tag{2.10}$$

From an intraperiod perspective, both contracting parties thus at least weakly benefit if the project becomes more lucrative. Additionally, the intraperiod motivation effect directly enhances the project's success rate in period t. Formally, by differentiation of (2.6), we get

$$\frac{\mathrm{d}S_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} = P'(e(b_t(\boldsymbol{r}_t)))\,\frac{\partial e(b_t(\boldsymbol{r}_t))}{\partial r_t}\,\frac{\partial b_t(\boldsymbol{r}_t)}{\partial r_t}\left(1 - S_{t+1}(\boldsymbol{b}_{t+1}(\boldsymbol{r}_{t+1}))\right) \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \\ > 0 & \text{if } r_t > \underline{r}_t. \end{cases}$$

We summarize these findings in the following proposition.

Proposition 4 (Intraperiod Effects).
If the revenue profile $\boldsymbol{r}_t$ satisfies $r_t > \underline{r}_t$, then an increase in the revenue r_t strictly increases the (i) bonus b_t^*, (ii) effort provision e_t^*, (iii) agents' expected payoffs v_t^*, (iv) principal's expected profit π_t^*, and (v) success rate S_t^*. If, on the contrary, $\boldsymbol{r}_t$ satisfies $r_t \leq \underline{r}_t$, then an increase in r_t has no effect in period t at all.

As we will show next, these beneficial intraperiod effects can, however, trigger adverse intertemporal effects in other periods that may prevail.

2.4.2 Interperiod Effects

While a change in revenue leaves future periods unaffected, intertemporal effects may arise in preceding periods. To grasp the underlying mechanisms, it is sufficient to analyze these effects between an arbitrary period t and its predecessor $t-1$. Since the interperiod effects are mediated by agents' effort provision and the principal's bonus design, we first examine how revenues affect efforts and bonuses.

Effort Provision and Bonuses. The next proposition describes how higher revenues in period t change equilibrium efforts and bonuses in $t-1$.

Proposition 5 (Interperiod Effort and Bonus Effects).
Let $T \geq t > 1$. Then the following holds.

(i) *Equilibrium efforts e^*_{t-1} and bonuses b^*_{t-1} in period $t-1$ satisfy*

$$\frac{\mathrm{d}e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \text{ or } r_{t-1} \leq \underline{r}_{t-1} \\ < 0 & \text{if } r_t > \underline{r}_t \text{ and } r_{t-1} > \underline{r}_{t-1}, \end{cases}$$

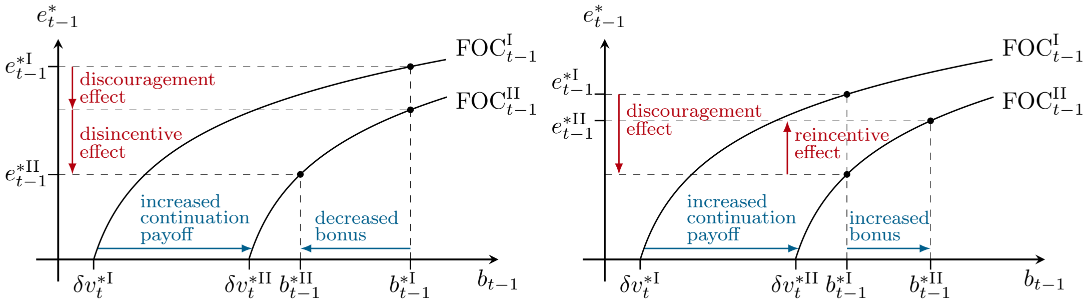
$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \text{ or } r_{t-1} \leq \underline{r}_{t-1} \\ \gtreqless 0 & \text{if } r_t > \underline{r}_t \text{ and } r_{t-1} > \underline{r}_{t-1}, \end{cases}$$

(ii) *In case $r_t > \underline{r}_t$ and $r_{t-1} > \underline{r}_{t-1}$, it holds that $\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})/\mathrm{d}r_t < 0$ if $r_{t-1} - \underline{r}_{t-1}$ is sufficiently small, while the converse holds if $r_{t-1} - \underline{r}_{t-1}$ is sufficiently large.*

Proposition 5 (i) states that intertemporal effects on efforts and bonuses only arise for $r_t > \underline{r}_t$ and $r_{t-1} > \underline{r}_{t-1}$, i.e., if efforts are positive in both periods. In that case, a higher revenue always leads to a lower effort provision in the preceding period. Intuitively, this holds true since a higher revenue in period t makes success in that period more attractive relative to success in preceding period. Consequently, the principal incentivizes lower efforts in $t-1$ in order to obtain a higher probability of reaching the relatively more attractive period t.

This reduction of efforts is the result of two (potentially countervailing) mechanisms. First, there is an intertemporal *discouragement effect*: a higher revenue in t increases the bonus the principal sets in t, which in turn increases the agents' continuation payoffs in the preceding period. Investing effort in period $t-1$ thus becomes less attractive for agents. A second mechanism arises since the principal anticipates the discouragement effect and, in general, adjusts the bonus b^*_{t-1}. Proposition 5 (ii) shows that the principal lowers the $t-1$ bonus if success in that period is rather unprofitable, i.e., if the revenue is rather low. We call this the *disincentive effect*. By contrast, the principal counteracts the discouragement effect by increasing the $t-1$ bonus if success in that period is sufficiently lucrative, i.e., if the revenue is rather high. We call this the *reincentive effect*.

Recognize that the discouragement effect and the disincentive effect work in the same direction, since both reduce agents' efforts in period $t-1$. By contrast, the reincentive effect counteracts the discouragement effect. Yet, as shown by Proposition 5 (i), the discouragement always dominates the reincentive effect. The reincentive effect hence only partly compensates the discouragement effect. Figures 2.1a and 2.1b depict these effects.



(a) Disincentive effect

(b) Reincentive effect

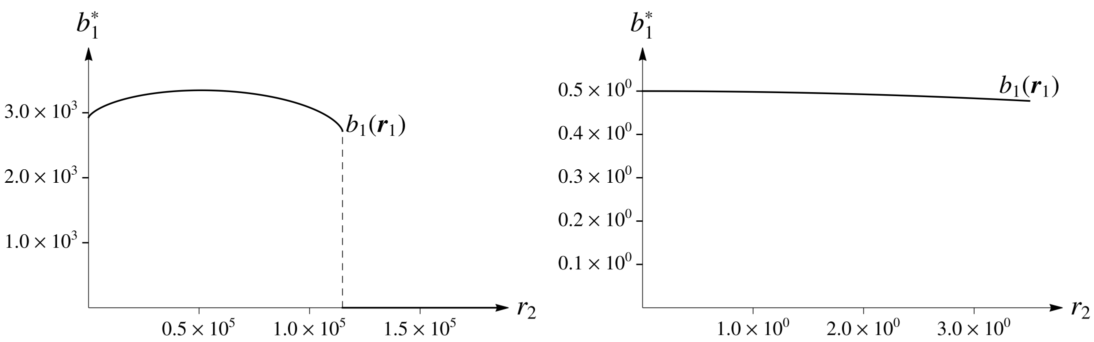
Figure 2.1. Interperiod effects on efforts in $t-1$ induced by an increased revenue in t.

We illustrate our findings on how higher revenues affect the equilibrium bonus design in the previous period with numerical examples.

Example 2 (Interperiod Effects on Bonuses)**.**
Consider the two-period model $(T=2)$ *with quadratic costs* $c(e_{t,i}) = (e_{t,i})^2$*. We demonstrate our results with two alternative success functions, namely, a non-linear and a linear one:*

$$P_N(e_t) = 1 - \frac{1}{1+ne_t}, \qquad \textit{and} \qquad P_L(e_t) = \begin{cases} ne_t & \text{if } e_t \le 1/n \\ 1 & \text{if } e_t > 1/n. \end{cases}$$

Figure 2.2a displays the non-linear case and shows that the reincentive effect arises if the revenue r_2 *is small (and the difference* $r_1 - \underline{r}_1 > 0$ *is thus large), while the disincentive effect arises for larger revenues* r_2*. Figure 2.2b displays the linear case. For the presented parameterization, which will be used in the examples throughout the paper, the difference* $r_1 - \underline{r}_1$ *is too low for the reincentive effect to arise such that the disincentive effect occurs for all revenues* r_2*.*



(a) Non-linear success function ($n = 1$, $r_1 = 10^5$, and $\delta = 0.95$)

(b) Linear success function ($n = 4$, $r_1 = 4$, and $\delta = 0.95$)

Figure 2.2. Interperiod effects on the bonus in period 1.

Expected Payoff. Against the background of the interperiod effects on equilibrium efforts and bonuses revealed above, we now raise the question: do agents inevitably benefit from an economically more favorable environment? We investigate how an increase in revenue changes agents' expected payoffs in the previous period. By the Envelope Theorem, the derivative of the expected payoff in period $t-1$ computes

$$\begin{aligned}\frac{\mathrm{d}v_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} = &\overbrace{P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t}}^{\gtreqless 0} + \overbrace{\big(1-P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))\big)\,\delta\,\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}}^{\geq 0} \\ &+ \underbrace{\Big(b_{t-1}(\boldsymbol{r}_{t-1}) - \delta v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))\Big)\frac{n-1}{n}P'(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\,\frac{\mathrm{d}e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t}}_{\leq 0}.\end{aligned} \tag{2.11}$$

The first summand on the r.h.s. of (2.11) results from the optimizing behavior of the principal. It captures the disincentive (reincentive) effect that harms (benefits) the agent's expected payoff, cf. Proposition 5 (ii). The second summand captures the intraperiod effect in t: for a higher revenue r_t, the principal offers higher bonus payments b_t^*, which increase the (discounted) continuation payoff δv_t^*. Finally, the third expression results from the *team externality*: since it is optimal for the other team members to reduce their efforts in $t-1$, the likelihood of receiving the bonus payment b_{t-1}^* decreases.

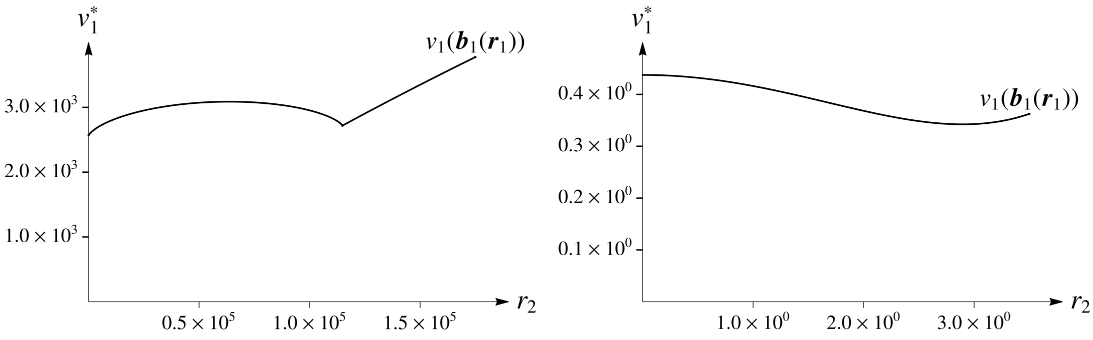
Henceforth, we say that a *payoff reversal* occurs in period $t-1$ if a higher revenue in t negatively affects the agents' expected payoffs in $t-1$:

$$\frac{\mathrm{d}v_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} < 0.$$

From (2.11), we can conclude that a payoff reversal can be induced by two mechanisms: a strong disincentive effect or a strong team externality. The latter occurs exclusively in team settings and requires that the reincentive effect is either absent or sufficiently weak. Payoff reversals due to the disincentive effect, however, may also arise in a single-agent setting. We provide a brief summary of these insights in the proposition below.

Proposition 6 (Payoff Reversal)**.**
Let $T \geq t > 1$. Then a payoff reversal in $t-1$ can occur due to the disincentive effect or due to the team externality.

Example 3 (Payoff Reversal).
Revisit the parameterization from Example 2. Figure 2.3a displays how the disincentive effect causes a payoff reversal in a single-agent setting, which arises in the left local neighborhood of the kink where $r_1 = \underline{r}_1$. Figure 2.3b shows that, in the linear setting, the expected payoff is locally decreasing in the second-period revenue r_2 due to the team externality.



(a) Non-linear success function ($n = 1$, $r_1 = 10^5$, and $\delta = 0.95$)

(b) Linear success function ($n = 4$, $r_1 = 4$, and $\delta = 0.95$)

Figure 2.3. Payoff reversals in period 1.

Expected Profit. Directing our attention to the contract-designing party, we next examine how the principal herself is affected by an increase in the project's revenue. The relationship between the expected profit in period $t-1$ and the revenue r_t is formally specified in the following lemma.

Lemma 6. *Let $T \geq t > 1$. Then the principal's expected profit satisfies*

$$\frac{\mathrm{d}\pi_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} = \overbrace{\delta\Big(1 - P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big)P(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))}^{\geq 0} \underbrace{-\delta\, n\, P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\,\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}}_{\leq 0}. \tag{2.12}$$

From (2.12), it is evident that the expected profit in $t-1$ is subject to two opposing effects. While higher revenues have a direct positive effect (cf. first line on the r.h.s. of (2.12)), setting incentives in $t-1$ becomes more costly for the principal owing to the agents' increased continuation payoff (cf. second line on the r.h.s. of (2.12)). Put differently, we know from Proposition 5 that, since agents correctly anticipate increased bonuses in t, they lower their efforts in the previous period, which negatively affect the principal's expected profit. If $P(e^*_{t-1})$ is sufficiently close to one, that is, if success is sufficiently likely in period $t-1$, then the latter effect dominates, and a *profit reversal* occurs. More precisely, we say that a profit reversal occurs in period $t-1$ if a higher revenue in period t negatively affects the principal's expected profit in period $t-1$:

$$\frac{\mathrm{d}\pi_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} < 0.$$

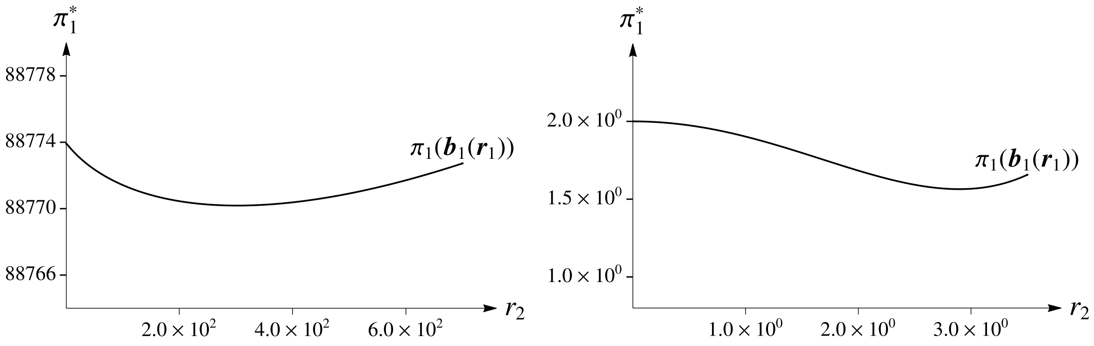
Theorem 3, which is a central result of this article, states a sufficient condition for the occurrence of such a profit reversal.

Theorem 3 (Profit Reversal).
Let $T \geq t > 1$ and suppose that $\boldsymbol{r}_{t-1}$ satisfies $r_t > \underline{r}_t$ and $r_{t-1} > \underline{r}_{t-1}$. Then a profit reversal occurs in $t-1$ if $r_{t-1} - \underline{r}_{t-1}$ is sufficiently large.

The intuition is as follows. If $r_{t-1} - \underline{r}_{t-1}$ is large, success in period $t-1$ is rather attractive for the principal. An increase in the revenue r_t causes the principal to implement a higher bonus in period t, which in turn lowers agents' efforts in $t-1$, cf. Propositions 4 and 5. Accordingly, success in the highly lucrative period $t-1$ becomes less likely, which ultimately harms the principal's expected profit.

The possibility of profit reversals is a novel and surprising finding. Although the principal has all the negotiation power and designs the bonuses in order to maximize her expected profit, she may actually suffer from a more lucrative economic environment. This intriguing phenomenon results from the sequential structure of the principal's decision problem under spot contracts (and, as will be shown later, persists when long-term contracts need to be renegotiation-proof). Since bonus payments are designed in the respective period, the principal cannot credibly account for negative intertemporal effects of higher bonus payments in earlier periods. Indeed, by the time the principal finds herself in the position to design the bonus, all previous periods are irrelevant. Agents, however, anticipate possibly high future bonuses via their continuation payoff and choose their efforts accordingly. A more profitable revenue profile can hence induce a less desirable intertemporal allocation of efforts and thereby harm the principal's expected profit.

Example 4 (Profit Reversal).
Revisit the parameterizations from Example 2. Figure 2.4 shows that if the revenue r_2 *is sufficiently small, i.e., if* $r_1 - \underline{r}_1$ *is sufficiently large, then a profit reversal emerges in period 1.*



(a) Non-linear success function ($n = 1$, $r_1 = 10^5$, and $\delta = 0.95$)

(b) Linear success function ($n = 4$, $r_1 = 4$, and $\delta = 0.95$)

Figure 2.4. Profit reversals in period 1.

Success Rate. Finally, we investigate how the project's revenue profile is linked to its probability of success. For the intraperiod effects, we have already established a positive relationship between r_t and the success rate S_t^*. However, the interperiod mechanisms explored in Proposition 5 may adversely affect the success rate S_{t-1}^* in the previous period. Formally, we say that a *success reversal* occurs in period $t-1$ if a higher revenue in period t negatively affects the project's success rate in period $t-1$, i.e., if

$$\frac{\mathrm{d}S_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} < 0.$$

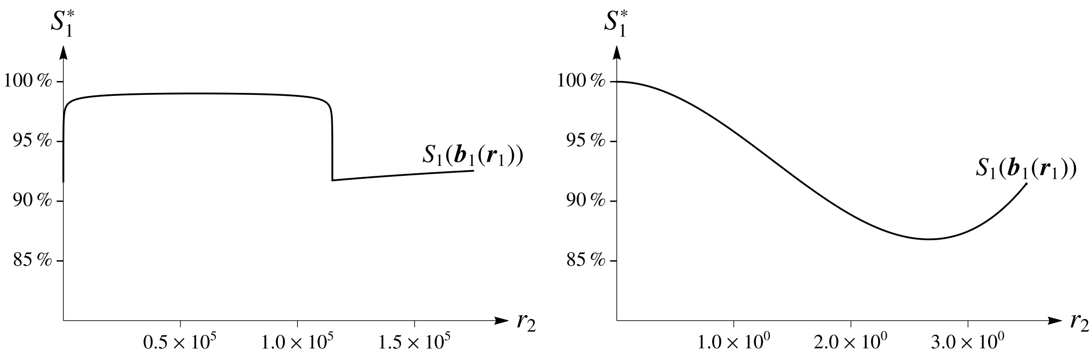
To comprehend the mechanisms underlying a success reversal, we state the following proposition.

Proposition 7 (Success Reversal).
Let $T \geq t > 1$. *A success reversal occurs if and only if*

$$\frac{\Big(1 - P(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))\Big)P'(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))}{\Big(1 - P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big)P'(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))} \cdot \frac{-\dfrac{\mathrm{d}e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t}}{\dfrac{\partial e(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\partial b_t}\dfrac{\partial \boldsymbol{b}_t(\boldsymbol{r}_t)}{\partial r_t}} > 1. \tag{2.13}$$

From Condition (2.13), it becomes apparent that success reversals can be induced by two mechanisms. First, a success reversal due to the *substitution of efforts* arises if the curvature of P is rather low. A success reversal occurs in this case for revenues that implement relatively high success probability in period t since the first factor on the l.h.s. of (2.13) becomes large. Intuitively speaking, a higher revenue r_t then causes that effort is substituted away from a period with a high probability of success $(t-1)$ to one that is unlikely to be reached (t). Second, a success reversal due to the *sensitivity of efforts* arises when an increase in the revenue r_t causes a strong decline in the implemented effort e^*_{t-1}, but only a slight increase in e^*_t such that the second factor in Condition (2.13) is large.

Example 5 (Success Reversal).
Revisit the parameterizations from Example 2. Figure 2.5a displays a success reversal due to the sensitivity of efforts, while Figure 2.5b depicts a success reversal due to the substitution of efforts.



(a) Non-linear success function ($n = 1$, $r_1 = 10^5$, and $\delta = 0.95$)

(b) Linear success function ($n = 4$, $r_1 = 4$, and $\delta = 0.95$)

Figure 2.5. Success reversals in period 1.

Occurrence of Reversals. The three reversals that we have discovered are intriguing phenomena. But are they rather rare anomalies or phenomena that occur regularly? In general, it depends on the success and cost function which parameter constellations give rise to payoff, profit, and success reversals. The following example illustrates that all three reversals can be relevant for a wide variety of parameterizations.

Example 6 (Occurrence of Reversals).
Revisit the parameterization from Example 2 with a linear success function. Figure 2.6 displays the combinations of first-period revenues and discount factors that lead to the different forms of reversals for some second-period revenues.

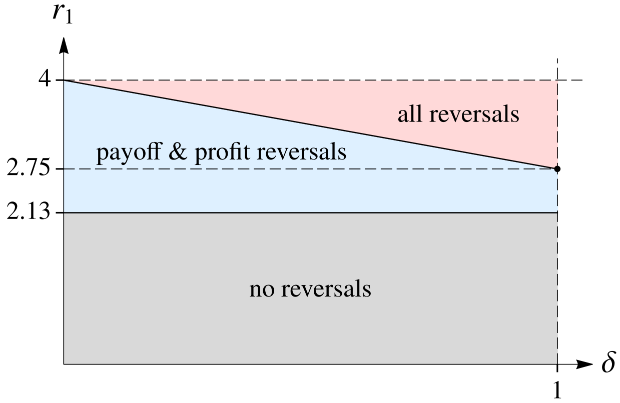


Figure 2.6. Parameterizations leading to reversals in period 1 (n=4).

2.4.3 Static vs. Dynamic Interaction

The decisive measures for the contracting parties are the expected payoff v_1^* (agents) and the expected profit π_1^* (principal), as they capture the *entire game* due to their recursive definition. Additionally, from a project-management perspective, the overall success rate S_1^* is an important indicator. To compare the effect of different revenue profiles on those measures, we introduce the notion of a *dominant revenue profile*.

Definition 4 (Dominant Revenue Profile).
The revenue profile $\boldsymbol{r}_1^{\mathrm{A}} = \left(r_1^{\mathrm{A}}, \dots, r_T^{\mathrm{A}}\right)$ *dominates the revenue profile* $\boldsymbol{r}_1^{\mathrm{B}} = \left(r_1^{\mathrm{B}}, \dots, r_T^{\mathrm{B}}\right)$ *if* $r_t^{\mathrm{A}} \geq r_t^{\mathrm{B}}$ *holds in all periods* $t \in \{1, \dots, T\}$ *and* $r_t^{\mathrm{A}} > r_t^{\mathrm{B}}$ *holds in at least one period* t.

For the static model, $T = 1$, we can immediately conclude that a dominant revenue profile induces a more preferable outcome for the contracting parties and the project itself. Formally, since $\underline{r}_1 = 0$, we can deduce from Proposition 4 that a higher revenue r_1 strictly increases v_1^*, π_1^*, as well as S_1^*.[12] This finding is straightforward because the detrimental interperiod effects discussed above are naturally

12. To see this, recognize that, in the static model, $r_{t+1} = 0$, $v_{t+1}^* = 0$, $\pi_{t+1}^* = 0$, and $S_{t+1}^* = 0$ for all $t \geq 1$.

absent in the static model. The conventional wisdom thus applies that a more lucrative environment is necessarily beneficial for the contracting parties and the probability of project success.

However, with dynamic interaction, $T > 1$, the interperiod effects come into play. The preceding analysis has demonstrated that an increase in revenues ambiguously affects the three measures under scrutiny. Indeed, a higher revenue in any period $t > 1$ may well decrease the principal's expected profit π_1^*, the agents' expected payoffs v_1^*, and the success rate S_1^*.[13] When comparing a dominant revenue profile $\boldsymbol{r}_1^{\mathrm{A}}$ with a dominated revenue profile $\boldsymbol{r}_1^{\mathrm{B}}$, we can thus conclude that $\boldsymbol{r}_1^{\mathrm{A}}$ may induce a *less favorable* outcome for all parties, i.e.,

$$v_1(\boldsymbol{b}_1(\boldsymbol{r}_1^{\mathrm{A}})) < v_1(\boldsymbol{b}_1(\boldsymbol{r}_1^{\mathrm{B}})), \quad \pi_1(\boldsymbol{b}_1(\boldsymbol{r}_1^{\mathrm{A}})) < \pi_1(\boldsymbol{b}_1(\boldsymbol{r}_1^{\mathrm{B}})), \quad S_1(\boldsymbol{b}_1(\boldsymbol{r}_1^{\mathrm{A}})) < S_1(\boldsymbol{b}_1(\boldsymbol{r}_1^{\mathrm{B}})).$$

This observation highlights the detrimental effects that may arise if the principal relies on spot contracts as an incentive mechanism. Although the principal faces a *more lucrative* economic environment, the resulting incentive mechanism may reduce project performance and leave all stakeholders worse off.

Example 7 (Dominant vs. Dominated Revenue Profiles).
Revisit the parameterization from Example 2 with a linear success function. Table 2.1 shows that the dominant revenue profile $\mathbf{r}_1^{\mathrm{A}}$ *leads to worse outcomes for both contracting parties and a lower probability of success than the dominated revenue profile* $\mathbf{r}_1^{\mathrm{B}}$.

Table 2.1. Comparison of revenue profiles ($n = 4$, $\delta = 0.95$).

$\boldsymbol{r}_1$	$\boldsymbol{b}_1^*$	v_1^*	π_1^*	S_1^*
$\boldsymbol{r}_1^{\mathrm{A}} = (4, 3)$	(0.48, 0.38)	0.34	1.57	87%
$\boldsymbol{r}_1^{\mathrm{B}} = (3.8, 1)$	(0.47, 0.13)	0.38	1.71	92%

However, apart from spot contracts, there are other contract classes that the principal can deploy as an incentive mechanism for her agents. The following section therefore examines the extent to which our previous results generalize to long-term contracts.

13. For an increase in r_2, this observation directly follows from our analysis above. However, due to the recursive definition of π_1^*, v_1^*, and S_1^*, payoff, profit, and success reversals from later periods can also have a detrimental effect up to the first period.

2.5 Long-Term Contracts

In many principal-agent relationships, long-term contracts constitute an alternative to short-term contracting via a sequence of spot contracts. A firm might, for example, conclude strategic long-term contracts with its suppliers. In the context of our model, we define a long-term contract as follows.

Definition 5 (Long-Term Contract)**.**
A long-term contract $\boldsymbol{b}_1 \in \mathbb{R}_+^{\mathrm{T}}$ *is a contract that stipulates all T bonus payments* $(b_1, \ldots, b_T)$ *at the beginning of the initial period* $t = 1$.

To investigate whether the detrimental effects of incentive mechanisms also emerge when the contracting parties deploy such long-term contracts, we must distinguish two cases according to the principal's commitment power.

Full Commitment. We first concentrate on settings where the principal has *full* commitment power, i.e., she can exclude any form of renegotiation in later periods, even if she would then benefit from such renegotiations. Given the revenue profile $\boldsymbol{r}_1$, the principal designs bonuses to maximize her expected profit from hiring the agents. Formally, her equilibrium long-term contract with full commitment $\boldsymbol{b}_1^{\mathrm{F}*} = (b_1^{\mathrm{F}*}, \ldots, b_T^{\mathrm{F}*})$ is defined by the solution to the T-dimensional optimization problem

$$\boldsymbol{b}_1^{\mathrm{F}}(\boldsymbol{r}_1) \; := \; \boldsymbol{b}_1^{\mathrm{F}*} \; = \; \underset{\boldsymbol{b}_1 \in \mathbb{R}_+^{\mathrm{T}}}{\operatorname{argmax}} \; \pi_1(\boldsymbol{b}_1).$$

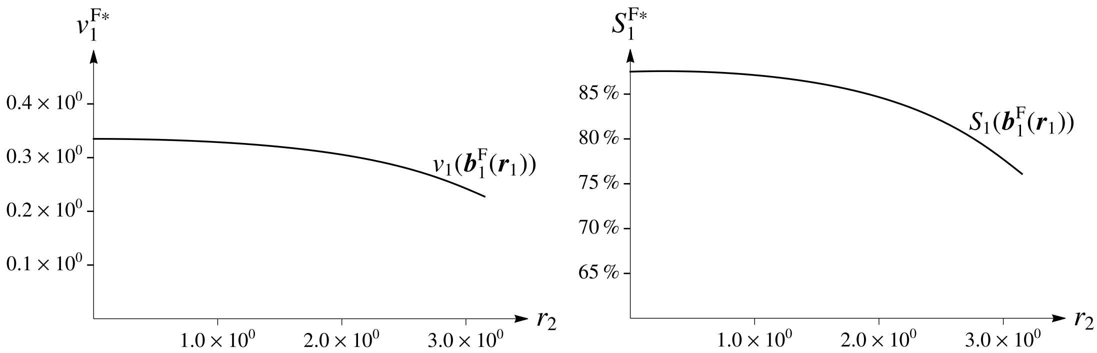
If the principal can fully commit to a bonus scheme, then a revenue profile $\boldsymbol{r}_1^{\mathrm{A}}$ that dominates the revenue profile $\boldsymbol{r}_1^{\mathrm{B}}$ generates at least a weakly higher expected profit, i.e., $\pi_1(\boldsymbol{b}_1^{\mathrm{F}}(\boldsymbol{r}_1^{\mathrm{A}})) \geq \pi_1(\boldsymbol{b}_1^{\mathrm{F}}(\boldsymbol{r}_1^{\mathrm{B}}))$.[14] The reason is straightforward: when faced with the dominant revenue profile $\boldsymbol{r}_1^{\mathrm{A}}$, the principal can design the same bonus scheme as when confronted with the dominated revenue profile $\boldsymbol{r}_1^{\mathrm{B}}$. She can thereby maintain the same efforts and thus success probabilities, but, at the same time, earn higher proceeds in the case of success. Hence, a profit reversal in $\pi_1(\boldsymbol{b}_1^{\mathrm{F}}(\boldsymbol{r}_1))$, the measure that captures the entire game, cannot arise with long-term contracts under full commitment. The principal is thus inevitably better off in an economically more favorable environment.[15]

14. The inequality is strict if $\boldsymbol{b}_1^{\mathrm{F}}(\boldsymbol{r}_1^{\mathrm{A}}))$ incentivizes a positive probability of success in at least one period where $r_t^{\mathrm{A}} > r_t^{\mathrm{B}}$ holds.

15. To be technically precise, the total differential of $\pi_1(\boldsymbol{b}_1^{\mathrm{F}}(\boldsymbol{r}_1))$ with respect to r_t is at least weakly positive for any $t \in \{1, \ldots, T\}$.

However, the same does not necessarily hold true for the agents' expected payoffs and the project's success rate. Since the principal solely focuses on maximizing *her* expected profit when determining the bonus payments, higher revenues may engender a design of long-term contracts that makes success less likely and harms the agents' expected payoffs. As shown in Example 8, success and payoff reversal also occur if a principal equipped with full commitment power sets incentives through long-term contracts.

Example 8 (Reversals Under Full Commitment)**.**
Revisit the parameterization from Example 2 with a linear success function. Let the principal have full commitment power and fix the revenue in period 1 to $r_1 = 3.5$. *Higher revenues in period 2 then lead to a contract design that reduces agents' expected payoffs (cf. Figure 2.7a) and the likelihood of project success (cf. Figure 2.7b).*



(a) Payoff Reversal **(b)** Success Reversal

Figure 2.7. Success and payoff reversals with long-term contracts.
($n = 4$, $r_1 = 3.5$, and $\delta = 0.95$)

Limited Commitment. We next turn our attention to the second case where, due to *limited* commitment power, the principal cannot rule out renegotiating the long-term contract in later periods. This case is highly relevant, since parties are usually able to change an existing contract if this benefits all parties involved. Following the renegotiation-proofness principle (Bolton, 1990; Dewatripont, 1989; Fudenberg & Tirole, 1990), we can, without loss of generality, restrict our attention to *renegotiation-proof long-term contracts* in the case of limited commitment. We use the following well-established definition of renegotiation-proof long-term contracts in finite-horizon games with discrete time (see, for example, Benoit and Krishna, 1993; Wang, 2000; Zhao, 2006).

Definition 6 (Renegotiation-Proofness).
The notion of renegotiation-proofness is defined recursively.

(i) A one-period contract is called renegotiation-proof if it is Pareto-efficient among all feasible and incentive-compatible one-period contracts.
(ii) A two-period contract is then called renegotiation-proof if it is Pareto-efficient among all feasible and incentive-compatible two-period contracts whose one-period continuations are renegotiation-proof, and so on.

One might now think that the principal can design long-term contracts that exploit payoff reversals or, more generally, specify rather low future bonuses in order to motivate agents to invest high efforts in early periods. However, the next lemma shows that the agents' payoffs as well as the principal's profits are rather restricted if they are generated by a renegotiation-proof long-term contract $\boldsymbol{b}_1^{\mathrm{L}} = (b_1^{\mathrm{L}}, \dots, b_T^{\mathrm{L}})$.

Lemma 7. *Let the sequence of equilibrium spot contracts $\boldsymbol{b}_1^*$ be given.*

(i) Any long-term contract $\boldsymbol{b}_1^{\mathrm{L}} = (b_1^{\mathrm{L}}, \dots, b_T^{\mathrm{L}})$ that is renegotiation-proof in the sense of Definition 6 satisfies

$$v_t(\boldsymbol{b}_t^{\mathrm{L}}) \geq v_t(\boldsymbol{b}_t^*), \qquad \text{and} \qquad \pi_t(\boldsymbol{b}_t^{\mathrm{L}}) \leq \pi_t(\boldsymbol{b}_t^*) \quad \text{for all } t \in \{1, \dots, T\}.$$

(ii) The bonus scheme $\boldsymbol{b}_1^*$ is a renegotiation-proof long-term contract in the sense of Definition 6.

Part *(i)* reveals that the payoffs obtained by agents under spot contracts create a lower bound for the payoffs that can be implemented by any renegotiation-proof contract. This, in particular, implies that it is not possible to motivate agents cheaply by specifying future bonuses that yield relatively low continuation payoffs. The lemma also reveals that the profits that the principal can obtain under spot contracts form an upper bound for the profits that can be implemented by any renegotiation-proof contract. Part *(ii)* now confirms that the principal can write a renegotiation-proof contract that actually yields those profits from the upper bound. Indeed, the long-term contract comprising the corresponding spot bonuses is renegotiation-proof.

Using Lemma 7, we can now state the central result of this section, namely, which long-term contract a principal with limited commitment offers in equilibrium.[16]

16. A large debt of gratitude is due to Professor Weinschenk for numerous discussions that led to Lemma 7 and Theorem 4.

Theorem 4 (Equilibrium Renegotiation-Proof Long-Term Contract)**.**
The sequence of equilibrium spot contracts constitutes the principal's unique equilibrium renegotiation-proof long-term contract, i.e., $\boldsymbol{b}_1^{L*} = \boldsymbol{b}_1^* = (b_1^*, \dots, b_T^*)$.

Theorem 4 states that, if the principal has limited commitment power, then she offers the contract $\boldsymbol{b}_1^*$, since, among all renegotiation-proof long-term contracts, $\boldsymbol{b}_1^*$ generates the highest expect profit in period 1. Consequently, the bonus payments stipulated by a single long-term contract coincide with those implemented by a sequence of T spot contracts. Our insights derived in the analysis of spot contracts therefore fully carry over to the class of renegotiation-proof long-term contracts. In other words, the surprising phenomena of payoff, profit, and success reversals that can harm all parties involved may occur just as well with such long-term contracts as with spot contracts.

2.6 Conclusion

This article has examined a fairly general dynamic model of project completion in the presence of moral hazard. Our analysis demonstrates that a change in the project's returns triggers intraperiod as well as interperiod effects. While the intraperiod effects are intuitive, in the sense that an increase in the project's revenue makes all parties better off, the interperiod effects are surprising. Increased revenues may not only lead to a design of incentives that makes overall success less likely, and reduce the agents' expected payoffs, but may, in fact, be detrimental to the principal herself. These success, payoff, and profit reversals obtain for a wide variety of conceivable contracting situations, since our results apply to both short-term spot contracts as well as renegotiation-proof long-term contracts. Their root cause is that the implemented incentive mechanisms fail to account for the adverse interperiod effect that higher incentives have on earlier periods. Indeed, agents' effort provision is not only determined by their current bonuses for success, but also by their anticipation of future incentives. Even a principal who can take these interdependencies perfectly into account – by writing long-term contracts under full commitment – may design incentive mechanisms that negatively impact the agents' payoffs and the probability of project success. Nonetheless, full commitment power constitutes a remedy for profit reversals.

Our paper reveals highly relevant, yet undiscovered, deleterious effects of incentive mechanisms that arise in equilibrium when a principal designs optimal contracts. These findings imply that the lucrativeness of a project is not necessarily a decisive measure. In dynamic contexts, the common wisdom fails that the more lucrative a project, the better it is for its stakeholders, and the more likely it is to succeed. This article thus provides a theoretical foundation for the frequently encountered phenomenon of failures and delays in lucrative large-scale projects reported, for instance, in *public infrastructure* (e.g., Flyvbjerg et al., 2003; Mittal et al., 2020; Steininger et al., 2021), *research and development* (e.g., Gupta and Wilemon, 1990; Lhuillery and Pfister, 2009; Radas and Bozic, 2012), as well as *IT projects* (e.g., Al-Ahmad et al., 2009; Brown et al., 2007; Whitney and Daniels, 2013).[17] To alleviate such problems, our results imply that more attention should be devoted to the intertemporal distribution of project revenues instead of their sheer size.

Clearly, the model does not come without limitations. While our assumptions on the cost function are standard, the symmetry of both agents and the success function is somewhat limiting. It is, however, necessary to keep the analytical complexity manageable. This case of asymmetric agents (e.g., skilled vs. unskilled workers or, alternatively, team leaders vs. regular employees) could nonetheless be a fruitful avenue for future research. While the basic mechanisms derived in this paper will also play a role with asymmetric agents, designing optimal incentives might become significantly more demanding for the principal.

17. Flyvbjerg et al. (2003, p. 6) even coined the term "*megaprojects paradox*" for the trend that, despite their poor performance, more and more megaprojects are carried out.

Appendix 2.A

Proof of Lemma 3.

Part (i) and *(ii)*. A formal proof is provided in Rauber and Weinschenk (2024).

Part (iii). Set $P(e_t) := p\left(e_t, \ldots, e_t\right)$. If $b_t > \delta v^*_{t+1}$, then $e(b_t, v^*_{t+1})$ is defined by the solution to (FOC$_t$). By Assumptions 3 and 4, differentiation of (FOC$_t$) directly shows that

$$\frac{\partial e(b_t, v^*_{t+1})}{\partial b_t} = \frac{\overbrace{-P'(e(b_t, v^*_{t+1}))}^{<0}}{\underbrace{P''(e(b_t, v^*_{t+1}))(b_t - \delta v^*_{t+1})}_{\leq 0} \underbrace{- nc''(e(b_t, v^*_{t+1}))}_{<0}} > 0. \tag{2.A.1}$$

Differentiating (2.A.1), we get

$$\begin{aligned}\frac{\partial^2 e(b_t, v^*_{t+1})}{\partial b_t^2} = &\frac{P'(\cdot)\left(P''(\cdot) + \frac{\partial e(b_t, v^*_{t+1})}{\partial b_t}\left[P'''(\cdot)(b_t - \delta v^*_{t+1}) - nc'''(e(b_t, v^*_{t+1}))\right]\right)}{\left[P''(\cdot)(b_t - \delta v^*_{t+1}) - nc''(e(b_t, v^*_{t+1}))\right]^2} \\ &- \frac{P''(\cdot)\frac{\partial e(b_t, v^*_{t+1})}{\partial b_t}\left[P''(\cdot)(b_t - \delta v^*_{t+1}) - nc''(e(b_t, v^*_{t+1}))\right]}{\left[P''(\cdot)(b_t - \delta v^*_{t+1}) - nc''(e(b_t, v^*_{t+1}))\right]^2},\end{aligned}$$

which, using (2.A.1), may be rewritten as

$$\begin{aligned}\frac{\partial^2 e(b_t, v^*_{t+1})}{\partial b_t^2} &= \frac{\overbrace{P'(\cdot)}^{>0}\left(\overbrace{2P''(\cdot)}^{\leq 0} + \overbrace{\frac{\partial e(b_t, v^*_{t+1})}{\partial b_t}}^{>0}\left[\overbrace{P'''(\cdot)(b_t - \delta v^*_{t+1}) - nc'''(e(b_t, v^*_{t+1}))}^{\leq 0}\right]\right)}{\underbrace{\left[P''(\cdot)(b_t - \delta v^*_{t+1}) - nc''(e(b_t, v^*_{t+1}))\right]^2}_{>0}} \\ &\leq 0.\end{aligned}$$

Hence, if $b_t > \delta v^*_{t+1}$, then $b_t \mapsto e(b_t, v^*_{t+1})$ is strictly increasing and weakly concave.

□

Proof of Lemma 4. First, consider the case $r_t > \underline{r}_t$. Then, to maximize the expected profit, the optimal bonus must satisfy $\delta v^*_{t+1} \leq b_t \leq (r_t - \delta \pi^*_{t+1})/n$, where $v^*_{t+1} = v_{t+1}(\boldsymbol{b}^*_{t+1})$ and $\pi^*_{t+1} = \pi_{t+1}(\boldsymbol{b}^*_{t+1})$. We then obtain the following limits for the l.h.s. of (2.4):

$$\lim_{b_t \to \delta v^*_{t+1}} \frac{\mathrm{d}\pi_t(b_t, \boldsymbol{b}^*_{t+1})}{\mathrm{d}b_t} = P'(0) \frac{\partial e(\delta v^*_{t+1}, \boldsymbol{b}^*_{t+1})}{\partial b_t} \left(r_t - n\,\delta v^*_{t+1} - \delta \pi^*_{t+1}\right) > 0$$

$$\lim_{b_t \to (r_t - \delta \pi^*_{t+1})/n} \frac{\mathrm{d}\pi_t(b_t, \boldsymbol{b}^*_{t+1})}{\mathrm{d}b_t} = -n P\Big(e((r_t - \delta \pi^*_{t+1})/n, \boldsymbol{b}^*_{t+1})\Big) < 0.$$

By the Intermediate Value Theorem, a solution $\delta v^*_{t+1} < b^*_t < (r_t - \delta \pi^*_{t+1})/n$ to (2.4) thus exists. We next show that the principal's objective function in (2.3) is strictly concave, implying that b^*_t is unique. Differentiation yields

$$\begin{aligned}\frac{\mathrm{d}^2\pi_t(b_t, \boldsymbol{b}^*_{t+1})}{\mathrm{d}b_t^2} = (r_t - nb_t - \pi^*_{t+1})\left[P''(\cdot)\left(\frac{\partial e(b_t, \boldsymbol{b}^*_{t+1})}{\partial b_t}\right)^2 + P'(\cdot)\frac{\partial^2 e(b_t, \boldsymbol{b}^*_{t+1})}{\partial b_t^2}\right] \\ - 2n\left(P'(\cdot)\frac{\partial e(b_t, \boldsymbol{b}^*_{t+1})}{\partial b_t}\right).\end{aligned}$$

By Lemma 3 (iii), $b_t \mapsto e(b_t, \boldsymbol{b}^*_{t+1})$ is strictly increasing and weakly concave on $\delta v_{t+1}(\boldsymbol{b}^*_{t+1}) \leq b_t \leq (r_t - \delta \pi_{t+1}(\boldsymbol{b}^*_{t+1}))/n$. Consequently, on that particular interval, it holds that

$$\frac{\mathrm{d}^2\pi_t(b_t, \boldsymbol{b}^*_{t+1})}{\mathrm{d}b_t^2} < 0, \tag{2.A.2}$$

showing that the objective function is indeed strictly concave. Hence, the optimal bonus $b^*_t = b_t(\boldsymbol{r}_t)$ in period t is uniquely determined.

Next, suppose that $r_t \leq \underline{r}_t$. Then the principal is better off if the game proceeds to the next period, implying that she optimally incentivizes zero effort in period t. Thus, according to our assumption, she offers the lowest possible bonus $b^*_t = 0$, which then implements zero effort provision.[18] Hence, $b_t(\boldsymbol{r}_t) = 0$ if $r_t \leq \underline{r}_t$. □

18. Formally, the first-order condition (2.4) is satisfied for any $b_t \leq \delta v^*_{t+1}$ and, in particular, for the bonus $b^*_t = 0$.

Proof of Lemma 5. For $r_t \leq \underline{r}_t$, we have $b_t(\boldsymbol{r}_t) \equiv 0$, which directly implies that

$$\frac{\partial b_t(\boldsymbol{r}_t)}{\partial r_t} = 0.$$

By contrast, for $r_t > \underline{r}_t$, implicit differentiation of (2.4) and using Lemma 3 as well as (2.A.2) then yields

$$\frac{\partial b_t(\boldsymbol{r}_t)}{\partial r_t} = -\frac{P'(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))\dfrac{\partial e(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\partial b_t}}{\dfrac{\mathrm{d}^2\pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}b_t^2}} > 0,$$

which completes the proof. □

Proof of Proposition 5.

Part (i). We distinguish between two cases, depending on the revenue profile $\boldsymbol{r}_{t-1}$.

Case 1. If $r_{t-1} \leq \underline{r}_{t-1}$, then $b_{t-1}(\boldsymbol{r}_{t-1}) = 0$ and $e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})) = 0$ by Lemma 4 and (2.5), respectively. Since $\underline{r}_{t-1}$ is strictly increasing in r_t, $r_{t-1} \leq \underline{r}_{t-1}$ continues to hold, such that we can then immediately infer that $\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})/\mathrm{d}r_t = 0$ and $\mathrm{d}e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))/\mathrm{d}r_t = 0$ must hold.

Case 2. If $r_{t-1} > \underline{r}_{t-1}$, then $e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))$ is defined by

$$\frac{P'(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))}{n}\Big(b_{t-1}(\boldsymbol{r}_{t-1}) - \delta v_t(b_{t-1}(\boldsymbol{r}_t))\Big) - c'(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))) = 0. \tag{2.A.3}$$

Differentiation of (2.A.3) with respect to r_t yields

$$\frac{\mathrm{d}e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} = \frac{\overbrace{-P'(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))}^{<0}\left(\dfrac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} - \delta\dfrac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}\right)}{\underbrace{P''(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big(b_{t-1}(\boldsymbol{r}_{t-1}) - \delta v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))\Big) - nc''(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))}_{<0}}. \tag{2.A.4}$$

Two subcases can occur in (2.A.4), depending on the revenue profile $\boldsymbol{r}_t$.

Case 2.1. If $r_t \leq \underline{r}_t$, then, by (2.9) and (2.10), we have

$$\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} = \frac{\mathrm{d}\pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} = 0. \tag{2.A.5}$$

In view of (2.A.5), we can deduce from the first-order condition (2.4) that

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} = 0. \tag{2.A.6}$$

Inserting (2.A.5) and (2.A.6) into (2.A.4), it now follows that

$$\frac{\mathrm{d}e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} = 0.$$

Case 2.2. If $r_t > \underline{r}_t$, then the derivative (2.A.4) is strictly negative if and only if

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} < \delta\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}. \tag{2.A.7}$$

To show that (2.A.7) holds, it is helpful to establish that for $r_{t-1} > \underline{r}_{t-1}$, we have

$$\frac{\partial e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial v_t^*} = -\delta\frac{\partial e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial b_{t-1}} \quad \text{and} \quad \frac{\partial^2 e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial v_t^*\,\partial b_{t-1}} = -\delta\frac{\partial^2 e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial b_{t-1}^2}. \tag{2.A.8}$$

These relationships are seen as follows. Differentiating (2.A.3) and comparing with (2.A.1) shows that

$$\begin{aligned}\frac{\partial e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial v_t^*} &= \frac{\delta P'(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{P''(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\big(b_{t-1}-\delta v_t(\boldsymbol{b}_t)\big) - nc''(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))} \\ &= -\delta\frac{\partial e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial b_{t-1}}.\end{aligned} \tag{2.A.9}$$

Moreover, from Schwarz's Theorem and (2.A.9), we can conclude that

$$\begin{aligned}\frac{\partial^2 e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial v_t^*\,\partial b_{t-1}} &= \frac{\partial^2 e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial b_{t-1}\,\partial v_t^*} \\ &= \frac{\partial}{\partial b_{t-1}}\left[-\delta\frac{\partial e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial b_{t-1}}\right] = -\delta\frac{\partial^2 e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial b_{t-1}^2}.\end{aligned}$$

Hence, (2.A.8) holds. We next prove that (2.A.7) indeed holds, using (2.A.8). Implicit differentiation of (2.4) yields

$$\begin{aligned}\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} =& \frac{-\chi(\boldsymbol{r}_{t-1})\Big[P''(\cdot)\frac{\partial e_{t-1}^*}{\partial b_{t-1}}\frac{\partial e_{t-1}^*}{\partial v_t^*} + P'(\cdot)\frac{\partial^2 e_{t-1}^*}{\partial v_t^*\,\partial b_{t-1}}\Big] + nP'(\cdot)\frac{\partial e_{t-1}^*}{\partial v_t^*}}{\chi(\boldsymbol{r}_{t-1})\Big[P''(\cdot)\Big(\frac{\partial e_{t-1}^*}{\partial b_{t-1}}\Big)^2 + P'(\cdot)\frac{\partial^2 e_{t-1}^*}{\partial b_{t-1}^2}\Big] - 2nP'(\cdot)\frac{\partial e_{t-1}^*}{\partial b_{t-1}}}\cdot\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} \\ &+ \frac{P'(\cdot)\frac{\partial e_{t-1}^*}{\partial b_{t-1}}}{\chi(\boldsymbol{r}_{t-1})\Big[P''(\cdot)\Big(\frac{\partial e_{t-1}^*}{\partial b_{t-1}}\Big)^2 + P'(\cdot)\frac{\partial^2 e_{t-1}^*}{\partial b_{t-1}^2}\Big] - 2nP'(\cdot)\frac{\partial e_{t-1}^*}{\partial b_{t-1}}}\cdot\delta\frac{\mathrm{d}\pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t},\end{aligned} \tag{2.A.10}$$

where, for notational simplicity,

$$\chi(\boldsymbol{r}_{t-1}) := r_{t-1} - nb_{t-1}(\boldsymbol{r}_{t-1}) - \delta\pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t)), \qquad \text{and} \qquad e^*_{t-1} = e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})).$$

Using (2.A.8), the derivative (2.A.10) may be rewritten as

$$\begin{aligned}\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} =& \frac{\chi(\boldsymbol{r}_{t-1})\left[P''(\cdot)\left(\frac{\partial e^*_{t-1}}{\partial b_{t-1}}\right)^2 + P'(\cdot)\frac{\partial^2 e^*_{t-1}}{\partial b^2_{t-1}}\right] - nP'(\cdot)\frac{\partial e^*_{t-1}}{\partial b_{t-1}}}{\chi(\boldsymbol{r}_{t-1})\left[P''(\cdot)\left(\frac{\partial e^*_{t-1}}{\partial b_{t-1}}\right)^2 + P'(\cdot)\frac{\partial^2 e^*_{t-1}}{\partial b^2_{t-1}}\right] - 2nP'(\cdot)\frac{\partial e^*_{t-1}}{\partial b_{t-1}}} \cdot \delta\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} \\ &+ \frac{P'(\cdot)\frac{\partial e^*_{t-1}}{\partial b_{t-1}}}{\chi(\boldsymbol{r}_{t-1})\left[P''(\cdot)\left(\frac{\partial e^*_{t-1}}{\partial b_{t-1}}\right)^2 + P'(\cdot)\frac{\partial^2 e^*_{t-1}}{\partial b^2_{t-1}}\right] - 2nP'(\cdot)\frac{\partial e^*_{t-1}}{\partial b_{t-1}}} \cdot \delta\frac{\mathrm{d}\pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}.\end{aligned} \tag{2.A.11}$$

Observe that the second summand in (2.A.11) is strictly negative, implying that

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} < \frac{\chi(\boldsymbol{r}_{t-1})\left[P''(\cdot)\left(\frac{\partial e^*_{t-1}}{\partial b_{t-1}}\right)^2 + P'(\cdot)\frac{\partial^2 e^*_{t-1}}{\partial b^2_{t-1}}\right] - nP'(\cdot)\frac{\partial e^*_{t-1}}{\partial b_{t-1}}}{\chi(\boldsymbol{r}_{t-1})\left[P''(\cdot)\left(\frac{\partial e^*_{t-1}}{\partial b_{t-1}}\right)^2 + P'(\cdot)\frac{\partial^2 e^*_{t-1}}{\partial b^2_{t-1}}\right] - 2nP'(\cdot)\frac{\partial e^*_{t-1}}{\partial b_{t-1}}} \cdot \delta\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}. \tag{2.A.12}$$

Since the first factor on the r.h.s. of (2.A.12) is between zero and one, we can now conclude that (2.A.7) indeed holds,

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} < \delta\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}.$$

Hence, the derivative (2.A.4) is strictly negative.

Summing up *Cases 1* and *2*, we have proven that

$$\frac{\mathrm{d}e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \text{ or } r_{t-1} \leq \underline{r}_{t-1} \\ < 0 & \text{if } r_t > \underline{r}_t \text{ and } r_{t-1} > \underline{r}_{t-1}, \end{cases}$$

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} \begin{cases} = 0 & \text{if } r_t \leq \underline{r}_t \text{ or } r_{t-1} \leq \underline{r}_{t-1} \\ \gtreqless 0 & \text{if } r_t > \underline{r}_t \text{ and } r_{t-1} > \underline{r}_{t-1}, \end{cases}$$

which completes the proof of *Part (i)*.

Part (ii). In the following, suppose that $r_t > \underline{r}_t$ and $r_{t-1} > \underline{r}_{t-1}$. We first show that if $r_{t-1} - \underline{r}_{t-1}$ is sufficiently small, then

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} < 0.$$

Note that, if $r_{t-1} - r_{t-1} \to 0$, then $\chi(\boldsymbol{r}_{t-1}) \to 0$ because

$$0 \leq \chi(\boldsymbol{r}_{t-1}) \leq r_{t-1} - \underline{r}_{t-1}.$$

For $\chi(\boldsymbol{r}_{t-1}) \to 0$, in turn, the derivative (2.A.11) becomes

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} = \frac{\delta}{2n}\left(n\frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} - \frac{\mathrm{d}\pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}\right). \tag{2.A.13}$$

From (2.A.13), (2.4), (2.8), (2.10), and the fact that

$$\frac{\partial b_t(\boldsymbol{r}_t)}{\partial r_t} < \frac{1}{2n},$$

we can conclude that (2.A.13) is strictly negative if the following sufficient condition is satisfied:

$$(2n-1)\Big(b_t(\boldsymbol{r}_t) - \delta v_{t+1}(\boldsymbol{b}_{t+1}(\boldsymbol{r}_{t+1}))\Big) \leq r_t - \underline{r}_t. \tag{2.A.14}$$

We next show that (2.A.14) indeed holds. Recognize that (2.A.14) is fulfilled in the limit $r_t \to \underline{r}_t$, since we then have $b_t(\boldsymbol{r}_t) - \delta v_{t+1}(\boldsymbol{b}_{t+1}(\boldsymbol{r}_{t+1})) \to 0$. In addition, note that the slope of the r.h.s. in (2.A.14) is strictly greater than the slope of its l.h.s. because

$$\begin{aligned}\frac{\partial}{\partial r_t}\left((2n-1)\Big(b_t(\boldsymbol{r}_t) - \delta v_{t+1}(\boldsymbol{b}_{t+1}(\boldsymbol{r}_{t+1}))\Big)\right) &= (2n-1)\frac{\partial b_t(\boldsymbol{r}_t)}{\partial r_t} \\ &< \frac{2n-1}{2n} < 1 = \frac{\partial}{\partial r_t}\big(r_t - \underline{r}_t\big).\end{aligned}$$

Hence, we can infer that (2.A.14) is satisfied for all $r_t > \underline{r}_t$, implying that

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} < 0$$

holds if $r_{t-1} - \underline{r}_{t-1}$ is sufficiently small.

As a last step, we show that if $r_{t-1} - \underline{r}_{t-1}$ is sufficiently large, then

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} > 0.$$

Since $\chi(\boldsymbol{r}_{t-1})$ is strictly increasing and unbounded, we obtain for $r_{t-1} - \underline{r}_{t-1} \to \infty$ that $\chi(\boldsymbol{r}_{t-1}) \to \infty$. Consequently, for $r_{t-1} - \underline{r}_{t-1} \to \infty$, (2.A.11) becomes

$$\frac{\mathrm{d}b_{t-1}(\boldsymbol{r}_{t-1})}{\mathrm{d}r_t} = \delta \frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} > 0.$$

The proof of *Part (ii)* is now complete. □

Proof of Lemma 6. The principal's expected profit in period $t-1$ is given by

$$\begin{aligned} \pi_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})) = & P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big(r_{t-1} - nb_{t-1}(\boldsymbol{r}_{t-1})\Big) \\ & + \Big(1 - P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big)\delta \pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t)). \end{aligned} \tag{2.A.15}$$

By the Envelope Theorem, the derivative of (2.A.15) computes

$$\begin{aligned} \frac{\mathrm{d}\pi_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} &= \Big(1 - P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big)\delta \frac{\mathrm{d}\pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t} \\ + P'(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))) & \frac{\partial e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\partial v_t^*} \frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}\Big(r_{t-1} - n\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}) - \delta \pi_t(\boldsymbol{b}_t(\boldsymbol{r}_t))\Big). \end{aligned} \tag{2.A.16}$$

Using (2.4), (2.10), and (2.A.9), we may then rewrite (2.A.16) as

$$\begin{aligned} \frac{\mathrm{d}\pi_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} = & \ \delta \Big(1 - P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big) P(e(\boldsymbol{b}_t(\boldsymbol{r}_t))) \\ & - \delta\, n\, P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))) \frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}, \end{aligned}$$

which is exactly (2.12). □

Proof of Theorem 3. From (2.4) it can be deduced that if $r_{t-1} - \underline{r}_{t-1} \to \infty$, then the net bonus in $t-1$ converges to infinity, $b_{t-1} - \delta v_t^* \to \infty$, implying that $P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))) \to 1$. As a consequence, the derivative (2.A.16) becomes

$$\frac{\mathrm{d}\pi_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} = -\delta n \frac{\mathrm{d}v_t(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\mathrm{d}r_t}. \tag{2.A.17}$$

From (2.9), we can conclude that (2.A.17) is strictly negative if $r_t > \underline{r}_t$. Hence,

$$\frac{\mathrm{d}\pi_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} < 0,$$

given that $r_t > \underline{r}_t$ and $r_{t-1} - \underline{r}_{t-1}$ sufficiently large. □

Proof of Proposition 7. Using the recursive definition in (2.6), the success rate in period $t-1$ reads

$$\begin{aligned} S_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})) &= P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})) \\ &+ \Big(1 - P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big)\Big[P(e(\boldsymbol{b}_t(\boldsymbol{r}_t)) + \Big(1 - P(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))\Big)S_{t+1}(\boldsymbol{b}_{t+1}(\boldsymbol{r}_{t+1}))\Big]. \end{aligned}$$

Differentiating with respect to r_t and straightforward rearranging yields

$$\begin{aligned} \frac{\mathrm{d}S_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} &= \Big(1 - S_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))\Big) \\ &\cdot \Bigg[\Big(1 - P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big)P'(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))\,\frac{\partial e(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\partial b_t}\frac{\partial \boldsymbol{b}_t(\boldsymbol{r}_t)}{\partial r_t} \\ &+ \Big(1 - P(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))\Big)P'(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\,\frac{\mathrm{d}e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t}\Bigg]. \end{aligned} \tag{2.A.18}$$

Since $\Big(1 - S_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))\Big) > 0$, we can conclude from (2.A.18) and a few further straightforward rearrangements that

$$\begin{aligned} &\frac{\mathrm{d}S_{t-1}(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t} < 0 \\ \iff &\frac{\Big(1 - P(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))\Big)P'(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))}{\Big(1 - P(e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1})))\Big)P'(e(\boldsymbol{b}_t(\boldsymbol{r}_t)))} \cdot \frac{-\dfrac{\mathrm{d}e(\boldsymbol{b}_{t-1}(\boldsymbol{r}_{t-1}))}{\mathrm{d}r_t}}{\dfrac{\partial e(\boldsymbol{b}_t(\boldsymbol{r}_t))}{\partial b_t}\dfrac{\partial \boldsymbol{b}_t(\boldsymbol{r}_t)}{\partial r_t}} > 1. \end{aligned}$$

which is exactly Condition (2.13) stated in the proposition. □

Proof of Lemma 7. To proof Lemma 7, we show that *Part (i)* and *(ii)* hold for every period $t = 1, \dots, T$.

Consider the **last period T**.
Part (i). Assume, for the sake of contradiction, that

$$v_T(b_T^{\mathrm{L}}) \;<\; v_T(b_T^*). \tag{2.A.19}$$

would hold. This requires that $b_T^{\mathrm{L}} \neq b_T^*$. Since, by Lemma 4, b_T^* is the unique maximizer of the principal's expected profit in T, this implies that

$$\pi_T(b_T^{\mathrm{L}}) \;<\; \pi_T(b_T^*). \tag{2.A.20}$$

Conditions (2.A.19) and (2.A.20) now yield a contradiction to renegotiation proofness in the sense of Definition 6: changing the implemented bonus form b_T^{L} to b_T^* would generate a Pareto improvement. Hence, for all renegotiation-proof contracts, we must have that

$$v_T(b_T^{\mathrm{L}}) \;\geq\; v_T(b_T^*), \qquad \text{and thus} \qquad \pi_T(b_T^{\mathrm{L}}) \;\leq\; \pi_T(b_T^*), \tag{2.A.21}$$

with both inequalities being strict if $b_T^{\mathrm{L}} > b_T^*$.
Part (ii). Since the bonus payment b_T^* is the unique maximizer of the principals profit, we have for any $\tilde{b}_T \neq b_T^*$ that

$$\pi_T(\tilde{b}_T) \;<\; \pi_T(b_T^*).$$

Hence, the contract b_T^* is Pareto-efficient and therefore renegotiation-proof, since any other bonus payment $\tilde{b}_T$ would leave the principal strictly worse off.

Now, consider the **second-last period $T-1$**.
Part (i). Assume, for the sake of contradiction, that

$$v_{T-1}(b_{T-1}^{\mathrm{L}}, b_T^{\mathrm{L}}) \;<\; v_{T-1}(b_{T-1}^*, b_T^*), \tag{2.A.22}$$

which directly implies that $(b_{T-1}^{\mathrm{L}}, b_T^{\mathrm{L}}) \neq (b_{T-1}^*, b_T^*)$. From (2.2) we can infer that, ceteris paribus, the principal's expected profit π_{t-1} is increasing in her expected profit π_t and non-increasing in the agents' payoff v_t of the subsequent period. Hence, with a slight abuse of notation, we can conclude from (2.2) that

$$\frac{\partial \pi_{t-1}}{\partial \pi_t} \;>\; 0, \qquad \text{and} \qquad \frac{\partial \pi_{t-1}}{\partial v_t} \;\leq\; 0. \tag{2.A.23}$$

Combining (2.A.21) and (2.A.23) yields

$$\pi_{T-1}(b^{\mathrm{L}}_{T-1}, b^{\mathrm{L}}_{T}) \;\leq\; \pi_{T-1}(b^{\mathrm{L}}_{T-1}, b^{*}_{T}) \;\leq\; \pi_{T-1}(b^{*}_{T-1}, b^{*}_{T}), \tag{2.A.24}$$

where the last inequality follows from the fact that b^{*}_{T-1} is the unique maximizer of the principal's expected profit provided that $b_T = b^{*}_{T}$. Since $(b^{\mathrm{L}}_{T-1}, b^{\mathrm{L}}_{T}) \neq (b^{*}_{T-1}, b^{*}_{T})$, at least one of the inequalities in (2.A.24) is strict. Hence, (2.A.22) and (2.A.24) imply a contradiction to renegotiation-proofness: the contract (b^{*}_{T-1}, b^{*}_{T}) has a one-period continuation b^{*}_{T} that is renegotiation-proof, and offering this contract allows an Pareto improvement compared to the contract $(b^{\mathrm{L}}_{T-1}, b^{\mathrm{L}}_{T})$. Hence, renegotiation-proof contracts $(b^{\mathrm{L}}_{T-1}, b^{\mathrm{L}}_{T})$ must satisfy $v_t(\boldsymbol{b}^{\mathrm{L}}_{t}) \geq v_t(\boldsymbol{b}^{*}_{t})$ for all $t \in \{T-1, T\}$ and thus $\pi_t(\boldsymbol{b}^{\mathrm{L}}_{t}) \leq \pi_t(\boldsymbol{b}^{*}_{t})$ for all $t \in \{T-1, T\}$.

Part (ii). By construction of b^{*}_{T-1}, it holds for any $\tilde{b}_{T-1} \neq b^{*}_{T-1}$ that

$$\pi_{T-1}(\tilde{b}_{T-1}, b^{*}_{T}) \;<\; \pi_{T-1}(b^{*}_{T-1}, b^{*}_{T}). \tag{2.A.25}$$

Moreover, since we have already shown that any renegotiation-proof bonus b^{L}_{T} satisfies $v_T(b^{\mathrm{L}}_{T}) \geq v_T(b^{*}_{T})$ and $\pi_T(b^{\mathrm{L}}_{T}) \leq \pi_T(b^{*}_{T})$, we can infer from (2.A.23) that

$$\pi_{T-1}(\tilde{b}_{T-1}, b^{\mathrm{L}}_{T}) \;\leq\; \pi_{T-1}(\tilde{b}_{T-1}, b^{*}_{T}). \tag{2.A.26}$$

Combining (2.A.25) and (2.A.26), we find that, compared to the contract (b^{*}_{T-1}, b^{*}_{T}), the principal is strictly worse off with any contract $(\tilde{b}_{T-1}, b^{\mathrm{L}}_{T})$, i.e., any other contract whose one-period continuation is renegotiation-proof. Therefore, the contract (b^{*}_{T-1}, b^{*}_{T}) is Pareto-efficient among these contracts and thus renegotiation-proof in the sense of Definition 6.

Repeating the above arguments until period $t = 1$ completes the proof. □

Proof of Theorem 4. Since the principal chooses a renegotiation-proof contract that maximizes her expected profit in the first period, it directly follows from Lemma 7 that $\boldsymbol{b}^{*}_{1}$ must be *an* equilibrium renegotiation-proof contract.

Its uniqueness can be shown by contradiction. Assume, for the sake of contradiction, that a second renegotiation-proof contract $\tilde{\boldsymbol{b}}_1$ with $\pi_1(\boldsymbol{b}^{*}_{1}) = \pi_1(\tilde{\boldsymbol{b}}_1)$ and $b^{*}_{t} \neq \tilde{b}_t$ for some $t \in \{1, \ldots, T\}$ would exist. Let τ be the last period where $b^{*}_{t} \neq \tilde{b}_t$ holds. Because b^{*}_{τ} is the unique maximizer given $\boldsymbol{b}_{\tau+1}$, we have that

$$\pi_\tau(\tilde{b}_\tau, \boldsymbol{b}^{*}_{\tau+1}) \;<\; \pi_\tau(b^{*}_{\tau}, \boldsymbol{b}^{*}_{\tau+1}). \tag{2.A.27}$$

Moreover, renegotiation-proofness then requires

$$v_\tau(\tilde{b}_\tau, \boldsymbol{b}^*_{\tau+1}) > v_\tau(b^*_\tau, \boldsymbol{b}^*_{\tau+1}). \tag{2.A.28}$$

From (2.A.27) and (2.A.28) together with (2.A.23), we obtain for period $\tau - 1$ that

$$\pi_{\tau-1}(\tilde{b}_{\tau-1}, \tilde{b}_\tau, \boldsymbol{b}^*_{\tau+1}) < \pi_{\tau-1}(\tilde{b}_{\tau-1}, b^*_\tau, \boldsymbol{b}^*_{\tau+1}) \leq \pi_{\tau-1}(b^*_{\tau-1}, b^*_\tau, \boldsymbol{b}^*_{\tau+1}), \tag{2.A.29}$$

where the last inequality is strict if $b^*_{\tau-1} \neq \tilde{b}_{\tau-1}$. Again, due to renegotiation-proofness, it must then hold that

$$v_{\tau-1}(\tilde{b}_{\tau-1}, \tilde{b}_\tau, \boldsymbol{b}^*_{\tau+1}) > v_{\tau-1}(b^*_{\tau-1}, b^*_\tau, \boldsymbol{b}^*_{\tau+1}). \tag{2.A.30}$$

Exploiting (2.A.23), (2.A.29) together with (2.A.30) imply that

$$\pi_{\tau-2}(\tilde{b}_{\tau-2}, \tilde{b}_{\tau-1}, \tilde{b}_\tau, \boldsymbol{b}^*_{\tau+1}) < \pi_{\tau-2}(\tilde{b}_{\tau-2}, b^*_{\tau-1}, b^*_\tau, \boldsymbol{b}^*_{\tau+1}) \leq \pi_{\tau-2}(\boldsymbol{b}^*_{\tau-2}).$$

By repeating the above arguments until period $t = 1$, we obtain that

$$\pi_1(\tilde{\boldsymbol{b}}_1) < \pi_1(\boldsymbol{b}^*_1),$$

which is a contradiction. Hence, there exists no other renegotiation-proof contract $\tilde{\boldsymbol{b}}_1$ that satisfies $\pi_1(\boldsymbol{b}^*_1) = \pi_1(\tilde{\boldsymbol{b}}_1)$ and $b^*_t \neq \tilde{b}_t$ for some t, i.e., $\boldsymbol{b}^*_1$ is the principal's unique equilibrium renegotiation-proof contract. □

References

Al-Ahmad, W., Al-Fagih, K., Khanfar, K., Alsamara, K., Abuleil, S., & Abu-Salem, H. (2009). A taxonomy of an it project failure: Root causes. *International Management Review*, *5*(1), 93–104. [72]

Altan, B. (2019). Dynamic moral hazard with sequential tasks. *Economics Letters*, *183*, 108606. [49]

Bellman, R. (1957). *Dynamic programming*. Princeton University Press. [53]

Benoit, J.-P., & Krishna, V. (1993). Renegotiation in finitely repeated games. *Econometrica*, *61*(2), 303–323. [69]

Bergemann, D., & Hege, U. (2005). The financing of innovation: Learning and stopping. *RAND Journal of Economics*, *36*(4), 719–752. [50]

Bolton, P. (1990). Renegotiation and the dynamics of contract design. *European Economic Review*, *34*, 303–310. [49, 69]

Bolton, P., & Dewatripont, M. (2005). *Contract theory*. MIT Press. [46]

Bonatti, A., & Hörner, J. (2011). Collaborating. *American Economic Review*, *101*(2), 632–663. [50]

Brown, S. A., Chervany, N. L., & Reinicke, B. A. (2007). What matters when introducing new information technology. *Communications of the ACM*, *50*(9), 91–96. [72]

Chandrasekher, M. (2015). Unraveling in a repeated moral hazard model with multiple agents. *Theoretical Economics*, *10*(1), 11–49. [50]

Che, Y.-K., & Yoo, S.-W. (2001). Optimal incentives for teams. *American Economic Review*, *91*(3), 525–541. [50]

Chiappori, P.-A., Macho, I., Rey, P., & Salanié, B. (1994). Repeated moral hazard: The role of memory, commitment, and the access to credit markets. *European Economic Review*, *38*(8), 1527–1553. [49]

Dewatripont, M. (1989). Renegotiation and information revelation over time: The case of optimal labor contracts. *Quarterly Journal of Economics*, *104*(3), 589–619. [49, 69]

Flyvbjerg, B., Bruzelius, N., & Rothengatter, W. (2003). *Megaprojects and risk: An anatomy of ambition*. Cambridge University Press. [72]

Fudenberg, D., Holmström, B., & Milgrom, P. (1990). Short-term contracts and long-term agency relationships. *Journal of Economic Theory*, *51*(1), 1–31. [49]

Fudenberg, D., & Tirole, J. (1990). Moral hazard and renegotiation in agency contracts. *Econometrica*, *58*(6), 1279–1319. [69]

Georgiadis, G. (2015). Projects and team dynamics. *Review of Economic Studies*, *82*(1), 187–218. [50]

Georgiadis, G. (2022). Contracting with moral hazard: A review of theory & empirics. *Working Paper. Available at SSRN 4196247*. [46]

Green, B., & Taylor, C. (2016). Breakthroughs, deadlines, and self-reported progress: Contracting for multistage projects. *American Economic Review*, *106*(12), 3660–3699. [49, 50]

Gupta, A. K., & Wilemon, D. L. (1990). Accelerating the development of technology-based new products. *California Management Review*, *32*(2), 24–44. [72]

Halac, M., Kartik, N., & Liu, Q. (2016). Optimal contracts for experimentation. *Review of Economic Studies*, *83*(3), 1040–1091. [49, 50]

Halac, M., Lipnowski, E., & Rappoport, D. (2021). Rank uncertainty in organizations. *American Economic Review*, *111*(3), 757–86. [46]

Hart, O. D., & Tirole, J. (1988). Contract renegotiation and coasian dynamics. *Review of Economic Studies*, *55*(4), 509–540. [49]

Holmström, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, *10*(1), 74–91. [46]

Holmström, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, *13*(2), 324–340. [46]

Holmström, B. (1983). Equilibrium long-term labor contracts. *Quarterly Journal of Economics*, *98*, 23–54. [49]

Holmström, B., & Milgrom, P. (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*, 24–52. [46]

Hörner, J., & Samuelson, L. (2013). Incentives for experimenting agents. *RAND Journal of Economics*, *44*(4), 632–663. [50]

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305–360. [46]

Laffont, J.-J., & Martimort, D. (2002). *The theoy of incentives: The principal agent problem*. Princeton University Press. [46]

Laffont, J.-J., & Tirole, J. (1990). Adverse selection and renegotiation in procurement. *Review of Economic Studies*, *57*(4), 597–625. [49]

Laux, C. (2001). Limited-liability and incentive contracting with multiple projects. *RAND Journal of Economics*, *32*(3), 514–526. [46]

Lhuillery, S., & Pfister, E. (2009). R&d cooperation and failures in innovation projects: Empirical evidence from french cis data. *Research Policy*, *38*(1), 45–57. [72]

Mason, R., & Välimäki, J. (2015). Getting it done: Dynamic incentives to complete a project. *Journal of the European Economic Association*, *13*(1), 62–97. [49]

Mayer, S. (2022). Financing breakthroughs under failure risk. *Journal of Financial Economics*, *144*(3), 807–848. [49, 50]

Mirrlees, J. A. (1976). The optimal structure of incentives and authority within an organization. *Bell Journal of Economics*, *7*(1), 105–131. [46]

Mittal, Y. K., Paul, V. K., Rostami, A., Riley, M., & Sawhney, A. (2020). Delay factors in construction of healthcare infrastructure projects: A comparison amongst developing countries. *Asian Journal of Civil Engineering*, *21*, 649–661. [72]

Moroni, S. (2022). Experimentation in organizations. *Theoretical Economics*, *17*(3), 1403–1450. [50]

Mylovanov, T., & Schmitz, P. W. (2008). Task scheduling and moral hazard. *Economic Theory*, *37*, 307–320. [46]

Ohlendorf, S., & Schmitz, P. W. (2012). Repeated moral hazard and contracts with memory: The case of risk-neutrality. *International Economic Review*, *53*(2), 433–452. [50, 51]

Piskorski, T., & Westerfield, M. M. (2016). Optimal dynamic contracts with moral hazard and costly monitoring. *Journal of Economic Theory*, *166*, 242–281. [46]

Radas, S., & Bozic, L. (2012). Overcoming failure: Abandonments and delays of innovation projects in smes. *Industry and Innovation*, *19*(8), 649–669. [72]

Rauber, T., & Weinschenk, P. (2024). Dynamic interaction and (in)effectiveness of financial incentives. *Working Paper. Available at SSRN 3987829*. [50, 73]

Rey, P., & Salanie, B. (1990). Long-term, short-term and renegotiation: On the value of commitment in contracting. *Econometrica*, *58*(3), 597–619. [49]

Rogerson, W. P. (1985). Repeated moral hazard. *Econometrica*, *53*(1), 69–76. [46]

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *American Economic Review*, *63*(2), 134–139. [46]

Salanié, B. (2005). *The economics of contracts: A primer*. MIT Press. [46]

Sannikov, Y. (2008). A continuous-time version of the principal-agent problem. *Review of Economic Studies*, *75*(3), 957–984. [46]

Steininger, B. I., Groth, M., & Weber, B. L. (2021). Cost overruns and delays in infrastructure projects: The case of stuttgart 21. *Journal of Property Investment & Finance*, *39*(3), 256–282. [72]

Toxvaerd, F. (2006). Time of the essence. *Journal of Economic Theory*, *129*(1), 252–272. [49]

Varas, F. (2018). Managerial short-termism, turnover policy, and the dynamics of incentives. *Review of Financial Studies*, *31*(9), 3409–3451. [49, 50]

Wang, C. (2000). Renegotiation-proof dynamic contracts with private information. *Review of Economic Dynamics*, *3*(3), 396–422. [69]

Weinschenk, P. (2016). Procrastination in teams and contract design. *Games and Economic Behavior*, *98*, 264–283. [50]

Weinschenk, P. (2021). On the benefits of time-inconsistent preferences. *Journal of Economic Behavior & Organization*, *182*, 185–195. [50]

Whitney, K. M., & Daniels, C. B. (2013). The root cause of failure in complex it projects: Complexity itself. *Procedia Computer Science*, *20*, 325–330. [72]

Winter, E. (2004). Incentives and discrimination. *American Economic Review*, *94*(3), 764–773. [46]

Zhao, R. R. (2006). Renegotiation-proof contract in repeated agency. *Journal of Economic Theory*, *131*(1), 263–281. [69]

Chapter 3

Designing Emissions Trading Schemes: Negotiations on the Amount & Allocation of Permits*

Joint with Fabian Naumann

Abstract

National free-riding incentives prevent necessary reductions in global emissions and thus paralyze combating climate change. By considering two countries that engage in bargaining over a joint emissions cap according to Rubinstein's (1982) alternating-offers model, we discover a simple mechanism to overcome free-riding incentives and achieve efficient outcomes. Provided that countries are sufficiently symmetric, allowing them to endogenously design a joint cap-and-trade system by negotiating the amount and allocation of permits yields the efficient level of emissions and maximizes welfare. In contrast, if the negotiating countries are too heterogeneous, the scope for implicit side payments in this system may not be sufficient to implement an efficient emissions cap.

Keywords: Emissions Trading Schemes, Alternating-Offers Model, Nash Bargaining Solution, Outside Option, Free-Riding.

JEL Classification: C78, H41, Q53, Q58.

Disclosure: A (very) preliminary version of the project was part of Naumann's (2022) dissertation and can be found in Appendix 3.A.2. Comparing that version to the one presented in this thesis demonstrates that the project has advanced extensively ever since.

*Acknowledgments. We thank Joshua Bißbort, Daniel Heyen, Çağıl Koçyiğit, Paul Ritschel, Robert Schmidt, Philipp Weinschenk, and Jan Wenzelburger, as well as the participants of research seminars at the Universities of Cambridge and Kaiserslautern-Landau, for helpful comments and suggestions.

3.1 Introduction

Climate change constitutes one of the most severe challenges currently facing humanity (see, among many others, Nordhaus, 2019; Stern, 2007; Weitzman, 2007). It is well-known that extensive global emissions of greenhouse gases such as carbon dioxide, methane, and nitrous oxide lead to an increase in the temperature at the Earth's surface, which in turn promotes weather and climate extremes worldwide. Some future changes in the climate system are already inescapable, but their extent can be limited via immediate and effective emissions cuts (IPCC, 2023). While lower global emissions are thus undoubtedly beneficial from a normative perspective[1], the incentive to free-ride on other countries' abatement activities has created a deadlock (Underdal et al., 2012). An emissions trading scheme ("cap-and-trade system") is a prominent policy instrument that, if designed appropriately, breaks the deadlock and yields efficient outcomes (Schmalensee & Stavins, 2017). However, determining, implementing, and enforcing an appropriate design is a major challenge for policy-makers (Egenhofer, 2007; Nordhaus, 2007; Stavins, 2008a; Weitzman, 2014). Our paper explores whether a simple mechanism can implement efficient outcomes: allowing countries to endogenously design a joint cap-and-trade system through negotiations. We thus shed light on the question of whether this simple procedure for designing emissions trading schemes can solve the problem of excessive greenhouse gas emissions.

We consider a simple two-country model in which each country bears individual costs for reducing emissions, while, at the same time, benefiting not only from its own abatement but also from abatement activities carried out in the other country. The countries are rational and negotiate according to the alternating offers a la Rubinstein (1982) with complete and perfect information. Due to the negligible friction in the bargaining process, the subgame perfect equilibrium in this dynamic game coincides with Nash's (1950) bargaining solution, i.e., the solution to a simple optimization problem.

Our analysis starts by exploring the benchmark scenario in which both countries deploy national caps. Due to the positive externality – countries benefit from each other's reduction in emissions – we find a strong free-riding incentive that leads to an inefficiently high level of emissions. We then direct attention to a setting where the countries set a joint emissions cap through negotiations. The basic mechanisms are explored in a stylized setting where the countries have agreed on setting up a

1. Stern (2008, p. 1) even refers to the excessive greenhouse gas emissions as *"the biggest market failure the world has seen."*

joint cap-and-system and negotiate only on the amount of certificates. To derive our main results, we then successively add more degrees of freedom by *(i)* allowing the countries not only to bargain on the amount of certificates but also on their initial allocation. Afterwards, we *(ii)* incorporate the possibility of strategic termination of the negotiations. That is, each country may actively decide to end the negotiations, resulting in national abatement activities. A comparison with the benchmark then yields our two main results.

First, enabling the countries to set up a joint emissions trading scheme and to bargain over its design allows them to overcome the free-riding incentive and implement the efficient emissions cap if the countries are sufficiently symmetric. That is, if they are sufficiently similar in terms of their cost and benefit structures as well as their initial emissions. This holds irrespective of the possibility of strategic opting out of the negotiations. Intuitively, the countries agree on the cap that maximizes overall welfare. The initial allocation of certificates is then used as implicit side payment to distribute the highest level of welfare equally across countries. Second, negotiations do not necessarily implement the efficient cap if countries' benefits resulting from this cap are too heterogeneous. The intuition is as follows. The scope for using the initial allocation as implicit side payment is limited as a country cannot receive more than the entire share of certificates. If the benefits from the efficient cap are sufficiently different, then even allocating all certificates to the low-benefit country is insufficient to generate equal welfare levels in both countries. The countries will then rather agree on another cap that is less efficient but yields a more equal distribution of individual welfare levels. As we argue afterwards, our two insights also carry over to settings in which more than two countries negotiate on the design of a multilateral cap-and-trade system.

Related Literature. Free-riding incentives have received particular attention in the literature as a primary force preventing emissions reduction and hindering strict international agreements to curb climate change (see, e.g., Barrett, 1994, 2003; Carraro and Siniscalco, 1993; Nordhaus, 2015). They result from the public good nature of abatement. While all countries benefit from lower global emissions, only those countries that actually reduce their emissions bear the associated costs. This situation gives rise to Hardin's (1968) infamous tragedy of commons: each country will leave costly abatement activities to the others. As a remedy for the free-riding problem, economic theory proposes *price-based* and *quantity-based* instruments (see, e.g., Mas-Colell et al., 1995; Nordhaus, 2007; Weitzman, 1974).[2]

2. See Aldy et al. (2003) for a more nuanced distinction and Goulder and Schein (2013) as well as Stavins (2022) for a comparison of the different approaches.

While price-based instruments originate from Pigou's (1920) taxation of externalities, quantity-based instruments build on the concept of tradable permits as proposed by Coase (1960) and Dales (1968). Both approaches are widespread policy instruments to address either national or international climate targets.

For price-based instruments, Weitzman (2014) explores a simple mechanism to alleviate the free-riding incentive. Given that the countries can commit to a single emissions price, they determine this price by pairwise majority voting, resulting in emissions close to the efficient level. Intuitively, due to global commitment to the emissions price, a country's additional costs from a higher emissions price are offset by its additional benefit arising since all other countries reduce their emissions at the same time in response to the higher price. However, for quantity-based instruments, Weitzman (2014, p. 31) suspects that *"even if there were a collective commitment to negotiate or vote on a second-stage worldwide total emissions cap, disagreements over the first-stage subdivision formula (...) would paralyze such a quantity-based approach."* Despite these doubts, our analysis reveals that a similar mechanism also exists for quantity-based instruments where the total amount of emissions is capped, and the emissions price is determined via trading, given that the countries are sufficiently symmetric. We thereby contribute to two strands of the emissions-trading literature, namely endogenous allowance choices and linking emissions trading schemes.

Previous research by Helm (2003) has shown that *endogenous allowance choices* made by countries do not automatically result in lower pollution levels, as environmentally more (less) concerned countries choose to pollute less (more) such that the environmental efforts offset. Using numerical simulations, Smead et al. (2014) investigate a setup where agents bargain over their share of the fixed total emissions. They find that negotiations tend to fail if too many agents request overproportional emissions shares, making the initial demand for those shares a key factor for successful negotiations. Our paper complements these results by identifying an insufficient scope for side payments as an additional mechanism that may prevent efficient endogenous allowance choices.

Considering situations where emissions trading systems have already been implemented, the literature on *linking emissions trading schemes* raises the question as to whether combining these systems is beneficial. Flachsland et al. (2009) analyze the benefits and drawbacks of linking, such as reduced volatility, strengthening the multilateral commitment versus expanded emission caps, abatement targets that are not in line with a burden-sharing approach, and declining national regulatory power. Doda and Taschini (2017) argue that linking becomes more advantageous the larger the jurisdictions' size and variances of benefit shocks, while a stronger correlation of these shocks and higher sunk costs of linking result in the opposite effect. Doda et al. (2019) find that multilateral linking can lead to tremendous efficiency gains, which arise equally from effort- and risk-sharing. However, Habla and Winkler (2018) demonstrate that strategic delegation hinders the linking of emissions trading schemes.[3] We contribute to this literature by showing that linking arises naturally via negotiations if the countries are sufficiently symmetric.

More generally, applying game theory to analyze negotiations on ecological agreements and the provision of environmental public goods has been a vibrant research area over the past two decades (Caparrós, 2016).[4] Bargaining models were deployed in the context of global north-south climate change negotiations (cf. Caparrós et al., 2004), air pollution (cf. Harstad, 2007), investment in green technologies (cf. Urpelainen, 2012), global biodiversity regulation (cf. Swanson and Groom, 2012), interregional water sharing (cf. Nehra and Caplan, 2022), and climate policies (cf. Harstad, 2023). The paper closest in spirit to ours is Dijkstra and Nentjes (2020), who analyze negotiations on tradable production certificates in a related model. While the structure of their game is different, their results also differ in so far as they find that bargaining *always* leads to efficient production levels. In spite of this plentiful literature, this paper is, to the best of our knowledge, the first to explore how two countries endogenously design a joint cap-and-trade system by bargaining over the amount and allocation of permits.

Outline. The remainder of this paper is organized as follows. In the next section, we introduce our two-country model and define a country's welfare as a function of the design of the emissions trading scheme. Section 3.3 analyzes two benchmark scenarios for our welfare analysis by considering a social planner and national caps. In Section 3.4, we explore the simplest version of cap negotiation, i.e., a setting where the countries can bargain on the emissions cap only. Section 3.5 then generalizes the preceding analysis by allowing the countries to bargain simultaneously on the emissions cap and the initial allocations of permits, while also incorporating the possibility of strategic opting out of the negotiations. A discussion of multilateral negotiations, i.e., the case of more than two countries, is provided in Section 3.6. Finally, Section 3.7 concludes our analysis by highlighting its implications. All proofs are provided in Appendix 3.A.1.

3. There is a strand in the political economy literature that examines how delegates in the context of international environmental agreements are chosen. The baseline shared here is that countries may select delegates that misrepresent their preferences (see, e.g.,Graziosi, 2009; Habla and Winkler, 2018; Segendorff, 1998). We, however, abstract from such considerations and assume that the countries' preferences are correctly represented in the negotiations.

4. For an excellent overview of the literature, see Caparrós (2016).

3.2 Model and Basic Insights

This section presents the underlying theoretical framework for describing a cap-and-trade system. Subsequently, we derive individual welfare for a country as a function of the scheme's design.

Framework. Consider two countries that emit greenhouse gases. Each country $i \in \{1,2\}$ carries out abatement activities a_i relative to its emissions level under "business as usual" $e_{i,0} \in \mathbb{R}_{++}$.[5] Abatement activities affect country i's welfare through three channels. First, there are benefits of overall abatement $B_i(\sum_i a_i)$. Reducing emissions has thus a positive externality since country i also benefits from the abatement made by country $-i$ and vice versa. Second, country i bears costs of its own abatement $C_i(a_i)$. Third, as countries participate in a cap-and-trade system, emissions trading additionally results in either revenues or costs, depending on whether i is a seller or buyer of permits. A country acts as a seller [buyer] of permits if its realized emissions e_i are lower [higher] than its initial endowment of permits $\bar{e}_i \in \mathbb{R}_+$. The emissions market price p is endogenously determined. Putting these components together, country i's welfare amounts to:

$$B_i(\textstyle\sum_i a_i) - C_i(a_i) + (\bar{e}_i - e_i) \cdot p, \qquad i = 1,2. \quad (3.1)$$

Note that the benefits of lower emissions are expressed in terms of emissions abatement, i.e., the higher the abatement, the lower the emissions and thus the higher the benefits. For the emissions cap of the entire scheme, $\bar{E}$, it holds that

$$\bar{E} = \sum_i \bar{e}_i, \quad \text{and} \quad \bar{e}_i = \mu_i \bar{E} \qquad i = 1,2, \quad (3.2)$$
$$\text{where} \quad \mu_2 = 1 - \mu_1,$$

and $\mu_1 \in [0,1]$ denoting the share of permits allocated to country 1. As realized emissions are determined by emissions under "business as usual" minus abatement, we can rewrite the country i's welfare in (3.1) as

$$B_i(\textstyle\sum_i a_i) - C_i(a_i) + (\mu_i \bar{E} - (e_{i,0} - a_i))p, \qquad i = 1,2. \quad (3.3)$$

Following the literature, in which abatement costs are commonly assumed to be quadratic (see, for instance, Barrett, 1994; Baudry et al., 2021; Gersbach and Hummel, 2016; McGinty, 2007; Weitzman, 1974, 2014), we impose the following assumptions on the abatement costs:

5. As we will see later, our assumptions ensure that $a_i \geq 0$, i.e., a country does not increase its emissions above the initial level generated under "business as usual".

Assumption 5 (Cost Function).
Country i's abatement costs are captured by a cost function $C_i : [0, e_{i,0}] \to \mathbb{R}_+$*, which is of the quadratic form*

$$C_i(a_i) = \frac{\zeta_i}{2} a_i^2, \quad where \quad \zeta_i > 0, \qquad i = 1,2. \tag{3.4}$$

Without loss of generality, the indices are such that country 1 has weakly higher marginal abatement costs, i.e., $\zeta_1 \geq \zeta_2$. We further normalize that $\zeta_1 + \zeta_2 = 1$ to simplify the exposition. By defining $A := \sum_i a_i$ and $E_0 := \sum_i e_{i,0}$, country i's benefit writes as $B_i(A)$ and satisfies the following assumptions.

Assumption 6 (Benefit Function).
Country i's benefit function $B_i : [0, E_0] \to \mathbb{R}_+$ *is twice continuously differentiable, strictly increasing, strictly concave, and satisfies*

$$B_i(0) = 0, \quad B_i'(0) = \infty, \quad B_i'(E_0) = 0, \quad and \quad B_i'' < -\zeta_i \zeta_{-i}^2, \qquad i = 1,2.$$

The assumptions on the marginal benefits ensure that reducing emissions relative to "business as usual" is beneficial for both countries. Moreover, the assumption on the second derivative is a technical one, guaranteeing that country i's welfare function, which will be introduced shortly, is concave and thus well-behaved.

Abatements & Welfare Functions. We start by deriving fundamental insights about the realized emissions market price and the abatement activities within a given cap-and-trade system. As we consider a unit-mass continuum of homogeneous price-taking firms in each country, a firm's problem consists of minimizing its costs under the emissions trading scheme. The optimization problem of a representative firm in country i is thus

$$\min_{0 \leq a_i \leq e_{i,0}} p \cdot (e_{i,0} - a_i) + C_i(a_i), \qquad i = 1,2.$$

From the corresponding first-order condition (FOC), we directly obtain the optimal abatement level:

$$-p + C_i'(a_i) = 0 \quad \Longleftrightarrow \quad a_i = \frac{p}{\zeta_i}, \qquad i = 1,2. \tag{3.5}$$

Since overall emissions are restricted by the emissions cap $\bar{E}$, market clearing in the emissions permit market requires that $\sum_i (e_{i,0} - a_i) = \bar{E}$. Inserting the abatements per country and solving for the emissions market price yields:

$$p = p(\bar{E}) := \zeta_1 \zeta_2 \big(E_0 - \bar{E}\big). \tag{3.6}$$

Plugging (3.6) into (3.5), we obtain the abatement activities and calculate their derivatives with respect to $\bar{E}$:

$$a_i = a_i(\bar{E}) := \zeta_{-i}\big(E_0 - \bar{E}\big), \quad \text{and} \quad \frac{\partial a_i(\bar{E})}{\partial \bar{E}} = -\zeta_{-i}, \qquad i = 1,2, \tag{3.7}$$

$$A = A(\bar{E}) := E_0 - \bar{E}, \quad \text{and} \quad \frac{\partial A(\bar{E})}{\partial \bar{E}} = -1. \tag{3.8}$$

Since country 2 represents the country with lower abatement costs ($\zeta_2 \leq \zeta_1$), it contributes more to total abatement ($a_2 \geq a_1$). This is intuitive and in accordance with economic insights on emissions trading. By inserting (3.6)–(3.8) in expression (3.3), we are now in the position to define country i's welfare as a function of the design of the cap-and-trade system, i.e., the amount and allocation of permits:

$$W_i(\bar{E}, \mu_1) := B_i(A(\bar{E})) - C_i(a_i(\bar{E})) + (\bar{E}\mu_i - e_{i,0} + a_i(\bar{E}))p(\bar{E}), \qquad i = 1,2, \tag{3.9}$$
$$\text{where} \quad \mu_2(\mu_1) := 1 - \mu_1.$$

3.3 Benchmarks

We next examine two benchmark scenarios, against which we evaluate the endogenous design of the cap-and-trade system through negotiations. First, we consider a social planner who designs a joint emissions cap. Second, we investigate how each country would set its national cap individually if there was no joint cap-and-trade system. By comparing both of these scenarios, we then identify the welfare-reducing effect resulting from free-riding in our model.

Social Planner. Let us first consider a social planner who seeks to maximize the overall welfare of both countries. It is apparent that, from a social planner's perspective, trading activities between the countries offset each other, which renders the allocation μ_1 irrelevant for overall welfare. Hence, the optimization problem faced by the social planner is solely to choose a cap that maximizes welfare. Formally, the efficient cap $\bar{E}^S$ is defined by the solution to:

$$\max_{0 \leq \bar{E} \leq E_0} W(\bar{E}) := \sum\nolimits_i W_i(\bar{E}, \mu_1). \tag{3.10}$$

The FOC to the social planner's problem is

$$\frac{dW(\bar{E})}{d\bar{E}} = \sum_i \frac{dW_i(\bar{E}, \mu_1)}{d\bar{E}} = 0. \tag{3.11}$$

By inserting (3.9) and using the deviates in (3.7)–(3.8), Equation (3.11) can be simplified to

$$\sum_i B_i'(A(\bar{E})) = C_i'(a_i(\bar{E})), \qquad i = 1,2. \tag{3.12}$$

Indeed, solely balancing marginal cost of abatement with the overall marginal benefit of abatement is what determines the efficient cap. The following lemma establishes the existence and uniqueness of the efficient cap $\bar{E}^S$.

Lemma 8. *There exists a unique efficient cap* $0 < \bar{E}^S < E_0$ *that solves Problem (3.10). It is determined by the solution to (3.12).*

Intuitively, both countries benefit from a marginal increase in the abatement irrespective of where the emissions have been reduced (see l.h.s. of (3.12)). Since permits are traded, marginal abatement costs of the countries equalize (see r.h.s. of (3.12)) such that any emissions target is met at the lowest cost, making a cap-and-trade system an efficient policy instrument to regulate pollution.[6] Welfare is then maximized at the efficient emissions level $\bar{E}^S$ that balances the total marginal benefits of abatement with its marginal costs. We define the maximum level of welfare generated by the efficient cap as

$$W^S := W(\bar{E}^S).$$

National Caps. Next, we turn our attention to a scenario in which both countries do *not* participate in a joint emissions trading scheme but deploy national regulations in the form of national emissions caps instead. Since each country implements its own emissions cap, the corresponding abatement for country i and the corresponding overall abatement are of the form

$$a_i(\bar{e}_i) := e_{i,0} - \bar{e}_i, \qquad i = 1,2, \tag{3.13}$$

$$A(\bar{e}_i, \bar{e}_{-i}) := \sum_i a_i(\bar{e}_i) = E_0 - \sum_i \bar{e}_i. \tag{3.14}$$

6. This holds true in the absence of transaction costs and imperfect competition (see, e.g., Hahn, 1984; Stavins, 1995).

Given the cap of the other country, country i now chooses its own cap to maximize its welfare, i.e., by solving

$$\max_{0 \leq \bar{e}_i \leq e_{i,0}} B_i(A(\bar{e}_i, \bar{e}_{-i})) - C_i(a_i(\bar{e}_i)), \qquad i = 1,2. \quad (3.15)$$

Using (3.13) and (3.14), country i's FOC can be written as

$$B_i'(A(\bar{e}_i, \bar{e}_{-i})) = C_i'(a_i(\bar{e}_i)), \qquad i = 1,2. \quad (3.16)$$

Lemma 9 now establishes the existence and uniqueness of a Nash equilibrium in this abatement game.

Lemma 9. *For each country $i \in \{1,2\}$, there exists a unique Nash equilibrium cap $0 \leq \bar{e}_i^N < e_{i,0}$ that solves Problem (3.15). The first inequality is strict for at least one country.*

Intuitively, in the Nash equilibrium, the cap $\bar{e}_i^N$ chosen country i is the best response – by equating marginal benefits and costs – to the cap $\bar{e}_{-i}^N$ implemented by the other country. Hence, no country has an incentive to deviate. The overall welfare generated by national caps is defined by

$$W^N := \sum_i W_i^N, \quad \text{where} \quad W_i^N := B_i(A(\bar{e}_i^N, \bar{e}_{-i}^N)) - C_i(a_i(\bar{e}_i^N)), \qquad i = 1,2.$$

Comparison. Naturally, the following question arises: how effective are national caps in reducing emissions and improving welfare in comparison to the efficient outcome generated by the social planner? The following proposition answers this question by comparing both benchmark scenarios in terms of implemented overall cap and welfare.

Proposition 8 (National Caps Versus Social Planner).
Compared to a social planner, national caps implement a strictly higher overall cap, $\sum_i \bar{e}_i^N > \bar{E}^S$, that leads to a strictly lower level of welfare, $W^N < W^S$.

Proposition 8 is the manifestation of the free-riding problem in our model. Intuitively, each country i has an incentive to free-ride on the abatement carried out by the other country, while, at the same time, implementing insufficient domestic abatement targets to reduce its own abatement costs. This results in total emissions that are too high from a societal perspective. Example 9 illustrates this finding.

Example 9 (Free-Riding).
Consider benefits that are captured by a benefit function of the form

$$B_i(A) = \beta_i(2\sqrt{A} - A), \quad \textit{where} \quad \beta_i > 2\zeta_i\zeta_{-i}^2, \quad \textit{and} \quad A \in [0,1].$$

Even for the symmetric case where $e_{1,0} = e_{2,0} = 0.5$*,* $\zeta_1 = \zeta_2 = 0.5$*, and* $\beta_1 = \beta_2 = 0.3$*, free-riding leads to emissions that are 35% above the efficient emissions cap.*

3.4 Negotiations on the Emissions Cap

Our analysis of the benchmarks raises the question of whether a cap-and-trade system can still overcome or at least mitigate the free-riding incentive in the absence of a social planner. In other words, is it more favorable from a societal perspective if the countries design the cap-and-trade system themselves rather than implementing national caps?

Probably the most natural way for countries to endogenously design the cap-and-trade system is through negotiations, which we will analyze next. To understand the basic mechanisms, it is illustrative to start with the simplest form of negotiations, where the countries only bargain over the emissions cap. That is, we consider a setting where the countries have already *(i)* agreed on setting up a joint cap-and-trade system and *(ii)* determined an initial allocation of permits $\hat{\mu}_1 \in [0,1]$.

(i) means that neither country can strategically opt out of the negotiation, which, for our bargaining model, implies that no country has an outside option. One rationale for this situation could be that governments have already committed to establishing a joint cap-and-trade system or that public pressure is forcing them to do so.

(ii) entails that the allocation of permits is exogenous from the perspective of the negations on the emissions cap. This allocation could, for instance, be determined through prior negotiations or *grandfathering*, i.e., proportional to the countries' emissions under "business as usual" going back to the concept of first possession and appropriation (cf. Epstein, 1979; Lueck, 1995; Rose, 1985).[7] As grandfathering is frequently practiced in emissions trading systems such as the US sulphur dioxide emissions trading program and the EU ETS (Woerdman et al., 2008), the assumption of an exogenous allocation is plausible from a practical point of view.

7. Analyses of grandfathering from an economic perspective can be found, for example, in Böhringer and Lange (2005), Damon et al. (2019), as well as Grimm and Ilieva (2013).

Nonetheless, it should be emphasized that *(i)* and *(ii)* will be dropped in the course of this paper in order to derive more general insights.

3.4.1 Bargaining Outcome

First, we need to specify the set of feasible bargaining solutions. It is straightforward that any bargaining has to result in a Pareto-efficient outcome. Otherwise, the parties could simply agree on another cap and thereby achieve a Pareto improvement. To construct the set of Pareto-efficient caps, $\mathscr{P}_{\hat{\mu}_1}$, it is necessary to determine which joint cap $\bar{E}_i$ country i prefers most as an outcome of the bargaining procedure. Country i would set a global cap that maximizes its welfare,

$$\max_{0\leq\bar{E}\leq E_0} \; W_i(\bar{E},\hat{\mu}_1), \qquad i = 1,2. \quad (3.17)$$

From differentiating (3.9) and simplifying, we obtain country i's FOC, which reads

$$-B_i'(A(\bar{E})) + \mu_i C_i'(a_i(\bar{E})) + x_{i,-i}(\bar{E}) \; C_i''(a_i(\bar{E}))\frac{\partial a_i(\bar{E})}{\partial \bar{E}} = 0, \qquad i = 1,2, \quad (3.18)$$
$$\text{where} \quad x_{i,-i}(\bar{E}) := \left(\mu_i\bar{E} - e_{i,0} + a_i(\bar{E})\right).$$

As (3.18) shows, from an individual perspective, trading activities *and* the initial distribution of permits matter for welfare. Indeed, $x_{i,-i}$ is the amount of permits passed from country i to $-i$, which can be both positive and negative depending on whether i sells or purchases permits from country $-i$. The l.h.s. of Equation (3.18) reveals three marginal effects that a higher cap has on country i's welfare. There are effects on marginal benefits (first summand) and marginal costs (second summand), as well as a trading effect (third summand). The following lemma now establishes the existence and uniqueness of a solution to Problem (3.17) and compares it to the social planner solution.

Lemma 10. *For each country $i \in \{1,2\}$, there exists a unique individually optimal cap $0 \leq \bar{E}_i < E_0$ that solves Problem (3.17). The first inequality is strict for at least one country. Moreover, it holds that min $\{\bar{E}_1, \bar{E}_2\} \leq \bar{E}^S \leq$ max $\{\bar{E}_1, \bar{E}_2\}$.*

Lemma 10 reveals that at most one country advocates complete decarbonization, while the other country prefers a positive level of global emissions. Positive caps are determined by the FOC (3.18) to balance marginal benefits, costs, and trading effects. A social planner, in comparison, would cap overall emissions at a level that lies between those caps optimal from an individual perspective. Since both countries' welfare functions are strictly concave in the implemented cap, we can directly use Lemma 10 to define the Pareto set $\mathscr{P}_{\hat{\mu}_1}$.[8]

Definition 7 (Pareto-Efficient Caps).
The set of Pareto-efficient caps, $\mathscr{P}_{\hat{\mu}_1} \subset \mathbb{R}_+$, is defined by

$$\mathscr{P}_{\hat{\mu}_1} := \left\{\bar{E} : \bar{E} \in \left[\min\{\bar{E}_1, \bar{E}_2\}, \max\{\bar{E}_1, \bar{E}_2\}\right]\right\}.$$

Figure 3.1 illustrates the idea of Definition 7 for the case where $\bar{E}_1 < \bar{E}_2$. Caps lower than $\bar{E}_1$ are not Pareto-efficient since a marginal increase in $\bar{E}$ results in a Pareto improvement, whereas for caps larger than $\bar{E}_2$, a Pareto improvement can be achieved by reducing $\bar{E}$. Only in the shaded area in the closed interval $[\bar{E}_1, \bar{E}_2]$, we find Pareto-efficient caps. An increase in the emissions cap is detrimental to country 1 here, whereas it benefits country 2.

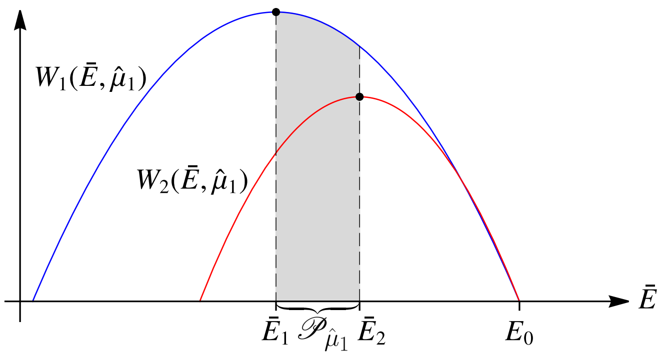


Figure 3.1. Individually optimal caps and Pareto set.

As a next step, we formalize the bargaining procedure. One of the most straightforward and intuitive ways to model a cap negotiation is according to Rubinstein's (1982) alternating-offers model, which applies to our setting as follows.[9]

Country i proposes a cap. Then country $-i$ can either accept this offer and the game ends or reject the offer and make a counteroffer after $\Delta > 0$ time units. In case of rejection, it is i's turn to decide whether to accept the counteroffer or to make a counter-counteroffer. This process continues until one country accepts the

8. Formally, concavity of W_i is shown as part of the proof of Lemma 10 in the Appendix 3.A.1.
9. See Osborne and Rubinstein (1990) and Muthoo (1999) for textbook as well as Roth (1985) and Binmore and Dasgupta (1987) for advanced treatments of bargaining theory.

proposed cap.[10] A prominent result in bargaining theory is that the subgame perfect equilibrium in the Rubinstein model converges to Nash's (1950) bargaining solution if $\Delta \to 0$ (Binmore, 1987; Binmore et al., 1986). Intuitively, in the words of Muthoo (1999, p. 52), $\Delta \to 0$ corresponds to a situation where *"the absolute magnitudes of the frictions in the bargaining process are small"*. Evidently, this is in accordance with our setup, as the bargaining process is substantially faster than the underlying process of climate change that requires the reduction of emissions. Even if the bargaining is extended by Δ due to the rejection of an offer, approximately the same benefits and costs can be attained through an agreement in the next round. For simplicity, we assume that the counties have the same time discount rate such that we can apply the symmetric Nash bargaining solution.[11] In our setting, the Nash bargaining solution $\bar{E}^B$ is defined as the solution to the following maximization problem:

$$\max_{\bar{E}} \mathcal{N}_{\hat{\mu}_1}(\bar{E}) := W_1(\bar{E}, \hat{\mu}_1) \cdot W_2(\bar{E}, \hat{\mu}_1), \qquad \text{s.t. } \bar{E} \in \mathcal{P}_{\hat{\mu}_1}, \tag{3.19}$$

where $\mathcal{N}$ is referred to as Nash product.[12] Figure 3.2 illustrates how the bargaining solution is determined. The purple line represents the Nash product. The Nash bargaining solution is the maximizer of the Nash product among the Pareto-efficient caps, which are represented by the solid part of the purple line.

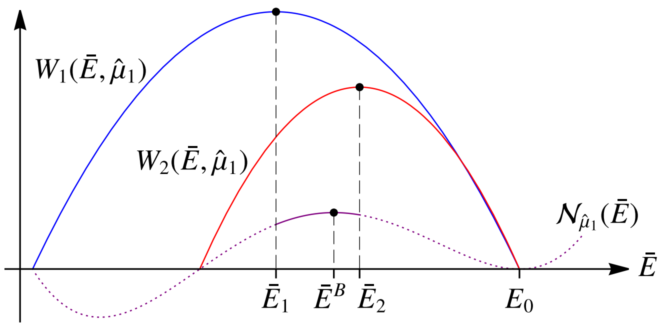


Figure 3.2. Nash bargaining solution.

10. Note that this standard version of the alternating-offers model does not incorporate strategic opting out of the bargaining. In our setup, the interpretation is that, while the counties have already agreed on creating a cap-and-trade system, they only bargain about the implemented cap.

11. Different discount rates shift bargaining power in favor of country i that possesses a lower discount rate, i.e., that is more patient. This leads to a bargaining outcome that is close to $\bar{E}_i$.

12. More precisely, $\mathcal{N}_{\hat{\mu}_1}(\bar{E}) = \big(W_1(\bar{E}, \hat{\mu}_1) - d_1\big) \cdot \big(W_2(\bar{E}, \hat{\mu}_1) - d_2\big)$. As d_i reflects welfare attained by country i if "business as usual" is maintained, it holds that $d_i = 0$ for all $i \in \{1, 2\}$ (cf. Binmore et al., 1986; Muthoo, 1999).

From differentiation, we obtain the FOC of the Nash product, which reads:

$$\frac{\mathrm{d}\mathcal{N}_{\hat{\mu}_1}(\bar{E})}{\mathrm{d}\bar{E}} = \sum_i \frac{\mathrm{d}W_i(\bar{E},\hat{\mu}_1)}{\mathrm{d}\bar{E}} \cdot W_{-i}(\bar{E},\hat{\mu}_1) = 0. \tag{3.20}$$

In our analysis, we exploit the following lemma.

Lemma 11. *There exists a unique Nash bargaining solution $0 \le \bar{E}^B < E_0$ that solves Problem (3.19). Any $\bar{E}^B > 0$ is determined by the unique solution to (3.20) in $\mathcal{P}_{\hat{\mu}_1}$.*

Although (3.20) has multiple solutions (cf. Figure 3.2), there exists at most one solution that is Pareto-efficient. If that solution indeed exists, then it defines the bargaining outcome $\bar{E}^B > 0$, while otherwise $\bar{E}^B = 0$ holds. Lemma 11 greatly helps us investigate the question of whether bargaining can implement the efficient cap or, more generally, whether the bargaining outcome may be welfare-enhancing compared to national caps. Since it was shown in Lemma 8 and 10 that $\bar{E}^S > 0$ and $\bar{E}^S \in \mathcal{P}_{\hat{\mu}_1}$, respectively, we can conclude that the bargaining procedure implements the efficient cap if and only if $\bar{E}^S$ solves (3.20).

3.4.2 Comparison to the Benchmarks

It is worth emphasizing that each total abatement in the joint cap-and-trade system is achieved with the optimal cost structure, namely with equal marginal cost in each country. Hence, if $\bar{E}^B = \bar{E}^S$, then this automatically implies that the bargaining solution leads to the greatest level of overall welfare. Proposition 9 now explores conditions under which bargaining indeed implements $\bar{E}^S$.

Proposition 9 (Bargaining over the Emissions Cap).
Consider two countries that agreed upon setting up a joint cap-and-trade system with an initial allocation $\hat{\mu}_1$ and bargain over the amount of permits. Then the following holds:

(i) If the countries are sufficiently symmetric in terms of benefits, costs, and initial emissions, then there exists a unique initial allocation $\mu_1^S \in [0,1]$ which yields $\bar{E}^B = \bar{E}^S$.
(ii) If, by contrast, $B_i(A(\bar{E}^S)) - B_{-i}(A(\bar{E}^S))$ sufficiently large, then $\bar{E}^B \neq \bar{E}^S$.

The intuition behind Proposition 9 is due to the efficiency-fairness trade-off in bargaining (see, e.g., Bertsimas et al., 2012; Dijkstra and Nentjes, 2020; Freeborn, 2023). The countries generally face a trade-off between “size of the cake” and “allocation of the cake”. On the one hand, they seek to maximize the overall welfare level that they can divide among themselves, i.e., they want to choose a cap close to $\bar{E}^S$. On the other hand, due to equal bargaining power, the countries want to implement a cap that leads to an equal split, i.e., that equalizes the counties' individual welfare levels.

Proposition 9 *(i)* reveals that if the countries are sufficiently symmetric in terms of cost- and benefit structures and initial emissions, then a unique allocation $\mu_1^S \in [0, 1]$ exists that completely resolves this trade-off. Given μ_1^S, agreeing on the efficient cap not only maximizes overall welfare but also induces equal welfare levels in both countries. In light of Proposition 8, this implies that if the countries initially agreed on an allocation of permits $\hat{\mu}_1$ sufficiently close to μ_1^S, then determining a joint cap via negotiations is indeed welfare-improving, since

$$W^S \geq W(\bar{E}^B) > W^N,$$

holds by continuity for $\hat{\mu}_1$ sufficiently close to μ_1^S. Put differently, if the countries are sufficiently symmetric and $\hat{\mu}_1$ is in the local neighborhood of μ_1^S, then endogenously designing the cap-and-trade system through cap negotiations alleviates or completely overcomes the free-ride incentive by implementing a stricter emissions cap and enhancing overall welfare compared to national caps. The welfare maximum W^S is however only achieved if the initial allocation $\hat{\mu}_1$ coincides with μ_1^S.

Proposition 9 *(ii)* states that if the countries' benefits obtained under the efficient cap are too different, then bargaining does not implement the social optimum. Intuitively, as indicated by Buchholz et al. (2005) and Caparrós (2016), the allocation of permits serves as an implicit side payment in an emissions trading system: with a higher share μ_i, country i has to purchase fewer certificates or receives additional revenue for selling the certificates, depending on whether i acts as buyer or seller of permits. However, the scope for providing side payments is limited as country i cannot receive more than the entire share of permits or less than no share. If the countries' benefits obtained under the efficient cap are too different, then the scope for providing implicit side payments is insufficient to fully resolve the trade-off, i.e., there is no allocation of permits for which the efficient cap also yields equal individual welfare levels. Hence, due to the prevailing trade-off, countries forego choosing the efficient cap and instead agree on a cap that leads to a more equal distribution of individual welfare levels.

Since an efficient emissions cap is implemented through an allocation that resolves the trade-off rather than allocating permits proportionally to initial emission levels, we can state the following corollary for grandfathering.

Corollary 4 (Grandfathering)**.**
Consider two countries that agreed upon setting up a joint cap-and-trade system with an initial allocation $\hat{\mu}_1$ and bargain over the amount of permits.
If $\hat{\mu}_1 \propto e_{1,0}/E_0$, then bargaining will generally lead to an inefficient emissions cap.

Example 10 illustrates our results thus far regarding the bargaining outcome.

Example 10 (Optimal Initial Allocation)**.**
Revisit the symmetric parameterization of Example 9. Figure 3.3 depicts the initial allocation μ_1^S for different β_2 and $\beta_1 = 0.3$. For completely symmetric countries, ($\beta_2/\beta_1 = 1$), bargaining implements the social optimum if and only if the countries initially agreed on an equal distribution of permits, $\hat{\mu}_1 = 0.5$. By contrast, for $\beta_2/\beta_1 < 0.8$ and $\beta_2/\beta_1 > 1.2$, the means of implicit side payments are insufficient to implement the efficient cap. Grandfathering does not implement the efficient cap except for the special case where $\beta_2/\beta_1 = 1$.

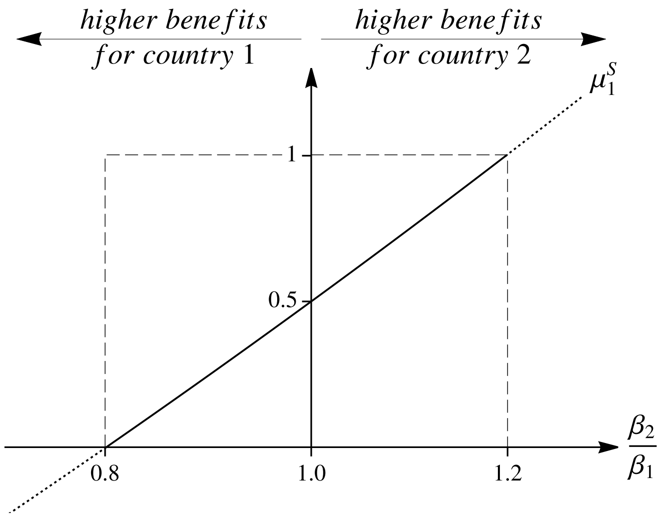


Figure 3.3. Optimal initial allocation.
($e_{1,0} = e_{2,0} = 0.5$, $\zeta_1 = \zeta_2 = 0.5$, and $\beta_1 = 0.3$)

3.5 Negotiations on the Cap and Allocation

Equipped with the inside that bargaining can be welfare-improving, we add a further degree of freedom by allowing countries to negotiate simultaneously on the emissions cap and the initial allocation of permits. We maintain the assumption that countries cannot strategically end the negotiation for the time being but will abandon this assumption in the course of this section. The Pareto set $\mathscr{P}$ now consists of tuples $(\bar{E}, \mu_1)$, i.e., combinations of a joint emissions cap and a corresponding allocation of these certificates among the countries. It can be defined as follows.

Definition 8 (Pareto-Efficient Tuples).
The set of Pareto-efficient tuples, $\mathscr{P} \subset \mathbb{R}_+ \times [0,1]$, is the set of all tuples $(\bar{E}, \mu_1)$ for which no other tuple $(\bar{E}', \mu_1')$ exists that satisfies

$$W_i(\bar{E}', \mu_1') \geq W_i(\bar{E}, \mu_1), \quad \textit{and} \quad W_{-i}(\bar{E}', \mu_1') > W_{-i}(\bar{E}, \mu_1)$$

for at least one $i \in \{1,2\}$, where $\mu_1, \mu_1' \in [0,1]$.

For the bargaining procedure, this implies that an offer in the alternating-offers model now consists of a tuple $(\bar{E}, \mu_1)$, i.e., a proposal about the amount of permits and their allocation among the countries. Due to the negligible friction in the bargaining process, we can exploit the relation between the subgame perfect equilibrium in this dynamic game and the static Nash bargaining approach again. Indeed, the bargaining solution $(\bar{E}^B, \mu_1^B)$ is defined by a solution to the following maximization problem:

$$\max_{(\bar{E}, \mu_1)} \mathscr{N}(\bar{E}, \mu_1) := W_1(\bar{E}, \mu_1) \cdot W_2(\bar{E}, \mu_1) \qquad \text{s.t. } (\bar{E}, \mu_1) \in \mathscr{P}. \tag{3.21}$$

Analyzing the Problem (3.21) leads to the following proposition.

Proposition 10 (Bargaining over the Emissions Cap and Initial Allocation).
Consider two countries that agreed upon setting up a joint cap-and-trade system and bargain over the amount and allocation of permits. Then the following holds:

(i) If the countries are sufficiently symmetric in terms of benefits, costs, and initial emissions, then $\bar{E}^B = \bar{E}^S$ and $\mu_1^B = \mu_1^S$.
(ii) If, by contrast, $B_i(A(\bar{E}^S)) - B_{-i}(A(\bar{E}^S))$ sufficiently large, then $\bar{E}^B \neq \bar{E}^S$.

Allowing the countries to negotiate simultaneously on the allocation of permits and the emissions cap yields some interesting results. Proposition 10 *(i)* shows that whenever the scope for setting side payments allows the countries to resolve the trade-off and implement the efficient cap, they will, in fact, design an allocation of permits to do so.[13] Put differently, if the countries are sufficiently symmetric, then setting up a cap-and-trade system and letting the countries bargain over the amount and allocation of permits completely removes the distortions created by the

13. Technically, the following relation holds $\bar{E}^B = \bar{E}^S \iff \mu_1^B = \mu_1^S \iff \mu_1^S \in [0,1]$.

free-riding incentive. It is worth emphasizing that the countries themselves then design an emissions cap exactly as it would have been done by a social planner or regulator with complete information.

Proposition 10 *(ii)* is due to the mechanism that we have already encountered: if the countries' benefits from the efficient cap are too heterogeneous, then the scope for providing side payments is insufficient to implement the efficient cap. This holds true irrespective of whether the allocation of permits is exogenously given or endogenously determined via bargaining.

It is worth discussing our results against the background of the famous Coase Theorem. As summarized by Harris and Roach (2022, p. 60), the theorem states that *"if property rights are well defined, and no significant transaction costs exist, an efficient allocation of resources will result even with externalities."* In our model, an efficient cap obtains, except for the case where the scope for side payments is exhausted, which can be interpreted as infinite transaction costs. While our results are thus in line with the Coase Theorem, it should be stressed that they are *not* a mere consequence of the theorem. First, property rights in our setup are ex ante not well defined, as the exact purpose of the negotiation is to determine these rights by specifying the amount of permits and their allocation among the countries. Second, even if this prerequisite was satisfied, Hahnel and Sheeran (2009) argue that formal bargaining models will not generally result in the efficient outcome predicted by the rather informal Coase Theorem: whether negotiations lead to an efficient outcome crucially depends on the bargaining procedure, countries' welfare functions and time preferences, as well as the information structure in the game.

The preceding analysis focused exclusively on situations where the countries were unable to terminate the negotiation strategically. While plausible for some settings, others may allow each side to strategically opt out and end the bargaining in disagreement. Metaphorically speaking, if one party decides to leave the negotiation table, then both parties are left with their *outside option*. Since the countries must rely on national caps in case setting up a joint cap-and-trade system fails, they are left with the Nash equilibrium caps described in Lemma 9. Hence, country i's outside option is simply W_i^N, i.e., the welfare level obtained in the national cap benchmark. As the friction in the bargaining process is negligibly small, insights from bargaining theory allow us to link the Nash bargaining solution to subgame perfect equilibrium in the alternating offers model extended by the possibility of strategic opting out for both parties. Following Binmore (1985), Binmore et al. (1986), and Muthoo (1999), the subgame perfect equilibrium in the extended alternating-offers model converges to the solution to the following maximization problem:

$$\max_{\bar{E},\mu_1} \mathcal{N}(\bar{E},\mu_1) = W_1(\bar{E},\mu_1) \cdot W_2(\bar{E},\mu_1)$$
$$\text{s.t.} \quad (\bar{E},\mu_1) \in \mathscr{P}, \quad W_1(\bar{E},\mu_1) \geq W_1^N, \quad W_2(\bar{E},\mu_1) \geq W_2^N. \tag{3.22}$$

The only difference to Problem (3.21) is that each country can secure itself a welfare level weakly greater than its outside option. Intuitively, a country would never accept a "bad" offer in the negotiation but instead strategically opt out and realize its outside option. Examining Problem (3.22) yields the following proposition.

Proposition 11 (Bargaining with Outside Option).
Consider two countries that bargain on setting up a joint cap-and-trade system. If the countries are sufficiently symmetric in terms of benefits, costs and initial emissions, then they agree on setting up a joint cap-and-trade system with $\bar{E}^B = \bar{E}^S$ *and* $\mu_1^B = \mu_1^S$.

Proposition 11 states if two countries are sufficiently symmetric, instead of implementing national caps, they will agree on setting up a joint cap-and-trade system with an emissions cap that is efficient. This leads to the highest possible level of welfare, which is equally distributed among the countries via the allocation of permits. Our analysis, therefore, points out a simple way to circumvent the inefficiency caused by the free-riding incentive: countries should be enabled to set up joint cap-and-trade systems and allowed to negotiate the cap and allocation of permits. This procedure then implements the social planner result, provided that the countries are sufficiently symmetric. Otherwise, the free-riding incentive might be so strong for a country and its outside option thus so attractive that the negotiated joint emissions cap is distorted away from the social optimum. This is illustrated in the following example.

Example 11 (Bargaining with Outside Option).
Revisit the symmetric parameterization of Example 9, where $\beta_1 = 0.3$.

- *(i) For* $\beta_2/\beta_1 \in [0.8, 0.97)$ *bargaining implements* $\bar{E}^S$*, while* μ_1 *is determined by providing a welfare level to country 1 that equals its outside option.*[14]
- *(ii) For* $\beta_2/\beta_1 \in [0.97, 1.03]$ *bargaining implements* $\bar{E}^S$ *and* μ_1 *is determined by equalizing the corresponding welfare levels in both countries.*
- *(iii) For* $\beta_2/\beta_1 \in (1.03, 2.05]$ *bargaining implements* $\bar{E}^S$*, while* μ_1 *is determined by providing a welfare level to country 2 that equals its outside option.*
- *(iv) By contrast, for* $\beta_2/\beta_1 > 2.05$ *bargaining does not implement* $\bar{E}^S$.

14. We do not examine ratios $\beta_2/\beta_1 < 0.8$, as they violate Assumption 6.

The example shows that the problem of insufficient side payments carries over to the presence of outside options. Even more surprisingly, comparing Examples 10 and 11 reveals that the presence of an outside option may ensure that the bargaining leads to an efficient cap for more asymmetric countries than it would be the case without an outside option. This is precisely the case for β_2/β_1 ratios between 1.2 and 2.05. We summarize this surprising finding in the following proposition.

Proposition 12 (Presence vs. Absence of an Outside Option).
If countries have the possibility of strategic opting out, then bargaining may implement an efficient cap for more asymmetric countries than it would be the case without an outside option.

The intuition is as follows. If country i's outside option is sufficiently attractive, then this eliminates the efficiency-fairness trade-off since the welfare level granted to country i equals its outside option. Whenever possible, the parties then agree on the efficient cap and an allocation of permits that provides county i with the welfare level of its outside option.[15] This bargaining outcome not only ensures county i's participation in the scheme but also maximizes the welfare left for country $-i$.

3.6 Multilateral Negotiations

Although our analysis was conducted in the two-country case for the sake of clarity, it easily generalizes to the case of n countries, where $n > 2$. Analogously to the two-country case without an outside option, bargaining implements the efficient cap if and only if a *feasible* allocation of permits exists that equates the welfare levels in all countries resulting from the efficient cap.[16] A feasible allocation is now characterized by

$$\sum_i \mu_i = 1, \qquad \text{and} \qquad \mu_i \in [0,1], \qquad\qquad i = 1, \dots, n.$$

In the symmetric case, allocating equal shares $\mu_i = 1/n$ to all countries n indeed constitutes the unique feasible allocation that equates the welfare levels from the

15. However, as can be seen in Part *(iv)* of Example 11, if the countries are too different, then there may not be a feasible allocation of the efficient amount of permits that provides country i with the welfare level of its outside option. In this case, the countries will not agree on the efficient cap but rather on one closer to country i's individually optimal cap $\bar{E}_i$.

16. Technically, the FOC of the social planner and the FOC of the Nash product with respect to the emissions cap coincide in this case for the efficient cap.

efficient cap such that bargaining leads to a design of the cap-and-trade system that reduces emissions to the efficient level. Hence, continuity implies that bargaining also implements the efficient cap if the countries are sufficiently symmetric. If, by contrast, $B_i(A(\bar{E}^S)) - B_j(A(\bar{E}^S))$ is sufficiently large for at least two countries i and j, then equating welfare levels resulting from the efficient cap would either require $\mu_i < 0$ or $\mu_j > 1$ such that no feasible allocation exists to do so. Bargaining will therefore not implement the efficient cap in this case. In the presence of an outside option, it additionally holds for symmetric countries that

$$W_i(\bar{E}^S, 1/n) = \frac{W^S}{n} > \frac{W^N}{n} = W_i^N, \qquad i = 1, \ldots, n.$$

We can thus infer that even with an outside option if the n countries are sufficiently symmetric, they will agree on setting up a cap-and-trade system and cap the emissions at the efficient level. Accordingly, our analysis carries over entirely to the case with more than two countries, i.e., multilateral negotiations on the design of a joint cap-and-trade system.

3.7 Conclusion

How can global greenhouse gas emissions be reduced to mitigate climate change? We have addressed this question by analyzing whether designing an emissions trading scheme through negotiation has the potential to enforce efficient emissions levels. Our analysis builds on a simple model with two countries that experience a positive externality from reducing emissions. Due to this externality, each country has an incentive to free-ride on the other country's abatement activities. In the case of national abatement activities, free-riding leads to an overall emissions level that exceeds the social optimum.

By applying insights from bargaining theory, we find that the ecological market failure resulting from the free-riding incentive may be eliminated by a simple mechanism derived from quantity-based instruments: enabling the countries to set up a joint cap-and-trade system and allowing them to bargain over the amount and allocation of certificates. If the countries are sufficiently symmetric, they agree to cap emissions at the efficient level. Since this efficient emissions level obtains with the optimal distribution of abatement activities among the countries, i.e., at the lowest cost, the endogenous cap maximizes overall welfare. The countries then use the allocation of certificates as an implicit side payment to distribute welfare equally among themselves. Surprisingly, an efficient cap may also be achieved for even more asymmetric countries if they can strategically terminate the bargaining and deploy national caps instead. However, if the countries are too different, then bargaining may not necessarily result in the efficient cap. In this case, the scope for implicit side payments through the initial allocation of certificates may be insufficient to make both countries agree on the efficient cap.

The implications of our analysis are quite striking. Even in the absence of a social planner, a joint cap-and-trade system may induce efficient outcomes. The sheer possibility of negotiating its design then induces cooperative behavior: the countries overcome the free-riding incentives, implement the efficient emissions cap, and distribute the resulting overall welfare among themselves in a fair way. Our results imply that negotiations are pivotal in efficiently designing cap-and-trade systems and should thus be encouraged. Moreover, they underline the importance of removing all sorts of barriers, such as transaction costs (cf. Montero, 1998), imperfections in the emissions market (cf. Stavins, 2008b), and conflicting national regulations (cf. Hahn and Stavins, 2011), that either hinder countries from setting up a joint cap-and-trade system or prevent them from linking existing schemes. However, designing a mechanism that implements efficient outcomes for strongly asymmetric countries, especially in the presence of outside options, is significantly more complex and constitutes a fruitful avenue for further research.

Appendix 3.A

The proofs for all of our main results are provided in Appendix 3.A.1. The (very) preliminary version of the project from Naumann's (2022) dissertation can be found in Appendix 3.A.2. A comparison to the version in this thesis shows how much the project has evolved over the last two years.

3.A.1 Proofs of the Main Results

Proof of Lemma 8. First, we establish the existence of a solution to (3.12). Since

$$\begin{aligned}\lim_{\bar{E}\to 0}\sum\nolimits_i B_i'(A(\bar{E})) - C_i'(a_i(\bar{E})) &= -\zeta_1\zeta_2 E_0 < 0,\\ \lim_{\bar{E}\to E_0}\sum\nolimits_i B_i'(A(\bar{E})) - C_i'(a_i(\bar{E})) &= \sum\nolimits_i B_i'(0) = \infty > 0,\end{aligned}$$

the Intermediate Value Theorem immediately implies the existence of a solution. Moreover, differentiation with respect to $\bar{E}$ yields

$$\sum\nolimits_i B_i''(A(\bar{E}))\frac{\partial A(\bar{E})}{\partial \bar{E}} - C_i''(a_i(\bar{E}))\frac{\partial a_i(\bar{E})}{\partial \bar{E}},$$

which, using (3.7)–(3.8), simplifies to

$$-\sum\nolimits_i B_i''(A(\bar{E})) + \zeta_1\zeta_2 > 0.$$

Hence, the solution is unique. Note also that our assumptions regarding the functional form of the cost and benefit functions immediately imply that $W(\bar{E})$ is strictly concave. Therefore, the solution $\bar{E}^S$ to the FOC is indeed a maximizer. □

Proof of Lemma 9. In the Nash equilibrium, both countries choose national caps that are best responses to each other. To derive the best responses, note that country i's objective function in (3.15) is concave in $\bar{e}_i$ such that if a solution to the FOC exists, then the solution indeed maximizes country i's welfare.

We consider the best response of country 1 first by investigating the limits of its FOC for a given $\bar{e}_2$

$$\begin{aligned}\lim_{\bar{e}_1\to e_{1,0}} B_1'(A(\bar{e}_1,\bar{e}_2)) - C_1'(a_1(\bar{e}_1)) &= B_1'(A(e_{1,0},\bar{e}_2)) > 0,\\ \lim_{\bar{e}_1\to 0} B_1'(A(\bar{e}_1,\bar{e}_2)) - C_1'(a_1(\bar{e}_1)) &= B_1'(A(0,\bar{e}_2)) - \zeta_1 e_{1,0} \gtreqless 0.\end{aligned}$$

Now, we need to distinguish two cases:

Case 1. If $B_1'(A(0,\bar{e_2}))-\zeta_1 e_{1,0}\geq 0$, then $\bar{e}_1^N=0$. Since, for country 2, it holds that

$$\lim_{\bar{e}_2\to e_{2,0}} B_2'(A(\bar{e}_2,0))-C_2'(a_2(\bar{e}_2)) \;=\; B_2'(A(e_{2,0},0)) \;>\; 0,$$
$$\lim_{\bar{e}_2\to 0} B_2'(A(\bar{e}_2,0))-C_2'(a_2(\bar{e}_2)) \;=\; -\zeta_2 e_{2,0} \;<\; 0.$$

The Intermediate Value Theorem implies the existence of a solution $0<\bar{e}_2^N<e_{2,0}$ to country 2's FOC. Moreover, differentiation with respect to $\bar{e}_2$ and simplifying yields

$$-B_2''(A(\bar{e}_2,0))+C_2''(a_2(\bar{e}_2)) \;=\; -B_2''(A(\bar{e}_2,0))+\zeta_2 \;>\; 0.$$

Hence, $\bar{e}_2^N$ is unique. In this case, the Nash equilibrium caps are thus uniquely determined and satisfy $\bar{e}_1^N=0$ and $0<\bar{e}_2^N<e_{2,0}$.

Case 2. If $B_1'(A(0,\bar{e_2}))-\zeta_1 e_{1,0}>0$, then the Intermediate Value Theorem implies the existence of a solution to country 1's FOC. Again, differentiating with respect to $\bar{e}_1$ and simplifying yields

$$-B_1''(A(\bar{e}_1,\bar{e}_2))+C_1''(a_1(\bar{e}_1)) \;=\; -B_1''(A(\bar{e}_1,\bar{e}_2))+\zeta_1 \;>\; 0.$$

such that the FOC has a unique solution that defines a best response function of the form $\bar{e}_1(\bar{e}_2)$. Plugging $\bar{e}_1(\bar{e}_2)$ into country 2's FOC and analyzing the limits yields

$$\lim_{\bar{e}_2\to e_{2,0}} B_2'(A(\bar{e}_2,\bar{e}_1(\bar{e}_2)))-C_2'(a_2(\bar{e}_2)) \;=\; B_2'(A(e_{2,0},\bar{e}_1(e_{2,0}))) \;>\; 0,$$
$$\lim_{\bar{e}_2\to 0} B_2'(A(\bar{e}_2,\bar{e}_1(\bar{e}_2)))-C_2'(a_2(\bar{e}_2)) \;=\; B_2'(A(0,\bar{e}_1(0)))-\zeta_2 e_{2,0} \;\gtreqless\; 0.$$

Two subcases need to be distinguished now:

Case 2.1. If $B_2'(A(0,\bar{e}_1(0)))-\zeta_2 e_{2,0}\geq 0$, then $\bar{e}_2^N=0$. Again, the Nash equilibrium caps are unique and satisfy $0<\bar{e}_1^N=\bar{e}_1(0)<e_{1,0}$ and $\bar{e}_2^N=0$.

Case 2.2. If $B_2'(A(0,\bar{e}_1(0)))-\zeta_2 e_{2,0}<0$, then, by the Intermediate Value Theorem, there exists a solution to country 2's FOC. Differentiation with respect to $\bar{e}_2$ and simplifying yields

$$-B_2''(A(\bar{e}_2,\bar{e}_1(\bar{e}_2)))\Big[1+\frac{\partial\bar{e}_1(\bar{e_2})}{\partial\bar{e}_2}\Big]+\zeta_2 \tag{3.A.1}$$

From differentiating country 1's FOC and simplifying, we get that

$$\frac{\partial\bar{e}_1(\bar{e_2})}{\partial\bar{e}_2} \;=\; \frac{B_1''(A(\bar{e}_1,\bar{e}_2))}{-B_1''(A(\bar{e}_1,\bar{e}_2))+\zeta_1}\in(-1,0) \qquad \text{for } 0<\bar{e}_1<e_{1,0}. \tag{3.A.2}$$

In view of (3.A.2), we find that the term in (3.A.1) is strictly positive. Hence, the solution to country 2's FOC is unique. Therefore, the unique Nash equilibrium caps in this case are $0 < \bar{e}_1^N = \bar{e}_1(\bar{e}_2^N) < e_{1,0}$ and $0 < \bar{e}_2^N < e_{2,0}$. □

Proof of Proposition 8. We distinguish two cases depending on whether $0 < \bar{e}_1^N, \bar{e}_2^N$ or $0 < \bar{e}_i^N$ and $0 = \bar{e}_{-i}^N$. For both cases, it is shown that $\bar{E}^S < \sum_i \bar{e}_i^N$ holds.

Case 1. We start by considering the case where $0 < \bar{e}_1^N, \bar{e}_2^N$. In this case, both FOCs in (3.16) hold with equality for the Nash equilibrium caps. Summing up these equations, we obtain from (3.16) that

$$\sum_i B_i'(A(\bar{e}_i^N, \bar{e}_{-i}^N)) = \sum_i C_i'(a_i(\bar{e}_i^N)) \tag{3.A.3}$$

must hold. Assume now, for the sake of contradiction, that $\bar{E}^S \geq \sum_i \bar{e}_i^N$ would hold. This implies that

$$\sum_i B_i'(A(\bar{e}_i^N, \bar{e}_{-i}^N)) \leq \sum_i B_i'(A(\bar{E}^S)) = C_i'(a_i(\bar{E}^S)) < \sum_i C_i'(a_i(\bar{e}_i^N)), \tag{3.A.4}$$

where the last inequality follows from the fact that $a_i(\bar{e}_i^N) \geq a_i(\bar{E}^S)$ as well as $a_{-i}(\bar{e}_{-i}^N) > 0$ must hold for at least one $i \in \{1, 2\}$ if $\bar{E}^S \geq \sum_i \bar{e}_i^N$. Comparing (3.A.3) and (3.A.4) yields a contradiction. We must thus have that $\bar{E}^S < \sum_i \bar{e}_i^N$.

Case 2. Now, consider the case where $0 < \bar{e}_i^N < e_{i,0}$ and $0 = \bar{e}_{-i}^N$. In this case, country i's FOCs in (3.16) holds with equality for the Nash equilibrium caps

$$B_i'(A(\bar{e}_i^N, 0)) = C_i'(a_1(\bar{e}_i^N)). \tag{3.A.5}$$

Assume, for the sake of contradiction, that $\bar{E}^S \geq \sum_i \bar{e}_i^N = \bar{e}_i^N$ would hold. This implies that

$$B_i'(A(\bar{e}_i^N, 0)) \leq B_i'(A(\bar{E}^S)) < \sum_i B_i'(A(\bar{E}^S)) = C_i'(a_i(\bar{E}^S)) \leq C_i'(a_i(\bar{e}_i^N)), \tag{3.A.6}$$

where the last inequality follows from the fact that $a_i(\bar{e}_i^N) \leq a_i(\bar{E}^S)$ must hold if $\bar{E}^S \geq \bar{e}_i^N$. Comparing (3.A.5) and (3.A.6) yields a contradiction. We must thus have that $\bar{E}^S < \sum_i \bar{e}_i^N$.

To see that $\bar{E}^S$ indeed induces a higher level of overall welfare, simply note that

$$W^S = W(\bar{E}^S) > W(\textstyle\sum_i \bar{e}_i^N) \geq W^N. \tag{3.A.7}$$

The first inequality follows from the fact that $\bar{E}^S < \sum_i \bar{e}_i^N$ where $\bar{E}^S$ is the unique maximizer of W. The second inequality holds since the cap and trade system realizes the benefits from capping the overall emission to the level $\sum_i \bar{e}_i^N$ at the lowest possible cost, i.e., a distribution of abatement activities among the countries that equates their marginal costs. □

Proof of Lemma 10. First, we show that Problem (3.17) is strictly concave. From differentiating the l.h.s. of (3.18), we get

$$\frac{\mathrm{d}^2 W_i(\bar{E},\hat{\mu}_1)}{\mathrm{d}\bar{E}^2} = -B_i''(A(\bar{E}))\frac{\partial A(\bar{E})}{\partial \bar{E}} + C_i''(a_i(\bar{E}))\frac{\partial a_i(\bar{E})}{\partial \bar{E}}\Big(2\mu_i + \frac{\partial a_i(\bar{E})}{\partial \bar{E}}\Big)$$
$$\text{where} \quad \mu_1 = \hat{\mu}_1, \quad \text{and} \quad \mu_2 = 1-\hat{\mu}_1,$$

since $C_i''' = \partial^2 a_i(\bar{E})/\partial \bar{E}^2 = 0$. Using (3.7), (3.8), and the fact that any $\mu_i \in [0,1]$, we obtain

$$\begin{aligned}\frac{\mathrm{d}^2 W_i(\bar{E},\hat{\mu}_1)}{\mathrm{d}\bar{E}^2} &= B_i''(A(\bar{E})) + C_i''(a_i(\bar{E}))\Big(-2\mu_i\zeta_{-i} + \zeta_{-i}^2\Big)\\ &\le B_i''(A(\bar{E})) + C_i''(a_i(\bar{E}))\zeta_{-i}^2\\ &= B_i''(A(\bar{E})) + \zeta_i\zeta_{-i} < 0.\end{aligned} \tag{3.A.8}$$

Next, we investigate whether the FOC has a solution by considering the limits

$$\lim_{\bar{E}\to E_0} -B_i'(A(\bar{E})) + \mu_i C_i'(a_i(\bar{E})) + x_{i,-i}(\bar{E})\;\; C_i''(a_i(\bar{E}))\frac{\partial a_i(\bar{E})}{\partial \bar{E}} = -\infty < 0,$$
$$\lim_{\bar{E}\to 0} -B_i'(A(\bar{E})) + \mu_i C_i'(a_i(\bar{E})) + x_{i,-i}(\bar{E})\;\; C_i''(a_i(\bar{E}))\frac{\partial a_i(\bar{E})}{\partial \bar{E}}$$
$$= \zeta_1\zeta_2 E_0\Big(\mu_i - \zeta_{-i} + \frac{e_{i,0}}{E_0}\Big) \gtreqless 0.$$

Now, we need to distinguish two cases:

Case 1. If $\mu_i - \zeta_{-i} + \frac{e_{i,0}}{E_0} > 0$, then the Intermediate Value Theorem implies the existence of solution $0 < \bar{E}_i < E_0$ to the FOC (3.18). Since Problem (3.17) is strictly concave, this solution must be unique.

Case 2. If $\mu_i - \zeta_{-i} + \frac{e_{i,0}}{E_0} \le 0$, then strict concavity of Problem (3.17) immediately implies that $\bar{E}_i = 0$. Using that $\zeta_{-i} = 1-\zeta_i$, $e_{i,0} = E_0 - e_{-i,0}$ and $\mu_{-i} = 1-\mu_i$, it holds for country $-i$ that

$$\begin{aligned}\mu_i - \zeta_{-i} + \frac{e_{i,0}}{E_0} &\le 0\\ \mu_i - (1-\zeta_i) - \frac{e_{-i,0}}{E_0} &\le -1\\ \mu_{-i} - \zeta_i + \frac{e_{-i,0}}{E_0} &\ge 1 > 0.\end{aligned}$$

Hence, if $\bar{E}_i = 0$ for country i, then the FOC (3.18) implies the existence of a unique solution $0 < \bar{E}_{-i} < E_0$ for country $-i$.

To compare $\bar{E}_i$ to $\bar{E}^S$, we need to consider the following three cases.

Case 1. If $0 < \bar{E}_i < \bar{E}_{-i}$, then, by strict concavity of W_i, we obtain the following limits

$$\begin{aligned} \lim_{\bar{E}\to\bar{E}_i} \frac{dW(\bar{E})}{d\bar{E}} &= \frac{dW_{-i}(\bar{E}_i,\hat{\mu}_1)}{d\bar{E}} > 0, \\ \lim_{\bar{E}\to\bar{E}_{-i}} \frac{dW(\bar{E})}{d\bar{E}} &= \frac{dW_i(\bar{E}_{-i},\hat{\mu}_1)}{d\bar{E}} < 0. \end{aligned} \tag{3.A.9}$$

Hence, the INTERMEDIATE VALUE THEOREM implies that $\bar{E}_i < \bar{E}^S < \bar{E}_{-i}$.

Case 2. If $0 = \bar{E}_i < \bar{E}_{-i}$, then Lemma 8 immediately implies that $\bar{E}_i < \bar{E}^S$, while the limit $\bar{E} \to \bar{E}_{-i}$ in (3.A.9) together with the concavity of W implies that $\bar{E}^S < \bar{E}_{-i}$. Thus, we have that $\bar{E}_i < \bar{E}^S < \bar{E}_{-i}$ again.

Case 3. In the trivial case $\bar{E}_i = \bar{E}_{-i}$, it is obvious that $\bar{E}_i = \bar{E}^S = \bar{E}_{-i}$ holds.

Since $i \in \{1,2\}$, combining *Case 1 – 3* yields

$$\min\{\bar{E}_1, \bar{E}_2\} \le \bar{E}^S \le \max\{\bar{E}_1, \bar{E}_2\},$$

which is the relation stated in Lemma 10. □

Proof of Lemma 11. Again, we need to distinguish three cases.

Case 1. Consider the case where $0 < \bar{E}_i < \bar{E}_{-i}$. Evaluating the derivative of the Nash product at the lower bound of $\mathscr{P}_{\hat{\mu}_1}$, we obtain, by definition of $\bar{E}_i$ and strict concavity of W_{-i}, that

$$\lim_{\bar{E}\to\bar{E}_i} \frac{d\mathscr{N}_{\hat{\mu}_1}(\bar{E})}{d\bar{E}} = \frac{dW_{-i}(\bar{E}_i,\hat{\mu}_1)}{d\bar{E}} \cdot W_i(\bar{E}_i,\hat{\mu}_1) > 0.$$

By contrast, evaluating the derivative of the Nash product at the upper bound of $\mathscr{P}_{\hat{\mu}_1}$, we obtain, by definition of $\bar{E}_{-i}$ and strict concavity of W_i, that

$$\lim_{\bar{E}\to\bar{E}_{-i}} \frac{d\mathscr{N}_{\hat{\mu}_1}(\bar{E})}{d\bar{E}} = \frac{dW_i(\bar{E}_{-i},\hat{\mu}_1)}{d\bar{E}} \cdot W_{-i}(\bar{E}_{-i},\hat{\mu}_1) < 0.$$

The INTERMEDIATE VALUE THEOREM implies that a solution $\bar{E}^B > 0$ to (3.20) in $\mathscr{P}_{\hat{\mu}_1}$ exists. To see that $\bar{E}^B$ is unique and indeed maximizes the Nash product, note first that $W_i > 0$, $dW_{-i}/d\bar{E} > 0$ and $dW_i/d\bar{E} < 0$ holds for all $\bar{E} \in (\bar{E}_i, \bar{E}_{-i})$. This requires that $W_{-i} > 0$ must hold for any solution $\bar{E}^B$ to (3.20). Hence, for the derivative of the l.h.s. of (3.20) evaluated at $\bar{E}^B$ it holds

$$\frac{d^2 \mathcal{N}_{\hat{\mu}_1}(\bar{E}^B)}{d\bar{E}^2} = \sum_i \left(\frac{\mathrm{d}^2 W_i(\bar{E}^B, \hat{\mu}_1))}{\mathrm{d}\bar{E}^2} \cdot W_{-i}(\bar{E}^B, \hat{\mu}_1)) + \frac{\mathrm{d}W_i(\bar{E}^B, \hat{\mu}_1)}{\mathrm{d}\bar{E}} \cdot \frac{\mathrm{d}W_{-i}(\bar{E}^B, \hat{\mu}_1)}{\mathrm{d}\bar{E}} \right)$$
$$< 0.$$

Graphically, for any solution $\bar{E}^B \in \mathcal{P}_{\hat{\mu}_1}$ to (3.20), the l.h.s. of (3.20) intersects with the r.h.s. (zero) with a negative slope such that only one solution exists, i.e., the solution to (3.20) is unique. Since the second derivative of the Nash product evaluated at $\bar{E}^B$ is negative, $\bar{E}^B$ is indeed a maximizer.

Case 2. Now, consider the case where $0 = \bar{E}_i < \bar{E}_{-i}$. The limit $\bar{E} \to \bar{E}_i$ then admits all signs, i.e., it might be positive or non-positive. In the positive case, $\bar{E}^B > 0$ is determined by the unique solution to (3.20) in $\mathcal{P}_{\hat{\mu}_1}$ following the arguments presented in case 1. In the non-positive case, no solution to (3.20) exists, and $\bar{E}^B = 0$ holds since the Nash product is decreasing on $\mathcal{P}_{\hat{\mu}_1}$.

Case 3. In the trivial case $\bar{E}_i = \bar{E}_{-i}$, it is obvious that $\bar{E}_i = \bar{E}^B = \bar{E}_{-i} > 0$ is the unique solution to (3.20) in $\mathcal{P}_{\hat{\mu}_1}$.

Combining *Case 1 – 3* and recognizing that $\bar{E}^B \leq \max\{\bar{E}_1, \bar{E}_2\} < E_0$ by Lemma 10 yields Lemma 11. □

Proof of Proposition 9. The proof proceeds in three steps. First, we show that

$$\bar{E}^S = \bar{E}^B \qquad \Longleftrightarrow \qquad \hat{\mu}_1 = \mu_1^S, \tag{3.A.10}$$

denoting μ_1^S the unique allocation that solves $W_i(\bar{E}^S, \hat{\mu}_1) = W_{-i}(\bar{E}^S, \hat{\mu}_1)$, i.e., that equally distributes the welfare level obtained under the efficient cap in both countries. We then establish that $\mu_1^S = 1/2$ in the symmetric case. Continuity thus implies that an allocation $\mu_1^S \in [0, 1]$ also exists if the countries are sufficiently symmetric, i.e., sufficiently similar in terms of benefits, costs, and initial emissions. Third, it is shown that $\hat{\mu}_1 \neq \mu_1^S$ if $B_i(A(\bar{E}^S)) - B_{-i}(A(\bar{E}^S))$ is sufficiently large.

Step 1. We start by establishing that

$$\bar{E}^S = \bar{E}^B \qquad \Longleftrightarrow \qquad W_i(\bar{E}^S, \hat{\mu}_1) = W_{-i}(\bar{E}^S, \hat{\mu}_1), \tag{3.A.11}$$

Note that $\bar{E}^S$ and $\bar{E}^B$ are unique (cf. Lemma 8 and 11). According to Lemma 8, it holds for the efficient cap that,

$$\sum_i \frac{dW_i(\bar{E}^S, \hat{\mu}_1)}{d\bar{E}} = 0. \tag{3.A.12}$$

Since $\bar{E}^S > 0$ by Lemma 8, we can infer from Lemma 11 that bargaining *can* result in the efficient cap, if and only if the bargaining outcome is determined by the FOC of the Nash product. In this case, $\bar{E}^B$ satisfies

$$\frac{d\mathcal{N}_{\hat{\mu}_1}(\bar{E})}{d\bar{E}} = \sum_i \frac{\mathrm{d}W_i(\bar{E}^B, \hat{\mu}_1)}{\mathrm{d}\bar{E}} \cdot W_{-i}(\bar{E}^B, \hat{\mu}_1) = 0. \tag{3.A.13}$$

If $\bar{E}^S = \bar{E}^B$, then Equation 3.A.13 implies that

$$\sum_i \frac{\mathrm{d}W_i(\bar{E}^S, \hat{\mu}_1)}{\mathrm{d}\bar{E}} \cdot W_{-i}(\bar{E}^S, \hat{\mu}_1) = 0$$

must also hold. Rearranging and inserting (3.A.12) yields that

$$\frac{\mathrm{d}W_i(\bar{E}^S, \hat{\mu}_1)}{\mathrm{d}\bar{E}}\Big(W_{-i}(\bar{E}^S, \hat{\mu}_1) - W_i(\bar{E}^S, \hat{\mu}_1)\Big) = 0, \qquad i = 1, 2,$$

which implies $W_i(\bar{E}^S, \hat{\mu}_1) = W_{-i}(\bar{E}^S, \hat{\mu}_1)$.

Now, consider the opposite direction. Multiplying both sides of (3.A.12), we obtain

$$W_i(\bar{E}^S, \hat{\mu}_1)\Big(\frac{dW_i(\bar{E}^S, \hat{\mu}_1)}{d\bar{E}} + \frac{dW_{-i}(\bar{E}^S, \hat{\mu}_1)}{d\bar{E}}\Big) = 0, \qquad i = 1, 2, \tag{3.A.14}$$

If $W_i(\bar{E}^S, \hat{\mu}_1) = W_{-i}(\bar{E}^S, \hat{\mu}_1)$, then (3.A.14) can be rewritten to

$$\sum_i \frac{\mathrm{d}W_i(\bar{E}^S, \hat{\mu}_1)}{\mathrm{d}\bar{E}} \cdot W_{-i}(\bar{E}^S, \hat{\mu}_1) = 0 \tag{3.A.15}$$

Comparing (3.A.13) to (3.A.15) immediately yields that $\bar{E}^S = \bar{E}^B$. We can therefore conclude that the relation stated in (3.A.11) holds.

To establish the following relation

$$W_i(\bar{E}^S, \hat{\mu}_1) = W_{-i}(\bar{E}^S, \hat{\mu}_1) \quad \Longleftrightarrow \quad \hat{\mu}_1 = \mu_1^S, \tag{3.A.16}$$

we need to show that a unique solution μ_1^S to $W_i(\bar{E}^S, \hat{\mu}_1) = W_{-i}(\bar{E}^S, \hat{\mu}_1)$ exists. To do so, note that the following limits obtain

$$\lim_{\hat{\mu}_1 \to \infty} W_1(\bar{E}^S, \hat{\mu}_1) - W_2(\bar{E}^S, \hat{\mu}_1) = +\infty,$$
$$\lim_{\hat{\mu}_1 \to -\infty} W_1(\bar{E}^S, \hat{\mu}_1) - W_2(\bar{E}^S, \hat{\mu}_1) = -\infty,$$

since $\bar{E}^S$ is fixed. Hence, the Intermediate Value Theorem implies the existence of a solution μ_1^S. From calculating the derivatives,

$$\frac{\mathrm{d}W_1(\bar{E}^S,\hat{\mu}_1)}{\mathrm{d}\mu_1} = p(\bar{E}^S)\bar{E}^S > 0, \quad \frac{\mathrm{d}W_2(\bar{E}^S,\hat{\mu}_1)}{\mathrm{d}\mu_1} = -p(\bar{E}^S)\bar{E}^S < 0, \tag{3.A.17}$$

we can conclude that the allocation μ_1^S is indeed unique. Combining (3.A.11) and (3.A.16) directly implies the relation in (3.A.10).

Step 2. From (3.9), we obtain that $W_i(\bar{E}^S,\hat{\mu}_1) = W_{-i}(\bar{E}^S,\hat{\mu}_1)$ if and only if $\hat{\mu}_1$ solves

$$\begin{aligned} B_i(A(\bar{E}^S)) - B_{-i}(A(\bar{E}^S)) + C_{-i}(a_{-2}(\bar{E}^S)) - C_i(a_i(\bar{E}^S)) &= -2x_{i,-i}C_i'(a_i(\bar{E}^S)) \\ \text{where} \quad x_{i,-i}(\bar{E}) := \left(\mu_i\bar{E}^S - e_{i,0} + a_i(\bar{E}^S)\right), \quad \mu_1 = \hat{\mu}_1, \quad &\text{and} \quad \mu_2 = 1-\hat{\mu}_1. \end{aligned} \tag{3.A.18}$$

In the symmetric case, we have have that $B_i(A) = B(A)$, $C_i(a_i) = C(a_i)$ and $e_{i,0} = e_{-i,0}$ for all $i \in \{1,2\}$. This implies that $a_i(\bar{E}^S) = a_{-i}(\bar{E}^S)$. Hence, the l.h.s. of (3.A.18) is zero. Inserting $x_{i,-i}$, using that $\bar{E}^S = 2(e_{i,0} - a_i(\bar{E}^S))$ and solving for $\hat{\mu}_1$ yields $\mu_1^S = 1/2$. By continuity, we can now infer that if the countries are sufficiently similar in terms of benefits, costs, and initial emissions, then $\mu_1^S \in [0,1]$.

Step 3. If $B_i(A(\bar{E}^S)) - B_{-i}(A(\bar{E}^S)) \to \infty$, then the l.h.s. of (3.A.18) converges to infinity since $C_i(a_i(\bar{E}^S)) \leq C_i(a_i(E_0))$ and $C_{-i}(a_{-i}(\bar{E}^S)) \leq C_{-i}(a_{-i}(E_0))$ are finite. Since all expressions on the r.h.s. of (3.A.18) besides μ_i are finite, $C_i'(a_i(\bar{E}^S)) \leq C_i'(a_i(E_0))$, $\bar{E}^S \leq E_0$, $a_i(\bar{E}^S) \leq a_i(E_0)$, $e_{i,0} \in \mathbb{R}_{++}$, we must have that $\mu_i \to -\infty$. This implies that

$$\mu_1^S \to \begin{cases} -\infty & \text{if } i = 1 \\ +\infty & \text{if } i = 2. \end{cases}$$

Hence, if $B_i(A(\bar{E}^S)) - B_{-i}(A(\bar{E}^S))$ sufficiently large, it holds that $\mu_1^S \notin [0,1]$ and thus that $\mu_1^S \neq \hat{\mu}_1 \in [0,1]$. By (3.A.10), $\mu_1^S \neq \hat{\mu}_1$ directly implies $\bar{E}^S \neq \bar{E}^B$. □

Proof of Corollary 4. Corollary 4 follows from the fact that $\bar{E}^B = \bar{E}^S$ if and only if $\hat{\mu}_1 = \mu_1^S$ by (3.A.10) and that μ_1^S is, in turn, determined by the unique solution to $W_i(\bar{E}^S,\hat{\mu}_1) = W_{-i}(\bar{E}^S,\hat{\mu}_1)$. Since W_i and W_{-i} also depend on benefit- and cost functions, this allocation generally deviates from an allocation that is solely determined on the basis of initial emissions. □

Proof of Proposition 10. We first establish that

$$\bar{E}^B = \bar{E}^S \quad \Longleftrightarrow \quad \mu_1^B = \mu_1^S \quad \Longleftrightarrow \quad \mu_1^S \in [0,1]. \tag{3.A.19}$$

To investigate whether a solution to Problem (3.21) implements $\bar{E}^B = \bar{E}^S$, we solve the relaxed problem

$$\max_{(\bar{E},\mu_1)} \mathscr{N}(\bar{E},\mu_1) \qquad \text{s.t. } 0 \le \mu_1 \le 1, \tag{3.A.20}$$

and show that the solution to the relaxed problem is in $\mathscr{P}$, i.e., corresponds to a solution to the original Problem (3.21). The Lagrangian of Problem (3.A.20) writes

$$\mathscr{L}(\bar{E},\mu_1) = W_1(\bar{E},\mu_1) \cdot W_2(\bar{E},\mu_1) - \lambda_1(\mu_1 - 1) + \lambda_2\mu_1.$$

The solution to the relaxed problem is determined by the system of FOCs

$$\frac{\mathrm{d}\mathscr{L}(\bar{E},\mu_1)}{\mathrm{d}\bar{E}} = \frac{\mathrm{d}W_1(\bar{E},\mu_1)}{\mathrm{d}\bar{E}} \cdot W_2(\bar{E},\mu_1) + W_1(\bar{E},\mu_1) \cdot \frac{\mathrm{d}W_2(\bar{E},\mu_1)}{\mathrm{d}\bar{E}} = 0, \tag{3.A.21}$$

$$\frac{\mathrm{d}\mathscr{L}(\bar{E},\mu_1)}{\mathrm{d}\mu_1} = \Big(W_2(\bar{E},\mu_1) - W_1(\bar{E},\mu_1)\Big)p(\bar{E})\bar{E} - \lambda_1 + \lambda_2 = 0, \tag{3.A.22}$$

and the complementary slackness conditions

$$\lambda_1(\mu_1 - 1) = 0, \qquad \text{and} \qquad \lambda_2\mu_1 = 0. \tag{3.A.23}$$

Now, three cases can occur.[17]

Case 1. If $\mu_1^S \in [0,1]$, then the tuple $(\bar{E}^S, \mu_1^S)$ is the unique solution to (3.A.21) and (3.A.22) by implementing

$$W_1(\bar{E}^S,\mu_1^S) = W_2(\bar{E}^S,\mu_1^S), \qquad \text{and} \qquad \frac{\mathrm{d}W_1(\bar{E}^S,\mu_1^S)}{\mathrm{d}\bar{E}} = \frac{\mathrm{d}W_2(\bar{E}^S,\mu_1^S)}{\mathrm{d}\bar{E}}.$$

To see that $(\bar{E}^S, \mu_1^S) \in \mathscr{P}$, note that $\bar{E}^S$ leads to the greatest level of overall welfare. Hence, for any other cap $\bar{E} \neq \bar{E}^S$, the total level of welfare is lower such that at least one country is worse off, i.e., setting $\bar{E} \neq \bar{E}^S$ does not constitute a Pareto improvement. For $\bar{E} = \bar{E}^S$, on the other hand, any other allocation $\mu_1 \neq \mu_1^S$ leaves one country worse off (cf. (3.A.17)), i.e., does not constitute a Pareto improvement.

17. Note that in the case in which no strictly positive solution to (3.A.21) exists, we have that $0 = \bar{E}^B \neq \bar{E}^S$ by Lemmas 8 and 11.

Thus, we must have that $(\bar{E}^S, \mu_1^S) \in \mathscr{P}$ such that $(\bar{E}^S, \mu_1^S)$ is also the solution to the original Problem (3.21) in this case.

$(\bar{E}^S, \mu_1^S)$ is indeed the global maximizer of the Nash product, which can be seen as follows. It is well known that the product of two positive real numbers is maximum when the numbers are equal, given that their sum is constant. In our setting, for any given $\bar{E}$, the corresponding level of total welfare is constant, i.e., independent of the actual allocation μ_1. Hence, given $\bar{E}$, the Nash product is maximum for the allocation that equates welfare levels in both countries. Moreover, $\bar{E}^S$ implements the highest positive level of total welfare. The tuple $(\bar{E}^S, \mu_1^S)$ must thus be a global maximizer since it equally distributes the highest level of total welfare.

Case 2. If $\mu_1^S > 1$, then $\bar{E}^S$ does not solve (3.A.21) since it holds for all $\mu_1 \in [0, 1]$ that

$$\frac{\mathrm{d}W_1(\bar{E}^S, \mu_1)}{\mathrm{d}\bar{E}} = \frac{\mathrm{d}W_2(\bar{E}^S, \mu_1)}{\mathrm{d}\bar{E}}, \quad \text{and} \quad W_1(\bar{E}^S, \mu_1) \neq W_2(\bar{E}^S, \mu_1).$$

Instead, we can read off (3.A.21) that the solution to the relaxed problem $\bar{E}^R$ must be in the open interval between $\bar{E}_1$ and $\bar{E}_S$. Moreover, (3.A.21) together with concavity of W_i implies that

$$W_2(\bar{E}^R, \mu_1^R) > W_1(\bar{E}^R, \mu_1^R) \tag{3.A.24}$$

holds for the solution to the relaxed problem. Hence, from (3.A.22) we obtain that $\lambda_1 > 0$. By (3.A.23), we can now infer that $\mu_1^R = 1$ and thus $\lambda_2 = 0$.

To see that the tuple $(\bar{E}^R, 1) \in \mathscr{P}$, first observe that for $\mu_1^R = 1$ any other cap from the open interval between $\bar{E}^R$ and $\bar{E}_2$ leaves country 1 strictly worse off. This is illustrated in Figure 3.A.1. Moreover, given any cap from the open interval between $\bar{E}^R$ and $\bar{E}_2$, any other allocation $\mu_1 < 1$ further reduces country 1's welfare. Hence, choosing any cap from the open interval between $\bar{E}^R$ and $\bar{E}_2$ and any $\mu_1 \in [0, 1]$ does not constitute a Pareto improvement. Second, since $\bar{E}^R$ is in the open interval between $\bar{E}_S$ and $\bar{E}_1$, any cap in the open interval between $\bar{E}_1$ and $\bar{E}^R$ leads to a lower level of overall welfare irrespective of the allocation μ_1, see Figure 3.A.1. Hence, for any allocation $\mu_1 \in [0, 1]$, any cap from this interval leaves at least one country worse off, i.e., does not constitute a Pareto improvement. The tuple $(\bar{E}^R, 1)$ is thus Pareto-efficient, that is $(\bar{E}^R, 1) \in \mathscr{P}$. The solution to the relaxed Problem (3.A.20) therefore corresponds to the solution to the original Problem (3.21) such that $\bar{E}^B = \bar{E}^R \neq \bar{E}^S$ and $\mu_1^B = \mu_1^R = 1 < \mu_1^S$.[18]

18. It can readily be verified that $(\bar{E}^R, 1)$ is indeed the maximizer of the relaxed problem.

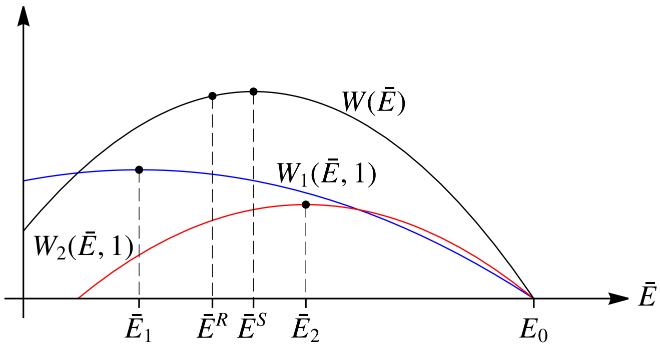


Figure 3.A.1. Pareto efficiency of the solution to the relaxed problem.

Case 3. If $\mu_1^S < 0$, then $\bar{E}^B = \bar{E}^R \neq \bar{E}^S$ and $\mu_1^B = \mu_1^R = 0 > \mu_1^S$ follows by analogous arguments.

Combining these three cases yields the relation stated in (3.A.19).

Part (i) of Proposition 10 now immediately follows from (3.A.19) and the fact that $\mu_1^S \in [0,1]$ if the countries are sufficiently symmetric, which we have shown in the proof of Proposition 9.

Part (ii) of Proposition 10 results from (3.A.19) and the fact that $\mu_1^S \notin [0,1]$ if the countries' benefits are sufficiently different, which we have already argued in the proof of Proposition 9. □

Proof of Proposition 11. We know from Proposition 10 that, in the absence of outside options, bargaining implements $(\bar{E}^S, \mu_1^S)$ where $\mu_1^S \in [0,1]$ if the countries are sufficiently symmetric. Hence, it remains to show that the tuple $(\bar{E}^S, \mu_1^S)$ also satisfies the additional constraints $W_1(\bar{E}, \mu_1) \geq W_1^N$ and $W_2(\bar{E}, \mu_1) \geq W_2^N$ for sufficiently symmetric countries. To do so, we show for the (completely) symmetric case where $B_i(A) = B(A)$, $C_i(a_i) = C(a_i)$ and $e_{i,0} = e_{-i,0}$ that $W_i(\bar{E}^S, \mu_1^S) > W_i^N$ holds for all $i \in \{1, 2\}$. Continuity thus implies that, as long as the countries are sufficiently symmetric, $(\bar{E}^S, \mu_1^S)$ satisfies the additional constraints, i.e., $(\bar{E}^S, \mu_1^S)$ is indeed the solution to Problem (3.22).

In the symmetric case without outside options, the bargaining outcome is $(\bar{E}^S, 1/2)$, which leads to the following welfare levels:

$$W_i(\bar{E}^S, 1/2) \;=\; W_{-i}(\bar{E}^S, 1/2) \;=\; \frac{W^S}{2}. \tag{3.A.25}$$

For national caps, on the other hand, symmetry leads to $\bar{e}_i^N = \bar{e}_{-i}^N$ and thus

$$W_i^N \;=\; W_{-i}^N \;=\; \frac{W^N}{2}. \tag{3.A.26}$$

In view of Proposition 8, combining (3.A.25) and (3.A.26) immediately implies that $W_i(\bar{E}^S, \mu_1^S) > W_i^N$ holds for all $i \in \{1, 2\}$ in the symmetric case. □

Proof of Proposition 12. Observing that bargaining over the amount and allocation of permits implements $\bar{E}^S$ if and only if $\mu_1^S \in [0, 1]$ by (3.A.19), the proposition follows immediately from comparing Examples 10 and 11. □

3.A.2 Preliminary Version from Naumann's (2022) Dissertation

Emissions Trading Schemes: Negotiations on the Emissions Cap

Fabian Naumann* Tom Rauber †

Abstract: Setting a sufficiently stringent emissions cap is a key factor in ensuring that an emissions trading system can effectively tackle climate change. The crucial question therefore becomes: what cap is implemented? In this paper, we consider an alternating-offers model in which two asymmetric countries have already committed to jointly implement an emissions trading scheme. We investigate whether bargaining over the emissions cap can result in the social emissions optimum and the reasons for deviations. We show that an initial endowment of emission rights based on historic emissions never results in the social optimum. However, other permit allocations exist which lead to the social optimum. In this case, the initial endowment can, to some extent, function relatively similar to a side payment, allowing efficiency and distribution to be separated. If the negotiating countries are too different, no allocation of allowances can lead to the socially optimal emissions level.

Keywords: Nash bargaining solution, Emissions trading schemes, Emissions cap

JEL Classification: C71; D62; H23; Q53

*Technische Universität Kaiserslautern (TUK), fabian.naumann@wiwi.uni-kl.de
†Technische Universität Kaiserslautern (TUK), tom.rauber@wiwi.uni-kl.de

1 Introduction

The ongoing climate change and its immense impact, which poses tremendous challenges to humanity, highlights market failure in the provision of global public goods.

Market failure in the presence of externalities is a well-studied problem in the literature. Coase (1960) shows that, once property rights are clearly defined, bargaining can lead to an efficient market outcome. This holds even in the presence of externalities and regardless of the initial allocation of property rights. First economic approaches to tackle the decline in environmental quality go back to the 1960s when Dales (1968) proposes a charging scheme in his seminal work. The suggested scheme limits the number of rights to pollute, issued by the government, thereby restricting environmental damages, such as to water and the atmosphere. Together with the theoretical foundation of markets in licenses and its cost efficiency by Montgomery (1972), this lays the theoretical basis for emissions trading systems as they are implemented nowadays. In light of the modern research that was built on that foundation, our paper can be viewed in the larger context of three strands of literature, namely *allowance choices* as well as *public good provision from a political perspective* and *linking emissions trading schemes*.

In a more recent paper, Helm (2003) shows that endogenous *allowance choices* by countries do not automatically result in lower pollution levels, as environmentally more (less) concerned countries choose to pollute less (more) and thus environmental efforts offset. It becomes clear that a transnational problem requires cooperation between countries. Smead et al. (2014) analyze a game, where agents bargain over their share of the fixed emission total including learning dynamics. They find that negotiations tend to fail if too many agents are faced with an under-proportional emissions share, making the initial demand a key factor for a successful negotiation.

Segendorff (1998) is the first to consider delegates in the context of international environmental agreements and represents the *public good provision from a political perspective.*[1] He finds that authorities choose delegates who misrepresent their preferences.[2]. Loeper (2017) analyzes international cooperation, where policymakers are elected by a country's population. Strategic behavior by voters leads to the election of policymakers who under-represent interests. A key finding is that the type of public good is relevant and a more convex demand function enhances the provision of a public good. Arvaniti and Habla (2021) also contribute to the "strategic delegation" literature, showing that delegates who misrepresent preferences lead to a situation in which it is not clear if and for whom cooperation is beneficial. Although we also explore how countries cooperate, in

[1] See also Siqueira (2003).
[2] For further contributuins on misrepresentation of preferences see also Crawford and Varian (1979), Jones (1989), and Burtraw (1992, 1993).

our case to define the emission cap in a cap and trade system, we abstract from situations where delegates falsely represent their countries preferences.

If national emissions trading systems have already been implemented, the question arises as to whether *linking existing systems* is advantageous. It becomes more advantageous the higher the jurisdictions' size and shock variances, while a higher correlation of shocks and sunk costs of linking show the opposite effect (Doda and Taschini, 2017). Flachsland et al. (2009) analyze from an economic, political, and regulatory perspective the benefits and disadvantages of linking, such as reduced volatility, strengthening the multilateral commitment versus expanded emission caps to obtain permit trade benefits and abatement targets that are not in line with a burden-sharing approach and decline in a country's regulatory power. Doda et al. (2019) find that multilateral linking can lead to tremendous efficiency gains which arise equally from effort and risk-sharing. However, Habla and Winkler (2018) demonstrate that strategic delegation hinders the linking of emissions trading schemes.

The paper closest in spirit to ours is Dijkstra and Nentjes (2020), which compares the Exchange-Matching-Lindahl (EML) solution (a bottom-up mechanism) and the Nash Bargaining solution (a top-down mechanism) for the provision of a public good. As the EML is lesser-known, we only briefly summarize this cooperation mechanism, as used by Dijkstra and Nentjes (2020). Under EML there is an exchange rate offered to countries, which specifies the ratio of global to national abatement. Now, given this exchange rate, countries declare their respective supply and demand for emissions reduction. They find a Pareto-efficient equilibrium, in which countries' demands are identical due to different exchange rates. Their results indicate that i) in a setting with two agents both mechanisms are equivalent, ii) in a setting with more than two agents, EML is beneficial for agents with high benefit and low costs, and iii) lower side payments under the EML mechanism.

In contrast to the existing literature, we analyze a negotiation in which the total cap of the emissions trading scheme is the outcome of the bargaining while the division of the total cap among the countries' is fixed. Our study is based on a two-country model, where we assume that asymmetric countries have already committed to jointly introduce a cap and trade system.[3] Using a model with alternating offers, we investigate under which conditions this bargaining process leads to a socially optimal emissions quantity and why it can deviate from it.

Our results show that the bargaining process can bring about a situation in which the negotiated emissions level equals the social optimum. If allowances are allocated based on historical emissions, the socially optimal level of emissions can never be achieved in our model environment. However,

[3] In a larger context, this could also be seen as two countries agreeing to link their emissions trading system and negotiate the overall cap.

the outcome of the bargaining may result in the social optimum if the allocation differs. In this case, the allocation of permits might be used as a compensation mechanism between countries. Furthermore, we find that if countries are too different, the redistribution of allowances reaches its limit and no allocation can lead to the social optimum. Nevertheless, we demonstrate that bargaining can result in a better solution compared to national emission trading schemes with national emission caps. Our work, thus, identifies reasons why the socially optimal emissions cap is not implemented in an emissions trading scheme. Although we made some strong assumptions, the model helps understand the reasons why emissions trading schemes might not set tight emission caps. In addition, the model can be extended to represent more realistic scenarios, such as an outside option or a risk of a breakdown in the bargaining process.

The paper is organized as follows. In the following section, we introduce the simple two-country model and provide basic insights about abatements as well as the emissions market price. Section 3 defines two benchmark scenarios, namely the social optimal emissions cap and national emission caps, for the welfare analysis. In Section 4 we analyze the cap negotiation, using an alternating-offers model, and compare results with the defined benchmarks. Finally, Section 5 concludes.

2 Model and Basic Insights

We briefly introduce the underlying theoretical framework in this section, deployed to describe a cap and trade system. Later, we endogenize the cap in this model by allowing the countries to negotiate.

Welfare Function

We define the welfare of a country $i \in [1, 2]$ as benefits of overall abatement $B(\sum_i a_i)$, assuming a positive externality (country i benefits from the abatement made by country $-i$), minus costs of abatement in a country $C(a_i)$. Because countries are linked through a cap and trade system, emissions trading results in either additional revenue or costs, depending on whether a country is a buyer or seller of permits. A country is a buyer (seller) of permits if actual emissions, e_i, are higher (lower) than its initial endowment of permits, $\bar{e}_i$, where p is the endogenous permit market price. For the emissions cap of the scheme $\bar{E}$ it holds that $\bar{E} = \bar{e}_1 + \bar{e}_2 = \mu\bar{E} + (1-\mu)\bar{E}$, where μ defines the permit share allocated to country 1.[4] Putting the components together the welfare function reads:

[4]We take μ as given. It could, for instance, be determined by emissions under business as usual, by a certain historical tradition, or by previous negotiations.

$$W_i(a_i, a_{-i}) = B_i(a_i, a_{-i}) - C_i(a_i) + (\bar{e}_i - e_i) \cdot p, \quad for\ all\ i = 1,2 \tag{1}$$

As realized emissions e_i are emissions under “business as usual” $e_{i,0}$ minus actual realized abatement a_i:[5], we can re-write the welfare function as follows.

$$W_i(a_i, a_{-i}) = B_i(a_i, a_{-i}) - C_i(a_i) + (\bar{e}_i - (e_{i,0} - a_i)) \cdot p, \quad for\ all\ i = 1,2 \tag{2}$$

Benefit and Cost Function

In the literature, linear or quadratic functions are often assumed for abatement benefits and costs, see, for instance, Weitzman (1974, 2014), Barrett (1994) and McGinty (2007). We assume a quadratic benefit function (3) and cost function (4) for each country, where the total abatement A is defined as $A \coloneqq \sum_i a_i$.

$$B_i(A) = \beta_i A - \frac{\delta_i}{2} A^2, \quad where\ \beta_i > 0\ and\ \delta_i > 0, \qquad for\ all\ i = 1,2 \tag{3}$$

$$C_i(a_i) = \frac{\zeta_i}{2} a_i^2, \qquad where\ \zeta_i > 0, \qquad for\ all\ i = 1,2. \tag{4}$$

In our analysis, we want to focus on the case that provides the most insights, where the two countries have different characteristics and, therefore, different goals in the reduction of emissions. Let country 1 have *i*) small emissions under “business as usual”, *ii*) high benefits, but also *iii*) high costs from abatement. Country 1 can be thought of as a country with a relatively small GDP, but rather higher exposure to the negative consequences of the emissions, e.g. because of geographical factors like a long coast line. County 2, on the contrary, has *i*) high emissions under “business as usual”, *ii*) low benefits, but also *iii*) low costs from abatement. This is an adequate representation of a country with a high GDP that is less affected by the negative consequences of the emissions. To reflect this in our model, we assume that $\zeta_1 > \zeta_2$, $\beta_1 > \beta_2$ while $\delta \coloneqq \delta_1 = \delta_2$, and $e_{2,0} = \kappa \cdot e_{1,0}$ where $\kappa > 1$.

Furthermore, in the main part of the paper, we focus on the most realistic scenarios where the countries' optimal caps are greater than zero, i.e., where complete decarbonization is never optimal. Formally, this requires $W_i(a_i, a_{-i})$ to be a concave function with its maximum greater than zero. As we see below, this is ensured for any given allocation μ by the following technical assumptions that we impose about the relation between the model parameters. These assumptions

[5] We assume $a_i \geq 0$. A country cannot increase its emissions above the initial level, the emissions “business as usual”.

moreover determine how the allocation μ changes country i's optimal cap, which allows us to avoid tedious case distinctions and focus on the cases where compelling results are obtained.

i) $\zeta_2 \left(\frac{\zeta_1}{\zeta_1+\zeta_2}\right)^2 < \delta.$

ii) $\frac{3\zeta_1+\zeta_2}{\zeta_1+3\zeta_2} < \kappa,$[6]

iii) $\frac{\beta_1}{e_{1,0}} < \frac{1}{2}\left[\delta + \frac{\zeta_1\zeta_2(2\zeta_1+\zeta_2)}{(\zeta_1+\zeta_2)^2}\right] + \frac{\kappa}{2}\left[\delta - \zeta_1\left(\frac{\zeta_2}{\zeta_1+\zeta_2}\right)^2\right].$

Abatements

We start by deriving some basic insights concerning the realized emission price and the abatement activities within an existing cap and trade system. As firms under regulation minimize their abatement costs, the minimization problem of a representative price-taking firm in each country reads

$$\min_{a_i} p \cdot (e_{i,0} - a_i) + \frac{\zeta_i}{2} a_i^2. \tag{5}$$

The corresponding FOC for a representative firm in country i reads as

$$a_i = \frac{p}{\zeta_i}, \tag{6}$$

and determines the optimal abatement activities carried out in that country. Since maximum emissions in total are limited to the overall emissions cap $\bar{E}$, market clearing in the emission permits market requires that $\sum_i (e_{i,0} - a_i) = \bar{E}$, where $e_{2,0} = \kappa e_{1,0}$. Inserting the abatements per country leads the emissions market price p:

$$p = \frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}\left((\kappa+1)e_{1,0} - \bar{E}\right). \tag{7}$$

The resulting abatement activities are:

$$a_1 = \frac{\zeta_2}{\zeta_1+\zeta_2}\left[(\kappa+1)e_{1,0} - \bar{E}\right], \tag{8}$$

$$a_2 = \frac{\zeta_1}{\zeta_1+\zeta_2}\left[(\kappa+1)e_{1,0} - \bar{E}\right], \tag{9}$$

$$A = a_1 + a_2 = (\kappa+1)e_{1,0} - \bar{E}. \tag{10}$$

Since country 2 represents the country with lower abatement costs ($\zeta_1 > \zeta_2$), this country also contributes more to total abatement ($a_2 > a_1$). This is in accordance with emissions trading and

[6] Note that this is a rather weak assumption, for $\zeta_2 \approx \zeta_1$ we get that $1 < \kappa$, while we get for $\zeta_2 \ll \zeta_1$ that $3 < \kappa$.

rather intuitive. Because emission permits can be traded, an international emissions trading scheme leads to emissions abatement in the most cost-effective way by equalizing marginal abatement costs.

3 Benchmarks

We define two benchmark scenarios with the help of which we can then evaluate the bargaining results of our model. In the first scenario, a *centralized* cap definition carried out by a social planner is analyzed, before investigating the *decentralized* processes in subsequent sections. In particular, analyzing the benchmarks allows us to compare the social welfare generated through bargaining to the welfare generated in those scenarios.

3.1 Social Optimum

Let us first assume that there exists a social planner who maximizes the welfare of both countries involved, which is $W(\bar{E}) = W_1(\bar{E}) + W_2(\bar{E})$. It is apparent that, from a centralized perspective, trading activities between the countries offset each other. Hence, the overall welfare optimized by the social planner only consists of benefits and costs and reads as:

$$W(\bar{E}) = B_1(A(\bar{E})) - C_1(a_1(\bar{E})) + B_2(A(\bar{E})) - C_2(a_2(\bar{E})). \tag{11}$$

The socially optimal cap satisfies the FOC, which is

$$W'(\bar{E}) = \left[B_1'(A(\bar{E})) + B_2'(A(\bar{E}))\right] \frac{\partial A}{\partial \bar{E}} - C_1'(a_1(\bar{E}))\frac{\partial a_1}{\partial \bar{E}} - C_2'(a_2(\bar{E}))\frac{\partial a_2}{\partial \bar{E}} = 0. \tag{12}$$

Intuitively, the abatement activities induced by the welfare maximizing cap $\bar{E}_S^*$ balance the marginal cost of abatement with the overall marginal benefit of abatement. Using (8)–(10) the FOC can be rewritten as

$$C_i'(a_i(\bar{E})) = B_1'(A(\bar{E})) + B_2'(A(\bar{E})) \quad i \in [1, 2]. \tag{13}$$

Since both countries benefit from a marginal increase in the abatement irrespectively where the emissions have been saved, we obtain the sum of the marginal benefits on the right-hand side of Equation (13). Since emissions rights can be traded, it leads to a situation where marginal abatement costs of participating countries equalize, i.e., $C_1'(a_1(\bar{E})) = C_2'(a_2(\bar{E}))$, and ultimately determine the resulting price on the certificate market. In general, emission trading schemes thereby ensure that the emissions target is met at lowest cost, which makes it an efficient policy instrument to regulate pollution.[7] Hence, a marginal increase in the abatement induces costs at the level represented on the left-hand side of Equation (13). The welfare is maximized at the socially optimal emissions level $\bar{E}_S^*$ where total marginal benefit is equal to marginal cost. For the specified benefit and cost functions, cf. (3) and (4), we can explicitly solve for the socially optimal cap, which is

$$\bar{E}_S^* = (\kappa + 1)e_{1,0} - \frac{\beta_1 + \beta_2}{2\delta + \frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}}. \tag{14}$$

The second term of Equation (14) can be interpreted as the abatement target implemented via the cap and trade system, which is subtracted from total emissions under business as usual. In our notation, we define the socially optimal welfare generated by $\bar{E}_S^*$ as

$$W_S^* := W\left(\bar{E}_S^*\right). \tag{15}$$

3.2 National Caps

Now, we turn to a decentralized scenario, in which countries do not participate in a joint emissions trading scheme but instead deploy national regulations, in form of national cap and trade systems. Because each country implement its own emissions cap, the corresponding abatements for the countries are

$$a_1 = e_{1,0} - \bar{e}_1, \tag{16}$$

$$a_2 = \kappa e_{1,0} - \bar{e}_2. \tag{17}$$

Given the cap of country $-i$, country i chooses its own cap to maximize its welfare, i.e., as solution to

$$\max_{\bar{e}_i} B_i(A(\bar{e}_i, \bar{e}_{-i})) - C_i(a_i(\bar{e}_i)) \tag{18}$$

Hence, the solution to (18) defines a reaction function of the form $\bar{e}_i(\bar{e}_{-i})$ for each country. Solving this system of equations leads to the Nash equilibrium, where

$$\bar{e}_{1,C}^* = e_{1,0} - \frac{\zeta_2\beta_1 + \delta(\beta_1 - \beta_2)}{\zeta_1\zeta_2 + \delta(\zeta_1 + \zeta_2)}, \tag{19}$$

$$\bar{e}_{2,C}^* = \kappa e_{1,0} - \frac{\zeta_1\beta_2 - \delta(\beta_1 - \beta_2)}{\zeta_1\zeta_2 + \delta(\zeta_1 + \zeta_2)}. \tag{20}$$

[7] This holds true in absence of transaction costs and imperfect competition (e.g., Hahn, 1984; Stavins, 1995).

Intuitively, the cap $e^*_{i,C}$ chosen by county i is the optimal response to the cap $e^*_{-i,C}$ chosen by country $-i$ such that no country has an incentive to deviate. To capture the global effect that those caps have on the overall emissions, we define $\bar{E}^*_C$ as sum of the individual caps, representing the resulting overall emissions level in the two country setting. Summing up Equation (19) and (20) and rearranging leads to

$$\bar{E}^*_C = \bar{e}^*_{1,C} + \bar{e}^*_{1,C} = (\kappa + 1)e_{1,0} - \frac{\zeta_2\beta_1 + \zeta_1\beta_2}{\zeta_1\zeta_2 + \delta(\zeta_1 + \zeta_2)} \tag{21}$$

Again, the second summand of Equation (19) can be interpreted as total abatement that is implemented via the decentralized CAPs. Similarly to the socially optimal welfare, the overall welfare generated by national caps is defined as

$$W^*_C \coloneqq W_1(\bar{e}^*_{1,C}) + W_2(\bar{e}^*_{2,C}). \tag{22}$$

4 Cap Negotiations

Now, we turn to the case where the countries have already agreed to commit in a cap and trade system and explore how they endogenously set the cap via negotiating.

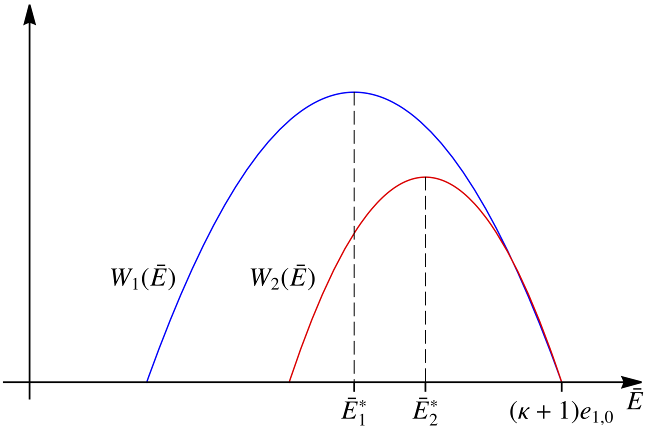
4.1 Basic Insights

To begin with, we specify the set of potential bargaining solutions and the effect of the model parameters. We, therefore, determine how each country i would optimally set the global cap $\bar{E}^*_i$. Since each country wants to maximize its welfare, the FOC of Equation (2) determines each country's desired cap. After simplifying the FOC writes as

$$B'_i(A(\bar{E})) = \mu_i C'_i(a_i(\bar{E})) + \left(\mu_i \bar{E} - e_{i,0} + a_i(\bar{E})\right) C''_i(a_i(\bar{E}))\frac{\partial a_i}{\partial \bar{E}}, \quad for\ all\ i = 1, 2, \tag{23}$$

where $\mu_1 = \mu$ and $\mu_2 = 1 - \mu$. The right-hand side of Equation (23) captures two marginal effects, a cost effect (first summand on the right-hand side) and a trading effect (second summand on the right-hand side). In optimum, these two effects equal marginal benefit (left-hand side of Eq. (23)). If a country is a seller (buyer) of certificates, the trading effect is positive (negative). Explicitly

Figure 1: Optimal caps for county 1 and 2.



solving Equation (23) for each country i leads to

$$\bar{E}_1^* = e_{1,0}(\kappa+1) - \frac{e_{1,0}\left[(\kappa+1)\mu - 1\right] + \frac{\zeta_1+\zeta_2}{\zeta_1\zeta_2}\beta_1}{2\mu + \frac{\zeta_1+\zeta_2}{\zeta_1\zeta_2}\left[\delta - \zeta_1\left(\frac{\zeta_2}{\zeta_1+\zeta_2}\right)^2\right]} \tag{24}$$

$$\bar{E}_2^* = e_{1,0}(\kappa+1) - \frac{e_{1,0}\left[1 - (\kappa+1)\mu\right] + \frac{\zeta_1+\zeta_2}{\zeta_1\zeta_2}\beta_2}{2(1-\mu) + \frac{\zeta_1+\zeta_2}{\zeta_1\zeta_2}\left[\delta - \zeta_2\left(\frac{\zeta_1}{\zeta_1+\zeta_2}\right)^2\right]} \tag{25}$$

We immediately find that $\bar{E}_1^* < \bar{E}_2^*$ for κ sufficiently large or for β_1 sufficiently greater than β_2. Intuitively, if country 1 has rather low emissions under "business as usual" or if its benefits from abatement are rather high compared to country 2, then country 1 advocates a lower cap, i.e., stronger abatement activities. Figure 1 illustrates each country's welfare dependent on the cap $\bar{E}$ in the case where $\bar{E}_1^* < \bar{E}_2^*$. As it can be seen, the set of Pareto efficient caps is given by the interval $[\bar{E}_1^*, \bar{E}_2^*]$. Since bargaining results in a Pareto efficient outcome, the bargaining solution is located in this closed interval. Moreover, if the countries agree to keep the current level of emission, i.e., a cap of $(\kappa+1)e_{1,0}$, both countries obtain a welfare of zero by the construction of the cost and benefit function. In general, the set of Pareto efficient caps, $\mathscr{P}$, from which we determine the Pareto efficient bargaining outcome, is given by the interval $[\min\{\bar{E}_1^*, \bar{E}_2^*\}, \max\{\bar{E}_1^*, \bar{E}_2^*\}]$.[8] The effect that the allocation of the certificates among the countries has on the optimal caps is specified in Lemma 1. For the proof see Appendix A.

Lemma 1 *If the share of country 1's certificates, μ, increases, then country 1's optimal cap decreases, $\frac{\partial \bar{E}_1^*}{\partial \mu} < 0$, while country 2's optimal cap increases, $\frac{\partial \bar{E}_1^*}{\partial \mu} > 0$.*

[8]Note that in the trivial case where $\min\{\bar{E}_1^*, \bar{E}_2^*\} = \max\{\bar{E}_1^*, \bar{E}_2^*\}$ no conflict of interests arises, and the countries agree upon a cap which is socially optimal.

Intuitively, since its emissions are significantly lower, country 1 wants to reduce the total amount of certificates to reduce the total emissions and to be able to sell its excess of certificates at a higher price if its share of certificates increases. By contrast, country 2 wants to compensate a lower share of certificates allocated to it by increasing the total amount of certificates and thereby reducing the market price.

4.2 Nash Bargaining Solution

One of the simplest ways to model a cap negotiation[9] is according to Rubinstein's (1982) alternating-offers model. Country i proposes a cap. Then country $-i$ can either accept this offer and the game ends or reject the offer and make a counter offer after $\Delta > 0$ time units. In case of rejection, it is i's turn to decide whether to accept the counteroffer or to make a counter-counteroffer. This process continues until one country accepts the proposed cap.[10] A prominent result in bargaining theory is that the subgame perfect equilibrium in the Rubinstein model converges to Nash's (1950) bargaining solution if the absolute magnitudes of the frictions in the bargaining process are small (Binmore et al., 1986; Binmore, 1987). Evidently, this is in accordance with our set-up as the bargaining process is substantially faster than climate change. Hence, even if the bargaining is extended by Δ due to the rejection of an offer almost the same benefits and costs can be reached through an agreement in the next round. For simplicity, we assume that the counties have the same discount rate such that we can apply the (symmetric) Nash bargaining solution.[11]

4.2.1 Definition

In our setting the Nash bargaining solution is defined as the solution of the following maximization problem:

$$\max_{\bar{E}\in\mathscr{P}} W_1\left(\bar{E}\right)\cdot W_2\left(\bar{E}\right), \tag{26}$$

where $W_1 \cdot W_2$ is referred to as Nash product. To facilitate exposition, we exploit the relation to the first-order condition specified in Lemma 2. For the proof see Appendix B.

Lemma 2 *The Nash bargaining solution $\bar{E}_N^*$ is unique and satisfies the first-order condition of the Nash product.*

[9]See Osborne and Rubinstein (1990) and Muthoo (1999) for textbook treatments of bargaining theory.

[10]Note that this standard version of the alternating-offers model does not incorporate the possibility of opting out of the bargaining. In our set-up, the interpretation is that while the counties have already agreed on creating a cap and trade system, they only bargain about the implemented cap.

[11]Different discount rates would shift bargaining power in favour of country i that possess a lower discount rate which in turn leads to a bargaining outcome that is close to $\bar{E}_i^*$.

Using Lemma 2, we obtain that $\bar{E}_N^*$ satisfies

$$W_1'\left(\bar{E}\right) + \frac{W_1\left(\bar{E}\right)}{W_2\left(\bar{E}\right)} \cdot W_2'\left(\bar{E}\right) = 0. \tag{27}$$

After inserting, rearranging and simplifying the FOC reads as follows:

$$C_1'(a_1(\bar{E}))\left[\mu + (1-\mu)\theta(\bar{E})\right] = B_1'(A(\bar{E})) + \theta(\bar{E})B_2'(A(\bar{E})) + x_{12}\frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}\left(1-\theta(\bar{E})\right), \tag{28}$$

$$\textit{where } x_{12} = \mu\bar{E} - (e_{1,0} - a_1(\bar{E})) \textit{ and } \theta(\bar{E}) = \frac{B_1(A(\bar{E})) - C_1(a_1(\bar{E})) + x_{12}C_1'(a_1(\bar{E}))}{B_2(A(\bar{E})) - C_2(a_2(\bar{E})) - x_{12}C_1'(a_1(\bar{E}))}.$$

x_{12} denotes the amount of certificates country 1 sells to country 2. In fact, x_{12} can also be negative implying that country 1 buys the corresponding amount of certificates from country 2. Hence, θ represents the ratio of country 1's welfare to country 2's welfare.

4.2.2 Comparison to the Social Optimum

Now we seek to explore the question whether the bargaining solution is socially optimal. It is worth emphasizing that each total abatement in the cap and trade system is achieved with the optimal cost structure, namely with equal marginal cost in each country. Hence, if $\bar{E}_N^* = \bar{E}_S^*$ then this automatically implies that the bargaining solution is socially optimal. Comparing the FOCs, leads to Lemma 3 for the proof see Appendix C.

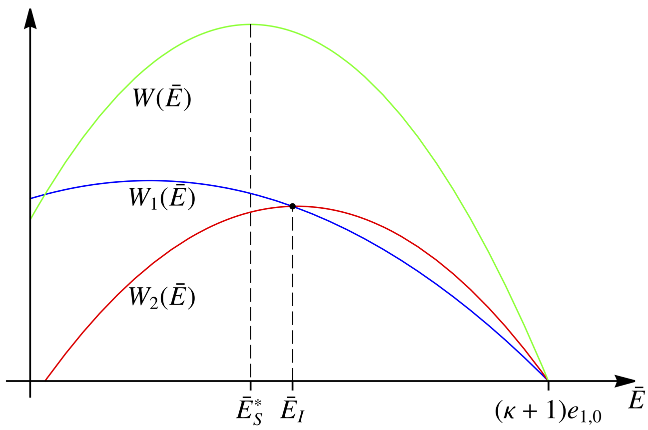
Lemma 3 *Bargaining implements the efficient cap if and only if* $\theta(\bar{E}_N^*) = 1$.

Intuitively, if $\bar{E}_N^*$ satisfies that $W_1\left(\bar{E}_N^*\right) = W_2\left(\bar{E}_N^*\right)$, then Equation (27) coincides with the the FOC for the socially optimal cap. Although the countries are different in terms of cost- and benefit structures, bargaining can only lead to a socially optimal outcome, if the bargaining solution generates the same welfare for both countries. As we investigate next, this is only satisfied for particular combinations of costs, benefits, emissions, and distribution of the certificates.

In our model it is reasonable to assume that the countries' cost and benefit structures as well as the emissions under "business as usual" are exogenously given and cannot be changed. The allocation of the certificates, however, is determined among the countries before the negotiation takes place. Therefore, the interesting question is, whether a allocation μ_S^* exists such that the consecutive bargaining results in the socially efficient outcome given the other parameters. Proposition 1 establishes the definition of the allocation μ_S^*. For the proof see Appendix D.

Proposition 1 *Bargaining implements the socially optimal cap,* $\bar{E}_N^* = \bar{E}_S^*$*, if and only if the allocation of shares is* μ_S^**, where*

Figure 2: Distribution of Welfare for $\beta_1 > \overline{\beta}_1$.



$$\mu_S^* = \frac{e_{1,0}\left[2\delta(\zeta_1 + \zeta_2) + \zeta_1\zeta_2\right] - \delta\frac{(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_2}(\beta_1 - \beta_2) - \frac{1}{4}\left[(3\zeta_1 + 5\zeta_2)\beta_1 - (\zeta_1 - \zeta_2)\beta_2\right]}{(\kappa + 1)e_{1,0}\left[2\delta(\zeta_1 + \zeta_2) + \zeta_1\zeta_2\right] - (\zeta_1 + \zeta_2)(\beta_1 + \beta_2)}. \tag{29}$$

Hence, if the ex-ante determined allocation of the certificates is μ_S^*, the bargaining then results in the socially optimal cap. For any other distribution of the certificates, the countries agree on a cap that is not optimal for the overall welfare. Furthermore, note that the numerator of Equation (29) is decreasing in β_1, this leads us to Corollary 1.

Corollary 1 *There exists a $\overline{\beta}_1$ such that $\mu_S^* = 0$, and μ_S^* is strictly decreasing for $\beta_1 \in [\beta_2, \overline{\beta}_1]$. For $\beta_1 > \overline{\beta}_1$ bargaining cannot implement $\bar{E}_S^*$.*

For the proof see Appendix E. Graphically, the mechanism is as follows, Equation (27) implies that bargaining leads to a socially optimal solution whenever $\bar{E}_I$, the cap for which W_1 and W_2 intersect, equals the socially optimal cap $\bar{E}_S^*$. As illustrated in Figure 2, if $\beta_1 > \overline{\beta}_1$, then we have that $\bar{E}_I > \bar{E}_S^*$ for $\mu = 0$. According to Lemma 1 an increase in μ shifts the maximum of W_1 up and to the left while the opposite effect occurs for W_2. Hence, increasing μ increases $\bar{E}_I$ while it has no effect on $\bar{E}_S^*$. $\beta_1 > \overline{\beta}_1$ therefore implies that $\bar{E}_I > \bar{E}_S^*$ for all possible distributions of certificates $\mu \in [0, 1]$.

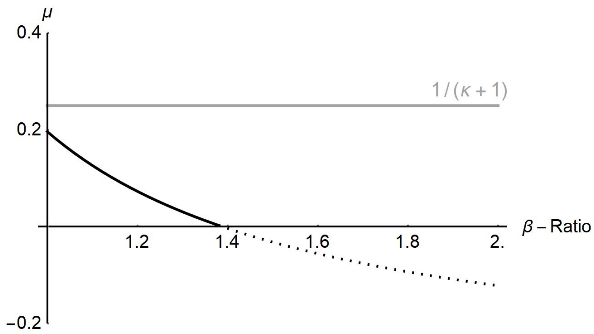
Intuitively, if β_1 is increased then country 1 benefits more from abatement. Hence, a lower cap is optimal for both, country 1 as well as for the total welfare. Country 2 needs to be compensated to agree on that lower cap in a bargaining. This compensation works via an increased share of certificates for country 2 (and therefore a lower share for country 1). Hence, as indicated by (Buchholz et al., 2005), the allocation of permits is used for implicit side payments in an emissions trading system. As a result, a country has to purchase fewer certificates or receive additional revenue for selling the certificates, depending on whether the country is a buyer or seller of allowances. If country 2 receives all the certificates, i.e., $\mu_S^* = 0$, then the entire scope for compensating country 2 is used. A further increase in β_1 would decrease the socially optimal cap, but this cap could not be implemented by the bargaining because there is no means for further compensating country 2. In particular, if the benefits for country 1 are too high, then there is no distribution of certificates among the countries that provides both countries with the same welfare. Consequently, bargaining does not lead to a socially optimal outcome in the case where the countries are strongly dissimilar, i.e., where the countries substantially differ in emission (κ sufficiently high) and benefits ($\beta_1 - \beta_2$ is sufficiently large).

Proposition1 together with Corollary 1 implies the following for the optimal allocation of certificates among the countries which is proved in Appendix F:

Corollary 2 *For the optimal allocation of certificates, it holds that* $\mu_S^* < 1/(\kappa + 1)$ *for all* $\beta_1 \in [\beta_2, \overline{\beta}_1]$.

Hence, allocating the certificates based on historical emissions, i.e., according to the proportion of the emissions under business as usual, does never lead to a socially optimal bargaining solution, see Figure 3. To ensure a welfare maximizing bargaining outcome the allocation of certificates needs to take not only the distribution of the emissions but also cost and benefit structures into account.

Figure 3: Optimal μ for $\kappa = 3$.

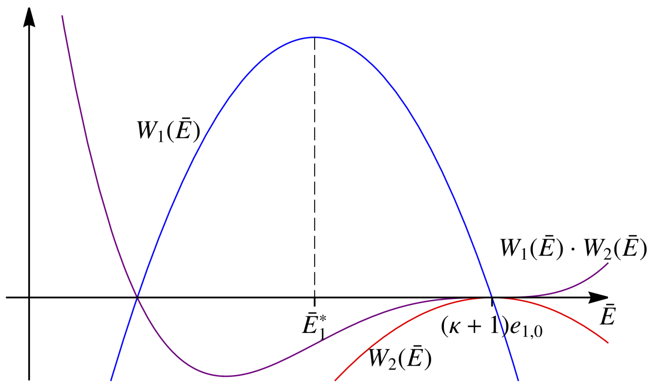


Parameter setting: $\zeta_1 = 0.9$, $\zeta_2 = 0.1$, $e_{1,0} = 100$, $\delta = 0.1$, $\beta_1 = 30$.

4.2.3 Comparison to the National Caps

Next, we compare the bargaining solution to the decentralized outcome where both countries deploy national caps. The welfare induced by the bargaining solution relies crucially on μ, i.e., the allocation of the certificates among the countries. As we have seen above, for $\beta_1 \in [\beta_2, \overline{\beta}_1]$

Figure 4: Optimal caps for county 1 and 2.



bargaining can result in the socially optimal solution, if $\mu = \mu_S^*$. In that case, we obtain that $W_N^* = W_S^*$. However, there also exists a μ_i^0 such that country i's optimal cap equals the total emissions under "business as usual". Interestingly, country i can enforce that cap in bargaining, i.e., the bargaining leads to $\bar{E}_N^* = (\kappa + 1)e_{1,0}$. According to Equations (8) and (9) this leads to zero abatement activities, which, in turn, induces a welfare of zero, $W_N^* = 0$. See Figure 4 for a graphical representation of the case where country 2 enforces the total emissions under "business as usual" as bargaining outcome. The intuition is as follows: since the distribution of certificates is in favour of $-i$, country i has nothing to gain in the bargaining. In fact, perpetually rejecting every offer made by $-i$ leaves country i better off, except for the offer of a cap that equals the emissions under "business as usual", which makes country i indifferent between accepting and rejecting the offer.[12]

Using Equations (24) and (25), we can explicitly solve $E_i^* = (\kappa + 1)e_{1,0}$ for μ_i^0, which yields

$$\mu_1^0 = \frac{1 - \left(\frac{\zeta_1 + \zeta_2}{\zeta_1 \zeta_2} \frac{\beta_1}{e_{1,0}}\right)}{\kappa + 1}, \tag{30}$$

$$\mu_2^0 = \frac{1 + \left(\frac{\zeta_1 + \zeta_2}{\zeta_1 \zeta_2} \frac{\beta_2}{e_{1,0}}\right)}{\kappa + 1}. \tag{31}$$

The finding that the bargaining results in emissions under "business as usual" and a welfare of zero, both individually and socially, immediately carries over to the case where a country demands a cap higher than "business as usual", $\bar{E}_i^* > (\kappa + 1)e_{1,0}$. Lemma 4 summarizes these results:

Lemma 4 *For $\mu \in [0, 1]$ the welfare generated by the bargaining solution is $W_S^* \geq W_N^* \geq 0$. If*

[12]In fact, the subgame perfect equilibrium in the Rubinstein alternating-offers is not unique in this case. Therefore, it is neither guaranteed that an agreement is struck in round 1 nor that it is reached at all. In terms of welfare, however, it is irrelevant whether countries remain in the status quo because they have agreed on it or because they permanently disagree on how to change the status quo.

$\mu \leq \mu_1^0$ *or* $\mu \geq \mu_2^0$ *bargaining results in* $\bar{E}_N^* = (\kappa+1)e_{1,0}$ *and* $W_N^* = 0$.

For *national CAPs*, however, each country obtains a welfare greater than zero. By definition of a Nash equilibrium country i's cap is the optimal response to the cap chosen by the other country $-i$. If country i deviates and chooses $\bar{e}_i = e_{i,0}$, i.e., zero abatement $a_i = 0$, instead of $\bar{e}_i = \bar{e}_{i,C}^*$ then this leads to a strictly reduced welfare for country i. But even this deviation generates welfare $W_i \geq 0$ since $\bar{e}_{-i,C}^* \leq e_{-i,0}$ because country i still benefits from the abatement of county $-i$ but does not experience any costs. Therefore, we obtain that $W_{i,C}^* > 0$ and $W_C^* = W_{1,C}^* + W_{2,C}^* > 0$.

As might be reasonably expected, the counter factual leads to a lower welfare compared to the social optimum, since the social planner takes the costs and benefits of both countries into account, whereas in the counterfactual scenario each country optimizes separately without considering cross-border benefits and cheap abatement options. This leads to two different effects.

First, the overall cap is *higher* in the counter factual scenario, i.e., the total abatement activities are lower. More precisely, using Equations (14) and (21), we can calculate the difference to the optimal cap as

$$\bar{E}_C^* - \bar{E}_S^* = \frac{\beta_1 + \beta_2}{2\delta + \frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}} \left[1 - \frac{1 + \frac{\delta}{\delta + \frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}}}{2 + \frac{(\zeta_1-\zeta_2)(\beta_1-\beta_2)}{\zeta_2\beta_1+\zeta_1\beta_2}} \right] > 0. \tag{32}$$

For $\beta_1 > \beta_2$, the difference is positive and increasing in β_1, which implies that the higher the difference $\beta_1 - \beta_2$, the higher the difference from the overall abatement activities in the counter factual scenario to the socially optimal abatement. Intuitively, since its benefits are increased country 1 sets a lower national cap $\bar{e}_{1,C}^*$ to implement higher national abatement. Country 2, on the contrary, can free ride on this abatement activities by setting a higher nationally cap $\bar{e}_{2,C}^*$ thereby reducing its cost. As a result, the total cap increases less than it would be socially optimal.

Second, the abatement activities induced by the $\bar{E}_C^*$ are *inefficient*, i.e., the same total abatement could be implemented with lower costs. A certain overall abatement is implemented efficiently if the abatement activities are distributed among the countries such that their marginal costs are equal. The cap and trade system provides the efficient distribution, where we have that

$$\frac{a_1}{a_2} = \frac{\zeta_2}{\zeta_1}. \tag{33}$$

For national CAPs, however, we obtain the following distribution among the countries:

$$\frac{a_1}{a_2} = \frac{\zeta_2\beta_1 + \delta(\beta_1 - \beta_2)}{\zeta_1\beta_2 - \delta(\beta_1 - \beta_2)} \tag{34}$$

For $\beta_1 > \beta_2$, country 1 abates too much, while country 2 abates too little. Moreover, this ratio is increasing in β_1, i.e., the higher difference $\beta_1 - \beta_2$ the higher the inefficiency. Lemma 5 summarizes the results for national CAPs.

Lemma 5 *The overall welfare of national CAPs, W_C^*, is greater than zero but smaller than the socially optimal, i.e., $0 < W_C^* < W_S^*$. The difference to the social optimum, $W_S^* - W_C^*$, increases in β_1.*

Combining Lemma 4 and Lemma 5 yields the following proposition.

Proposition 2 *For $\mu \in [0, 1]$ and $\beta_1 \in [\beta_2, \overline{\beta}_1]$, we have that $W_N^* > W_C^*$ if μ is sufficiently close to μ_S^* and $W_N^* < W_C^*$ if μ is sufficiently close to either μ_1^0 or μ_2^0.*

This is particularly the case, when the countries' benefit structures differ substantially, i.e., β_1 close to $\overline{\beta}_1$. National CAPs then lead to a significantly lower welfare compared to the social optimum due to *i*) substantial deviations from the optimal abatement *ii*) highly inefficient distribution of the abatement activities among the countries. Therefore, bargaining also leads to a significantly better outcome than national caps, if the ex-ante determined allocation of certificates is significantly close to zero.

5 Conclusion

In our two-country model, countries have already committed to jointly implement an emissions trading scheme and bargain over the total emissions cap. We model the cap negotiation as an alternating-offers model, following Rubinstein (1982), and show that even for moderate differences in baseline emissions the initial endowment of permit rights based on historical emissions never leads to the social optimum. However, we can determine allocations that lead to the optimum, where the allocation of allowances can be seen as a form of compensation payment between countries. In this case, the country with high emissions but low benefits and costs of abatement receives a higher share of allowances, resulting in a more stringent emissions cap. This compensation mechanism is only possible to a certain extent and depends on the countries' benefit functions.

If the countries' benefit structures of abatement are too different ($\beta_1 > \bar{\beta}_1$) we cannot find an allocation of allowances between countries that implements the social optimum. This results from the fact that the redistribution of the initial endowment is limited, as a country cannot receive more than the total quantity of allowances.

Although it is not always possible to achieve the socially optimal level of emissions, from a global perspective, bargaining can lead to a better solution than having national emission trading schemes with a national emissions cap instead.

We are aware that we have made a strong modeling assumptions, namely the countries have already agreed to implement a joint emissions trading system in a sense that they cannot opt out of the bargaining. Further research is necessary to verify our results in extensions of the alternating-offers model that allow the bargaining to end without an agreement. The most plausible approach is by allowing the parties to strategically opt out of the bargaining such that national caps are implemented. Integrating this outside option still allows to exploit the relation between the subgame perfect equilibrium in the alternating-offers model and the Nash bargaining solution. Merely the set of feasible caps for an agreement is weakly smaller, since these caps must not only be element of $\mathscr{P}$ but also ensure that each country gets a weakly higher welfare than in the case of national caps (Binmore, 1985; Muthoo, 1999).

Another alternative approach for allowing for the collapse of the bargaining is the integration of the risk of a random breakdown in the sense that one party gets suddenly fed up and leaves the negotiating table such that national caps are implemented [13]. Again, it is possible to exploit the relation to the Nash bargaining solution, where the Nash product now takes the form $(W_1 - d_1) \cdot (W_2 - d_2)$. In this setting, the disagreement point (d_1, d_2) is calculate from the countries welfare in case of national caps and the arrival rates of the breakdown (Binmore et al., 1986; Muthoo, 1999).

[13] Essentially this can be seen as an "agreement to disagree", which is plausible in behavioural settings (Binmore et al., 1986), while it is inconsistent for rational agents with common knowledge (Aumann, 1976).

References

Arvaniti, M., Habla, W., 2021. The political economy of negotiating international carbon markets. Journal of Environmental Economics and Management 110, 102521.

Aumann, R., 1976. Agreeing to disagree. Annals of Statistics 4, 1236–1239.

Barrett, S., 1994. Self-enforcing international environmental agreements. Oxford economic papers , 878–894.

Binmore, K., 1985. Bargaining and coalitions. Cambridge University Press. p. 269–304.

Binmore, K., 1987. Nash bargaining theory II. Blackwell. pp. 61–76.

Binmore, K., Rubinstein, A., Wolinsky, A., 1986. The nash bargaining solution in economic modelling. Rand Journal of Economics 17, 176–188.

Buchholz, W., Haupt, A., Peters, W., 2005. International environmental agreements and strategic voting. Scandinavian Journal of Economics 107, 175–195.

Burtraw, D., 1992. Strategic delegation in bargaining. Economics Letters 38, 181–185.

Burtraw, D., 1993. Bargaining with noisy delegation. Rand Journal of Economics , 40–57.

Coase, R.H., 1960. The problem of social cost. Journal of Law and Economics 3, 1–44.

Crawford, V.P., Varian, H.R., 1979. Distortion of preferences and the nash theory of bargaining. Economics Letters 3, 203–206.

Dales, J., 1968. Pollution, property & prices; an essay in policy-making and economics. University of Toronto Press.

Dijkstra, B.R., Nentjes, A., 2020. Pareto-efficient solutions for shared public good provision: Nash bargaining versus exchange-matching-lindahl. Resource and Energy Economics 61, 101179.

Doda, B., Quemin, S., Taschini, L., 2019. Linking permit markets multilaterally. Journal of Environmental Economics and Management 98, 102259.

Doda, B., Taschini, L., 2017. Carbon dating: When is it beneficial to link etss? Journal of the Association of Environmental and Resource Economists 4 (3), 701–730.

Flachsland, C., Marschinski, R., Edenhofer, O., 2009. To link or not to link: benefits and disadvantages of linking cap-and-trade systems. Climate Policy 9, 358–372.

Habla, W., Winkler, R., 2018. Strategic delegation and international permit markets: Why linking may fail. Journal of environmental economics and management 92, 244–250.

Hahn, R.W., 1984. Market power and transferable property rights. The Quarterly Journal of Economics 99, 753–765.

Helm, C., 2003. International emissions trading with endogenous allowance choices. Journal of Public Economics 87, 2737–2747.

Jones, S.R., 1989. Have your lawyer call my lawyer: Bilateral delegation in bargaining situations. Journal of Economic Behavior & Organization 11, 159–174.

Loeper, A., 2017. Cross-border externalities and cooperation among representative democracies. European Economic Review 91, 180–208.

McGinty, M., 2007. International environmental agreements among asymmetric nations. Oxford Economic Papers 59, 45–62.

Montgomery, W.D., 1972. Markets in licenses and efficient pollution control programs. Journal of Economic Theory 5, 395–418.

Muthoo, A., 1999. Bargaining theory with applications. Cambridge University Press.

Nash, J.F., 1950. The bargaining problem. Econometrica 18(2), 155–162.

Osborne, M.J., Rubinstein, A., 1990. Bargaining and markets. Academic Press.

Rubinstein, A., 1982. Perfect equilibrium in a bargaining model. Econometrica 50, 97–109.

Segendorff, B., 1998. Delegation and threat in bargaining. Games and Economic Behavior 23, 266–283.

Siqueira, K., 2003. International externalities, strategic interaction, and domestic politics. Journal of Environmental Economics and Management 45, 674–691.

Smead, R., Sandler, R.L., Forbes, P., Basl, J., 2014. A bargaining game analysis of international climate negotiations. Nature Climate Change 4, 442–445.

Stavins, R.N., 1995. Transaction costs and tradeable permits. Journal of Environmental Economics and Management 29, 133–148.

Weitzman, M.L., 1974. Prices vs. quantities. Review of Economic Studies 41, 477–491.

Weitzman, M.L., 2014. Can negotiating a uniform carbon price help to internalize the global warming externality? Journal of the Association of Environmental and Resource Economists 1, 29–49.

Appendix

A Proof of Lemma 1

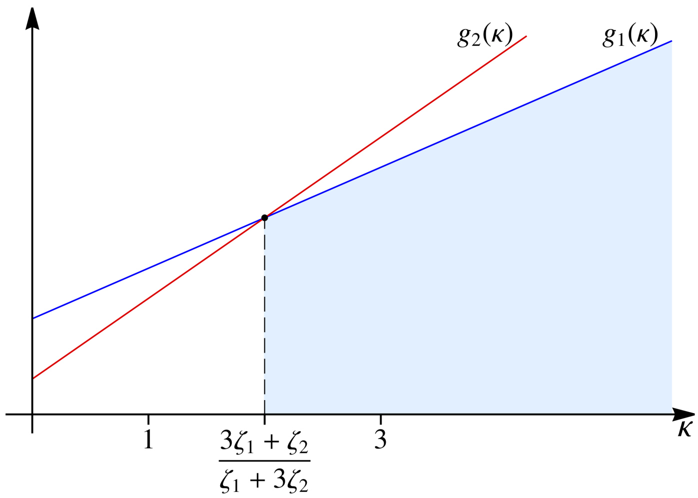
Differentiating (24) and (25) with respect to μ yields

$$\frac{\partial \bar{E}_1^*}{\partial \mu} > [<] 0 \iff \frac{\beta_1}{e_{1,0}} > [<] \overbrace{\frac{1}{2}\left[\delta + \frac{\zeta_1\zeta_2(2\zeta_1+\zeta_2)}{(\zeta_1+\zeta_2)^2}\right] + \frac{\kappa}{2}\left[\delta - \zeta_1\left(\frac{\zeta_2}{\zeta_1+\zeta_2}\right)^2\right]}^{g_1(\kappa):=} \tag{35}$$

$$\frac{\partial \bar{E}_2^*}{\partial \mu} > [<] 0 \iff \frac{\beta_2}{e_{1,0}} < [>] \overbrace{\frac{1}{2}\left[\delta - \zeta_2\left(\frac{\zeta_1}{\zeta_1+\zeta_2}\right)^2\right] + \frac{\kappa}{2}\left[\delta + \frac{\zeta_1\zeta_2(\zeta_1+2\zeta_2)}{(\zeta_1+\zeta_2)^2}\right]}^{g_2(\kappa):=}. \tag{36}$$

The light blue area in Figure 5 shows the set of feasible $\beta_1/e_{1,0}$ given Assumptions (ii) and (iii). Hence, we have that $\partial \bar{E}_1^*/\partial\mu < 0$. Since we additional get for $\kappa > (3\zeta_1+\zeta_2)/(\zeta_1+3\zeta_2)$ that $g_2(\kappa) > g_1(\kappa) > \beta_1/e_{1,0} > \beta_2/e_{1,0}$, we can immediately conclude that $\partial \bar{E}_2^*/\partial\mu > 0$.

Figure 5: Feasible parameter sets of $\beta_1/e_{1,0}$.



□

B Proof of Lemma 2

Let us first consider the case where $\bar{E}_1^* < \bar{E}_2^* < (\kappa+1)e_{1,0}$, see Figure 1. Evaluating the derivative of the Nash product at the lower bound of $\mathscr{P}$, we obtain, by definition of $\bar{E}_1^*$, that

$$W_1'\left(\bar{E}_1^*\right)\cdot W_2\left(\bar{E}_1^*\right) + W_1\left(\bar{E}_1^*\right)\cdot W_2'\left(\bar{E}_1^*\right) = W_1\left(\bar{E}_1^*\right)\cdot W_2'\left(\bar{E}_1^*\right) > 0. \tag{37}$$

By contrast, evaluating the derivative of the Nash product at the upper bound of $\mathscr{P}$, we obtain, by definition of $\bar{E}_2^*$, that

$$W_1'\left(\bar{E}_2^*\right)\cdot W_2\left(\bar{E}_2^*\right) + W_1\left(\bar{E}_2^*\right)\cdot W_2'\left(\bar{E}_2^*\right) = W_1'\left(\bar{E}_2^*\right)\cdot W_2\left(\bar{E}_2^*\right) < 0. \tag{38}$$

The Intermediate Value Theorem implies that a solution to the FOC exists. Moreover, we have that $W_1'\left(\bar{E}\right)\cdot W_2\left(\bar{E}\right)$ as well as $W_1\left(\bar{E}\right)\cdot W_2'\left(\bar{E}\right)$ are decreasing for $\bar{E}\in\mathscr{P}$, such that the Nash product is strictly concave on that interval. Thus, the FOC has a unique solution for $\bar{E}\in\mathscr{P}$

which constitutes a global maximum.
Next consider the case where $\bar{E}_1^* < (\kappa+1)e_{1,0} \le \bar{E}_2^*$, see Figure 4. In that case, we have that

$$W_1'\Big((\kappa+1)e_{1,0}\Big)\cdot W_2\Big((\kappa+1)e_{1,0}\Big) + W_1\Big((\kappa+1)e_{1,0}\Big)\cdot W_2'\Big((\kappa+1)e_{1,0}\Big) = 0. \tag{39}$$

Furthermore, by construction, the Nash product is zero for $(\kappa+1)e_{1,0}$ and strictly negative for all other $\bar{E} \in \mathscr{P}$. Hence, $(\kappa+1)e_{1,0}$ constitutes the unique global maximum for $\bar{E} \in \mathscr{P}$ and it satisfies the FOC.
For the cases where $\bar{E}_2^* < \bar{E}_1^* < (\kappa+1)e_{1,0}$ and $\bar{E}_2^* < (\kappa+1)e_{1,0} \le \bar{E}_1^*$, the same arguments apply only the indices are reversed.[14] □

C Proof of Lemma 3

First, note that $\bar{E}_S^*$ and $\bar{E}_N^*$ are unique by Equation 14 and Lemma 2, respectively. According to Lemma 2, the Nash bargaining solution satisfies

$$\begin{aligned} &C_1'(a_1(\bar{E}_N^*))\left[\mu + (1-\mu)\theta(\bar{E}_N^*)\right] \\ &= B_1'(A(\bar{E}_N^*)) + \theta(\bar{E}_N^*)B_2'(A(\bar{E}_N^*)) + x_{12}\frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}\left(1-\theta(\bar{E}_N^*)\right), \end{aligned} \tag{40}$$

while it holds for the efficient cap that

$$C_1'(a_1(\bar{E}_S^*)) = B_1'(A(\bar{E}_S^*)) + B_2'(A(\bar{E}_S^*)). \tag{41}$$

If we have that $\theta(\bar{E}_N^*) = 1$, then Equation (40) simplifies to

$$C_1'(a_1(\bar{E}_N^*)) = B_1'(A(\bar{E}_N^*)) + B_2'(A(\bar{E}_N^*)). \tag{42}$$

Comparing to Equation 41 immediately yields that $\bar{E}_S^* = \bar{E}_N^*$. Now, consider the opposite direction. If we have that $\bar{E}_S^* = \bar{E}_N^*$, then Equation 41 implies that

$$C_1'(A(\bar{E}_N^*)) = B_1'(A(\bar{E}_N^*)) + B_2'(A(\bar{E}_N^*)).$$

A comparison to Equation 40 directly reveals that we must have $\theta(\bar{E}_N^*) = 1$. □

D Proof of Proposition 1

First, we establish a relation that we will use throughout the proof:

$$x_{12} = \mu\bar{E} - (e_{1,0} - a_1(\bar{E})) = -(1-\mu)\bar{E} + (\kappa e_{1,0} - a_2(\bar{E})). \tag{43}$$

Now we start the proof by using the definition of $\theta(\bar{E}) = 1$ if and only if $\bar{E}_S^* = \bar{E}_N^*$. Inserting the definition of $\theta(\bar{E})$ yields

$$\frac{B_1(A(\bar{E}_N^*)) - C_1(a_1(\bar{E}_N^*)) + x_{12}C_1'(a_1(\bar{E}_N^*))}{B_2(A(\bar{E}_N^*)) - C_2(a_2(\bar{E}_N^*)) - x_{12}C_1'(a_1(\bar{E}_N^*))} = 1. \tag{44}$$

Rearranging this expression and using the functional forms of $B_i(A(\bar{E}))$ and $C_i(a_i(\bar{E}))$ leads to

$$-\frac{(\zeta_1+\zeta_2)(\beta_1-\beta_2)}{2\zeta_1\zeta_2} - \frac{\left[(\kappa+1)e_{1,0} - \bar{E}_N^*\right](\zeta_1-\zeta_2)}{4(\zeta_1+\zeta_2)} = x_{12}. \tag{45}$$

Inserting Equation (43) and rearranging for μ yields

$$1 + \frac{a(\bar{E}_N^*) - \kappa e_{1,0}}{\bar{E}} - \frac{(\zeta_1+\zeta_2)(\beta_1-\beta_2)}{2\zeta_1\zeta_2\bar{E}_N^*} - \frac{\left[(\kappa+1)e_{1,0} - \bar{E}_N^*\right](\zeta_1-\zeta_2)}{4(\zeta_1+\zeta_2)\bar{E}_N^*} = \mu. \tag{46}$$

[14] Note that the cases where $(\kappa+1)e_{1,0} \le \bar{E}_1^* \le \bar{E}_2^*$ and $(\kappa+1)e_{1,0} \le \bar{E}_2^* \le \bar{E}_1^*$ cannot occur in our model due to the quadric structure of benefits and costs.

Since we have that $\bar{E}_N^* = \bar{E}_S^*$, we can insert Equation (14). Simplifying finally leads to the desired result:

$$\mu_S^* = \frac{e_{1,0}\left[2\delta(\zeta_1+\zeta_2)+\zeta_1\zeta_2\right] - \delta\frac{(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_2}(\beta_1-\beta_2) - \frac{1}{4}\left[(3\zeta_1+5\zeta_2)\beta_1-(\zeta_1-\zeta_2)\beta_2\right]}{(\kappa+1)e_{1,0}\left[2\delta(\zeta_1+\zeta_2)+\zeta_1\zeta_2\right]-(\zeta_1+\zeta_2)(\beta_1+\beta_2)}. \tag{47}$$

□

E Proof of Corollary 1

First, we turn to $\overline{\beta}$. Note that the numerator of (29) is decreasing in β_1. Hence, $\mu_S^* = 0$ if and only if

$$e_{1,0}\left[2\delta(\zeta_1+\zeta_2)+\zeta_1\zeta_2\right] - \delta\frac{(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_2}(\beta_1-\beta_2) - \frac{1}{4}\left[(3\zeta_1+5\zeta_2)\beta_1-(\zeta_1-\zeta_2)\beta_2\right] = 0 \tag{48}$$

Solving Equation (48) for β_1 leads us to

$$\overline{\beta}_1 = \frac{4\zeta_1\zeta_2 e_{1,0}\left[2\delta(\zeta_1+\zeta_2)+\zeta_1\zeta_2\right]+\beta_2\left[4\delta(\zeta_1+\zeta_2)^2+\zeta_1\zeta_2(\zeta_1-\zeta_2)\right]}{\zeta_1\zeta_2(3\zeta_1+5\zeta_2)+4\delta(\zeta_1+\zeta_2)^2} \tag{49}$$

Second, note that the denominator of (29) is positive if only if:

$$\frac{\beta_1+\beta_2}{e_{1,0}} < (\kappa+1)\left(2\delta+\frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}\right) \tag{50}$$

Since we have that

$$\begin{aligned}
\frac{\beta_1+\beta_2}{e_{1,0}} &< 2\frac{\beta_1}{e_{1,0}} \\
&< \left[\delta+\frac{\zeta_1\zeta_2(2\zeta_1+\zeta_2)}{(\zeta_1+\zeta_2)^2}\right]+\kappa\left[\delta-\zeta_1\left(\frac{\zeta_2}{\zeta_1+\zeta_2}\right)^2\right] \\
&= \left[\delta+\zeta_2\left(\frac{\zeta_1}{(\zeta_1+\zeta_2)}\right)^2+\frac{\zeta_1\zeta_2}{(\zeta_1+\zeta_2)}\right]+\kappa\left[\delta-\zeta_1\left(\frac{\zeta_2}{\zeta_1+\zeta_2}\right)^2\right] \\
&< \left[2\delta+\frac{\zeta_1\zeta_2}{(\zeta_1+\zeta_2)}\right]+\kappa\left[\delta-\zeta_1\left(\frac{\zeta_2}{\zeta_1+\zeta_2}\right)^2\right] \\
&< (\kappa+1)\left(2\delta+\frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}\right),
\end{aligned}$$

Inequality (50) holds and we must have that the denominator of (29) is always positive. Because $g_1(\kappa)$ is increasing in κ, cf. Appendix A while $\overline{\beta}_1/e_{1,0}$ is independent of κ we must necessarily have that $\overline{\beta}_1/e_{1,0} < g_1(\kappa)$ for κ sufficiently large. In other words, for κ sufficiently large, we obtain feasible ratios $\beta_1/e_{1,0}$ where $\mu_S^* < 0$.

To determine $\partial\mu_S^*/\partial\beta_1$ we need to apply the quotient rule. The sign of the derivative, however, is determined by sign of the numerator of the resulting quotient. Hence, we get that

$$\begin{aligned}
\frac{\partial\mu_S^*}{\partial\beta_1} \overset{\text{sign}}{=} &-\left[\frac{\delta(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_1}+\frac{1}{4}(3\zeta_1+5\zeta_2)\right]\left[(\kappa+1)\eta-(\zeta_1+\zeta_2)(\beta_1+\beta_2)\right] \\
&+\left[(\zeta_1+\zeta_2)\right]\left[\eta-\frac{\delta(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_2}-\frac{1}{4}\left[(3\zeta_1+5\zeta_2)\beta_1-(\zeta_1-\zeta_2)\beta_2\right]\right]
\end{aligned} \tag{51}$$

where we defined $\eta = e_{1,0}[2\delta(\zeta_1+\zeta_2)+\zeta_1\zeta_2]$. Now, note that the r.h.s. of 51 is linear in β_1, i.e. it has a single root. Hence, we can conclude that μ_S^* is decreasing in β_1 on the entire interval $[\beta_2, \overline{\beta}_1]$, if we have that $\partial\mu_S^*/\partial\beta_1 < 0$ for $\beta_1 \to \beta_2$ and for $\beta_1 \to \overline{\beta}_1$.[15] First, we turn to $\beta_1 \to \overline{\beta}_1$. By definition of $\overline{\beta}_1$, we get that

$$\lim_{\beta_1 \to \overline{\beta}_1} \frac{\partial\mu_S^*}{\partial\beta_1} \overset{\text{sign}}{=} -\left[\frac{\delta(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_1} + \frac{1}{4}(3\zeta_1+5\zeta_2)\right]\left[(\kappa+1)\eta - (\zeta_1+\zeta_2)(\overline{\beta}_1+\beta_2)\right] < 0. \tag{52}$$

Second, let us analyze $\beta_1 \to \beta_2$, where we obtain that

$$\begin{aligned}\lim_{\beta_1 \to \beta_2} \frac{\partial\mu_S^*}{\partial\beta_1} \overset{\text{sign}}{=} &- \left[\frac{\delta(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_1} + \frac{1}{4}(3\zeta_1+5\zeta_2)\right]\left[(\kappa+1)\eta - (\zeta_1+\zeta_2)2\beta_2\right] \\ &+ \left[(\zeta_1+\zeta_2)\right]\left[\eta - \frac{\delta(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_2} - \frac{7}{4}\zeta_1\beta_2 - \frac{1}{4}\zeta_2\beta_2\right].\end{aligned} \tag{53}$$

To establish the negative sign, it is sufficient to show that the factors in the first line in 53 are greater than the factors in the second line. For the first factor, we obtain that

$$\begin{aligned}\frac{\delta(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_1} + \frac{1}{4}(3\zeta_1+5\zeta_2) &> \zeta_1 + \frac{1}{4}(3\zeta_1+5\zeta_2) \\ &= \frac{7}{4}\zeta_1 + \frac{5}{4}\zeta_2 \\ &> \zeta_1 + \zeta_2.\end{aligned} \tag{54}$$

Comparing the second factors leads to

$$(\kappa+1)\eta - (\zeta_1+\zeta_2)2\beta_2 > \eta - \frac{\delta(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_2} - \frac{7}{4}\zeta_1\beta_2 - \frac{1}{4}\zeta_2\beta_2, \tag{55}$$

which can be rearranged to

$$\frac{4\,\kappa\,\eta}{\zeta_1+7\zeta_2} + \frac{\delta(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_2(\zeta_1+7\zeta_2)} > \beta_2. \tag{56}$$

To see that Inequality 56 is indeed always satisfied, note that

$$\begin{aligned}\frac{4\,\kappa\,\eta}{\zeta_1+7\zeta_2} + \frac{\delta(\zeta_1+\zeta_2)^2}{\zeta_1\zeta_2(\zeta_1+7\zeta_2)} &> \frac{4\,\kappa\,\eta}{\zeta_1+7\zeta_2}, \\ &= 4\,\kappa\,e_{1,0}\left[2\delta\,\overbrace{\frac{(\zeta_1+\zeta_2)}{\zeta_1+7\zeta_2}}^{>\frac{1}{4}} + \frac{\zeta_1\zeta_2}{\zeta_1+7\zeta_2}\right], \\ &> \frac{e_{1,0}}{2}\left[4\delta\,\kappa + 8\kappa\frac{\zeta_1\zeta_2}{\zeta_1+7\zeta_2}\right],\end{aligned} \tag{57}$$

which we can further estimate downwards, since $\kappa > 1$, to

$$\begin{aligned}&> \frac{e_{1,0}}{2}\left[2\delta\,\kappa + \delta + \zeta_2\left(\frac{\zeta_1}{\zeta_1+\zeta_2}\right)^2 + \frac{\zeta_1\zeta_2}{\frac{1}{8}\zeta_1+\frac{7}{8}\zeta_2}\right], \\ &> \frac{e_{1,0}}{2}\left[\kappa\left[\delta - \zeta_1\left(\frac{\zeta_2}{\zeta_1+\zeta_2}\right)^2\right] + \delta + \zeta_2\left(\frac{\zeta_1}{\zeta_1+\zeta_2}\right)^2 + \frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}\right], \\ &= \frac{e_{1,0}}{2}\left[\delta + \frac{\zeta_1\zeta_2(2\zeta_1+\zeta_2)}{(\zeta_1+\zeta_2)^2} + \kappa\left[\delta - \zeta_1\left(\frac{\zeta_2}{\zeta_1+\zeta_2}\right)^2\right]\right], \\ &> \beta_1 = \beta_2.\end{aligned}$$

Hence, we have that

$$\lim_{\beta_1 \to \beta_2} \frac{\partial\mu_S^*}{\partial\beta_1} < 0,$$

which implies that μ_S^* is decreasing in β_1 on the entire interval $[\beta_2, \overline{\beta}_1]$. □

[15] We focus on the interesting case where $\overline{\beta}_1$ is feasible, i.e. where $\overline{\beta}_1/e_{1,0} < g_1(\kappa)$.

F Proof of Corollary 2

Since we have already established that $\frac{\partial \mu_S^*}{\partial \beta_1} < 0$ for $\beta_1 \in [\beta_2, \overline{\beta}_1]$, we need to show now that

$$\lim_{\beta_1 \to \beta_2} \mu_S^* < \frac{1}{\kappa + 1}. \tag{58}$$

Taking the limits and rearranging leads us to

$$\frac{3\zeta_1 + \zeta_2}{\zeta_1 + 3\zeta_2} < \kappa \tag{59}$$

Which is satisfied according to our assumptions. □

G Proof of Lemma 5

Using the Envelope Theorem, we get for W_S^* that

$$\frac{\partial W_S^*}{\partial \beta_1} = (\kappa + 1)\, e_{1,0} - \bar{E}_S^*. \tag{60}$$

By contrast, differentiating and simplifying yields for W_C^* that

$$\frac{\partial W_C^*}{\partial \beta_1} = (\kappa + 1)\, e_{1,0} - \bar{E}_C^* - \frac{\zeta_2 \beta_1 [\zeta_1 \zeta_2 + 2\delta(\zeta_1 + \zeta_2)] + \delta^2 (\beta_1 - \beta_2)(\zeta_1 + \zeta_2)}{[\zeta_1 \zeta_2 + \delta(\zeta_1 + \zeta_2)]^2}. \tag{61}$$

Subtracting (61) from (60) then leads to

$$\frac{\partial W_S^*}{\partial \beta_1} - \frac{\partial W_C^*}{\partial \beta_1} = \bar{E}_C^* - \bar{E}_S^* + \frac{\zeta_2 \beta_1 [\zeta_1 \zeta_2 + 2\delta(\zeta_1 + \zeta_2)] + \delta^2 (\beta_1 - \beta_2)(\zeta_1 + \zeta_2)}{[\zeta_1 \zeta_2 + \delta(\zeta_1 + \zeta_2)]^2}. \tag{62}$$

Since we have already established in (32) that $\bar{E}_C^* - \bar{E}_S^* > 0$, we can immediately conclude that

$$\frac{\partial W_S^*}{\partial \beta_1} - \frac{\partial W_C^*}{\partial \beta_1} > 0, \tag{63}$$

or, in other words, that the difference $W_S^* - W_C^*$ is increasing in β_1. □

References

Aldy, J. E., Barrett, S., & Stavins, R. N. (2003). Thirteen plus one: A comparison of global climate policy architectures. *Climate Policy*, *3*(4), 373–397. [89]

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, *46*(Supplement 1), 878–894. [89, 92]

Barrett, S. (2003). *Environment and statecraft: The strategy of environmental treaty-making*. Oxford University Press. [89]

Baudry, M., Faure, A., & Quemin, S. (2021). Emissions trading with transaction costs. *Journal of Environmental Economics and Management*, *108*, 102468. [92]

Bertsimas, D., Farias, V. F., & Trichakis, N. (2012). On the efficiency-fairness trade-off. *Management Science*, *58*(12), 2234–2250. [101]

Binmore, K. (1985). Bargaining and coalitions. In A. Roth (Ed.), *Game-theoretic models of bargaining* (pp. 269–304). Cambridge University Press. [105]

Binmore, K. (1987). Nash bargaining theory ii. In K. Binmore & P. Dasgupta (Eds.), *The economics of bargaining* (pp. 61–76). Blackwell. [100]

Binmore, K., & Dasgupta, P. (1987). *The economics of bargaining*. Blackwell. [99]

Binmore, K., Rubinstein, A., & Wolinsky, A. (1986). The nash bargaining solution in economic modelling. *Rand Journal of Economics*, *17*(2), 176–188. [100, 105]

Böhringer, C., & Lange, A. (2005). On the design of optimal grandfathering schemes for emission allowances. *European Economic Review*, *49*(8), 2041–2055. [97]

Buchholz, W., Haupt, A., & Peters, W. (2005). International environmental agreements and strategic voting. *Scandinavian Journal of Economics*, *107*(1), 175–195. [102]

Caparrós, A., Péreau, J.-C., & Tazdaït, T. (2004). North-south climate change negotiations: A sequential game with asymmetric information. *Public Choice*, *121*(3–4), 455–480. [91]

Caparrós, A. (2016). Bargaining and international environmental agreements. *Environmental and Resource Economics*, *65*(1), 5–31. [91, 102]

Carraro, C., & Siniscalco, D. (1993). Strategies for the international protection of the environment. *Journal of Public Economics*, *52*(3), 309–328. [89]

Coase, R. H. (1960). The problem of social cost. *Journal of Law and Economics*, *3*, 1–44. [90]

Dales, J. H. (1968). *Pollution, property & prices; an essay in policy-making and economics*. University of Toronto Press. [90]

Damon, M., Cole, D. H., Ostrom, E., & Sterner, T. (2019). Grandfathering: Environmental uses and impacts. *Review of Environmental Economics and Policy*, *13*(1), 23–42. [97]

Dijkstra, B. R., & Nentjes, A. (2020). Pareto-efficient solutions for shared public good provision: Nash bargaining versus exchange-matching-lindahl. *Resource and Energy Economics*, *61*, 101179. [91, 101]

Doda, B., Quemin, S., & Taschini, L. (2019). Linking permit markets multilaterally. *Journal of Environmental Economics and Management*, *98*, 102259. [91]

Doda, B., & Taschini, L. (2017). Carbon dating: When is it beneficial to link etss? *Journal of the Association of Environmental and Resource Economists*, 4(3), 701–730. [90]

Egenhofer, C. (2007). The making of the eu emissions trading scheme: Status, prospects and implications for business. *European Management Journal*, *25*(6), 453–463. [88]

Epstein, R. A. (1979). Possession as the root of title. *Georgia Law Review*, *13*, 1221–43. [97]

Flachsland, C., Marschinski, R., & Edenhofer, O. (2009). To link or not to link: Benefits and disadvantages of linking cap-and-trade systems. *Climate Policy*, *9*(4), 358–372. [90]

Freeborn, D. (2023). Efficiency and fairness trade-offs in two player gargaining games. *European Journal for Philosophy of Science*, *13*(49), 1–23. [101]

Gersbach, H., & Hummel, N. (2016). A development-compatible refunding scheme for a climate treaty. *Resource and Energy Economics*, *44*, 139–168. [92]

Goulder, L. H., & Schein, A. R. (2013). Carbon taxes versus cap and trade: A critical review. *Climate Change Economics*, *4*(3), 1350010. [89]

Graziosi, G. R. (2009). On the strategic use of representative democracy in international agreements. *Journal of Public Economic Theory*, *11*(2), 281–296. [91]

Grimm, V., & Ilieva, L. (2013). An experiment on emissions trading: The effect of different allocation mechanisms. *Journal of Regulatory Economics*, *44*(3), 308–338. [97]

Habla, W., & Winkler, R. (2018). Strategic delegation and international permit markets: Why linking may fail. *Journal of Environmental Economics and Management*, *92*, 244–250. [91]

Hahn, R. W. (1984). Market power and transferable property rights. *Quarterly Journal of Economics*, *99*(4), 753–765. [95]

Hahn, R. W., & Stavins, R. N. (2011). The effect of allowance allocations on cap-and-trade system performance. *Journal of Law and Economics*, *54*(4), S267–S294. [109]

Hahnel, R., & Sheeran, K. A. (2009). Misinterpreting the coase theorem. *Journal of Economic Issues*, *43*(1), 215–238. [105]

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*(3859), 1243–1248. [89]

Harris, J. M., & Roach, B. (2022). *Environmental and natural resource economics: A contemporary approach*. Routledge. [105]

Harstad, B. (2007). Harmonization and side payments in political cooperation. *American Economic Review*, *97*(3), 871–889. [91]

Harstad, B. (2023). Pledge-and-review bargaining: From kyoto to paris. *Economic Journal*, *133*(651), 1181–1216. [91]

Helm, C. (2003). International emissions trading with endogenous allowance choices. *Journal of Public Economics*, *87*(12), 2737–2747. [90]

IPCC. (2023). Summary for policymakers. *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 1–34. [88]

Lueck, D. (1995). The rule of first possession and the design of the law. *Journal of Law and Economics*, *38*(2), 393–436. [97]

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. Oxford University Press. [89]

McGinty, M. (2007). International environmental agreements among asymmetric nations. *Oxford Economic Papers*, *59*(1), 45–62. [92]

Montero, J.-P. (1998). Marketable pollution permits with uncertainty and transaction costs. *Resource and Energy Economics*, *20*(1), 27–50. [109]

Muthoo, A. (1999). *Bargaining theory with applications*. Cambridge University Press. [99, 100, 105]

Nash, J. F. (1950). The bargaining problem. *Econometrica*, *18*(2), 155–162. [88, 100]

Naumann, F. (2022). *Four essays in energy and environmental economics* [Doctoral Thesis]. Technische Universität Kaiserslautern. https://nbn-resolving.de/urn:nbn:de:hbz:386-kluedo-70667. [87, 110, 122]

Nehra, A., & Caplan, A. J. (2022). Nash bargaining in a general equilibrium framework: The case of a shared surface water supply. *Water Resources and Economics*, *39*, 100206. [91]

Nordhaus, W. D. (2007). To tax or not to tax: Alternative approaches to slowing global warming. *Review of Environmental Economics and Policy*, *1*(1), 26–44. [88, 89]

Nordhaus, W. D. (2015). Climate clubs: Overcoming free-riding in international climate policy. *American Economic Review*, *105*(4), 1339–1370. [89]

Nordhaus, W. D. (2019). Climate change: The ultimate challenge for economics. *American Economic Review*, *109*(6), 1991–2014. [88]

Osborne, M. J., & Rubinstein, A. (1990). *Bargaining and markets*. Academic Press. [99]

Pigou, A. (1920). *The economics of welfare*. McMillan & Co. [90]

Rose, C. M. (1985). Possession as the origin of property. *University of Chicago Law Review*, *52*(1), 73–88. [97]

Roth, A. E. (1985). *Game-theoretic models of bargaining*. Cambridge University Press. [99]

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, *50*(1), 97–109. [87, 88, 99]

Schmalensee, R., & Stavins, R. N. (2017). Lessons learned from three decades of experience with cap and trade. *Review of Environmental Economics and Policy*, *11*(1), 59–79. [88]

Segendorff, B. (1998). Delegation and threat in bargaining. *Games and Economic Behavior*, *23*(2), 266–283. [91]

Smead, R., Sandler, R. L., Forbes, P., & Basl, J. (2014). A bargaining game analysis of international climate negotiations. *Nature Climate Change*, *4*(6), 442–445. [90]

Stavins, R. N. (1995). Transaction costs and tradeable permits. *Journal of Environmental Economics and Management*, *29*(2), 133–148. [95]

Stavins, R. N. (2008a). Addressing climate change with a comprehensive us cap-and-trade system. *Oxford Review of Economic Policy*, *24*(2), 298–321. [88]

Stavins, R. N. (2008b). A meaningful u.s. cap-and-trade system to address climate change. *Harvard Environmental Law Review*, *32*(2), 293–371. [109]

Stavins, R. N. (2022). The relative merits of carbon pricing instruments: Taxes versus trading. *Review of Environmental Economics and Policy*, *16*(1), 62–82. [89]

Stern, N. H. (2008). The economics of climate change. *American Economic Review: Papers & Proceedings*, *98*(2), 1–37. [88]

Stern, N. H. (2007). *The economics of climate change: The stern review*. Cambridge University Press. [88]

Swanson, T., & Groom, B. (2012). Regulating global biodiversity: What is the oroblem? *Oxford Review of Economic Policy*, *28*(1), 114–138. [91]

Underdal, A., Hovi, J., Kallbekken, S., & Skodvin, T. (2012). Can conditional commitments break the climate change negotiations deadlock? *International Political Science Review*, *33*(4), 475–493. [88]

Urpelainen, J. (2012). Technology investment, bargaining, and international environmental agreements. *International Environmental Agreements: Politics, Law and Economics*, *12*, 145–163. [91]

Weitzman, M. L. (1974). Prices vs. quantities. *Review of Economic Studies*, *41*(4), 477–491. [89, 92]

Weitzman, M. L. (2007). A review of the stern review on the economics of climate change. *Journal of Economic Literature*, *45*(3), 703–724. [88]

Weitzman, M. L. (2014). Can negotiating a uniform carbon price help to internalize the global warming externality? *Journal of the Association of Environmental and Resource Economists*, *1*(1/2), 29–49. [88, 90, 92]

Woerdman, E., Arcuri, A., & Clò, S. (2008). Emissions trading and the polluter-pays principle: Do polluters pay under grandfathering? *Review of Law & Economics*, 4(2), 565–590. [97]

Chapter 4

Banking Competition and Capital Dependence of the Production Sector: Growth and Welfare Implications*

Joint with Paul Ritschel

Abstract

This article explores how competition among banks affects economic growth and welfare in an overlapping-generations model with rational expectations. We show that monopolistic banking can outperform its competitive counterpart in terms of both growth and welfare if the production sector is capital-intensive. In contrast, competition is beneficial if production is labor-intensive. The capital dependence of the production sector determines the relative importance of agents' saving incentives and banks' intermediation margins for capital accumulation. The theory of the second best explains the welfare-enhancing effect of a monopoly since capital accumulation is generally inefficient. We link these findings to empirical evidence.

Keywords: Financial Intermediation, Banking Competition, Overlapping Generations, Financial Dependence, Economic Growth, Risk Sharing.

JEL Classification: D53, E32, E44, G21, O41.

---

*ACKNOWLEDGMENTS. Would would like to thank Thorsten Beck, Dominique Demougin, Fabian Herweg, Maximilian Kähny, Çağıl Koçyiğit, Jan Münning, Guillermo Ordoñez, Philipp Weinschenk, and Jan Wenzelburger for valuable comments and ideas, as well as Aiswarya Nair for being an excellent research assistant. This paper has also benefited from the audiences at the 2023 Africa Meeting of the Econometric Society in Nairobi and the 2023 Asia Meetings of the Econometric Society in Beijing and Singapore, as well as research seminars at the University of Kaiserslautern-Landau and the University of Luxembourg.

4.1 Introduction

Economic growth is of central interest to economists. Although discussed controversially in the past, it is now virtually undisputed in modern macroeconomics that the financial sector of an economy also influences the development of the real sector (Levine, 1997). As banks continue to play a pivotal role in the global financial system, a sound understanding of the impact of banking on economic growth is essential. In this context, the optimal degree of competition among banks has been subject to a contentious debate over the past three decades, both in theoretical and empirical literature (Coccorese, 2017). While some contributions in economic theory advocate competition (see, e.g., Guzman, 2000; Hamada et al., 2018; Smith, 1998), others applaud the monopolistic pole of the competitive continuum for promoting growth (for example, Cetorelli, 1997; Cetorelli and Peretto, 2000, 2012).

This article contributes to the literature by explaining the ambiguous relationship between interbank competition and economic growth. We specify a simple condition under which a banking monopoly induces greater growth and higher levels of welfare than a perfectly competitive banking system and vice versa. Monopolization can be beneficial if the economy's production sector is highly capital-dependent and the dividend payments of banks sufficiently constrained. However, if production is labor-intensive, then perfect competition maximizes growth and welfare. The underlying mechanism works as follows. In our overlapping generations model, capital accumulation is endogenous and banks contribute directly to capital accumulation via their equity, a channel that complements their intermediation of agents' deposits. Competition among banks has a twofold impact on capital accumulation, as it determines agents' saving incentives on the one hand and banks' investments through their intermediation margins on the other. For labor-intensive production technologies, household wage income accounts for a large share of total income in the economy, so that capital accumulation is primarily determined by the incentives to save. Since competitive banking sets high-powered incentives, it generates greater economic growth than its monopolistic counterpart. For capital-intensive technologies, on the contrary, wage income is relatively low such that the provision of saving incentives is of minor importance. Instead, it is the opulent capital investments of a monopolistic bank that drive economic growth.

The welfare-enhancing effect of a monopoly can be explained by the theory of the second best due to Lipsey and Lancaster (1956). Since savings depend on preferences, agents will generally fail to implement efficient capital accumulation. Introducing an additional source of market failure, namely a banking monopoly, may then lead to a more desirable overall outcome since the monopolist's investments contribute to capital accumulation and improve its efficiency. By taking a dynamic perspective on banking competition, this paper is, to the best of our knowledge, the first to discover that imperfect competition can engender positive implications for economic growth and welfare that result from allocative effects only.

Our results constitute a novel theoretical foundation for an empirical puzzle concerning the interplay of banking competition and economic growth. Both Cetorelli and Gambera (2001) and Hoxha (2013) find that banking concentration promotes growth in financially dependent industries. By contrast, it has a growth-depressing effect in industries with low financial dependence. Maudos and Fernandez de Guevara (2006) show that banks' exercise of market power is positively correlated with economic growth in industrialized economies. On the other hand, Deidda and Fattouh (2005) find that the correlation between banking concentration and growth is negative solely in low-income countries. This result is complemented by those of Beck et al. (2004), who discover that banking concentration implies financing obstacles for firms in developing countries only. Clearly, these empirical insights are consistent with our theoretical results because industrialized countries tend to have capital-intensive branches of industry while low-income countries are typically dominated by labor-intensive manufacturing.

Three further valuable insights can be derived from our model. First, we show that the provision of risk sharing by financial intermediation boosts the long-run growth of an economy, independently of the degree of competition between banks. In particular, an economy with banks, even if it is endowed with significantly less initial capital, will eventually outgrow an economy without banks. This result helps explain why under-developed countries that do not have a well-functioning banking system might experience development traps. The second insight is that risk sharing, regardless of the competition among banks, has no impact on the qualitative patterns of economic growth. In particular, it cannot adversely affect the dynamics of an economy that otherwise exhibits monotonic growth only. Finally, comparing monopolistic and competitive banking, the competitive banking system turns out to be more prone to banking crises due to lower equity reserves and higher susceptibility to forecast errors. Our model thus provides support for the competition-fragility hypothesis debated in the finance literature, cf. Freixas and Rochet (2008).

Methodology. Our analysis builds on a simple overlapping-generations (OLG) model with two-period lived agents and financial intermediation as a vehicle to share risk.[1] Profit-maximizing banks offer risk-averse agents deposit contracts that

1. Our overlapping-generations framework is related to the one used in Allen and Gale (1997) and Banerji et al. (2004).

insure against idiosyncratic risk in the production sector. Following Böhm and Wenzelburger (1999), we analyze the economic dynamics using a sequential modeling approach that allows to invoke well-known results from dynamical systems theory. The basic mechanism of capital accumulation is described by an economic law and the formation of expectations is stipulated by a forecasting rule. A time-discrete dynamical system then obtains by combing the economic law with a forecasting rule. In the bulk of this article, we focus on rational expectations, although the modeling approach is more generic as it also allows for subjective (erroneous) beliefs.

Related Literature. The literature on the finance-growth nexus dates back to Schumpeter (1911), who held the view that financial services facilitate the initiation of entrepreneurial activity (King & Levine, 1993b, 1993c). However, it was not until the 1950s that the importance of finance for economic growth was explicitly noted by Gurley and Shaw (1955, p. 516), who suspected *"an inadvertent undervaluation by economists of the role that finance plays in determining the pace and pattern of growth."* The presumed correlation between finance and economic growth was then indeed confirmed in the pioneering empirical contributions by Goldsmith (1969), McKinnon (1973), and Shaw (1973).[2] Economic theory, on the other hand, attributes the promotion of growth to the fundamental functions that banks fulfill in an economy (see, for example, Bencivenga and Smith, 1991; Greenwood and Jovanovic, 1990; Greenwood and Smith, 1997; Obstfeld, 1994; Pagano, 1993). Particularly important functions are the provision of liquidity (cf. Bencivenga and Smith, 1991; Diamond and Dybvig, 1983), the informational advantage of financial intermediaries (cf. Hellwig, 1991), and risk sharing (cf. Bencivenga and Smith, 1991). Risk sharing is the pivotal function in our model.

In both empirical and theoretical economic research, it is therefore generally accepted that financial intermediation and especially banking spurs economic growth. However, there is no clear-cut consensus on the optimal degree of competition between banks. Early contributions, as for example Greenwood and Jovanovic (1990) and Bencivenga and Smith (1991), share the bottom line that market imperfections in the banking sector tend to inhibit economic growth, or, in the words of King and Levine (1993b, p. 515), *"financial sector distortions can reduce the rate of economic growth."* Market power, above all, has attracted special attention ever since. Smith (1998) attributes imperfect interbank competition a negative impact on economic growth due to increased financing costs for firms. Apart from that,

2. More recent studies that find a positive correlation between financial intermediation and economic growth are King and Levine (1993a), Levine and Zervos (1998), and Beck et al. (2000). For excellent overviews of the empirical literature, see Levine (2005) and Aziakpono (2011).

he demonstrates that monopolistic banking can increase the severity of business cycles. As pointed out above, a banking monopoly can neither amplify nor generate business cycles in our model. Complementing Smith's results, Guzman (2000) finds that monopolistic banking reduces economic growth by tightening credit rationing. Finally, Hamada et al. (2018) show that less competition adversely affects economic growth through lower deposit rates.

Contrary findings were first brought forward by Cetorelli (1997) and Cetorelli and Peretto (2000, 2012). The authors argue that reduced competition between banks can be beneficial for economic growth in environments with substantial idiosyncratic risk. The mechanism is that competition reduces the banks' incentives to engage in relationship lending, which is detrimental to economic growth. However, this mechanism relies on informational asymmetries between banks and borrowers as well as spill-over effects of financial intermediation on the performance of the productive sector. Both of these features are absent in our model because there is symmetric information and the economy's productivity is unaffected by financial intermediation.

Besides the effect on economic growth, there is a small canvas of literature that attributes monopolistic banking a positive impact on financial stability. Hellmann et al. (2000), Keeley (1990), and Allen and Gale (2000) show that fierce competition induces banks to take more risk, thereby reducing the stability of the financial system. This result is consistent with the findings in our paper, as the model shows that competitive banks are more susceptible to forecast errors than a monopolist, see Section 4.6. Nonetheless, the competition-fragility hypothesis is not undisputed among economists. Boyd and De Nicolo (2005), for instance, argue that market power leads to higher loan rates, which imply higher bankruptcy risks for borrowers.

Outline. The remainder of this paper is organized as follows. The next section lays out the basic OLG framework, describes the financial intermediation components of the model, and introduces all necessary assumptions. In Section 4.3, we analyze capital accumulation and define dynamical systems. Subsequently, we present our main findings concerning economic growth and welfare in Section 4.4. Section 4.5 examines the qualitative dynamics of the model and whether financial intermediation may generate endogenous fluctuations. Section 4.6 is a brief digression discussing the implications for financial stability. We extend the model by incorporating dividend payments in Section 4.7 before we conclude. All proofs are provided in Appendix 4.A. Throughout this article, we will compare three economies in a *ceteris-paribus* manner: an economy without financial intermediation, one with perfect competition among banks, and one reigned by a monopolist.

4.2 Basic Model

4.2.1 Prerequisites

Consider an overlapping-generations model with one sector and markets for capital, labor, and a perishable good that can be consumed and invested. Time is discrete and indexed by $t = 0, 1, \dots, \infty$. At the beginning of each period t, a new generation comprising a unit-mass continuum of homogeneous agents is born. Agents live for two periods, referred to as *young* and *old*. In period $t = 0$, there exists an initial old generation endowed with capital $K_0 > 0$. The economy accommodates financial intermediaries in the form of risk-neutral commercial banks, described in more detail below.

Preferences. Agents are risk-averse and value both youthful consumption $c^1 \geq 0$ and old-age consumption $c^2 \geq 0$. Their intertemporal preferences are represented by a life-cycle utility function $U : \mathbb{R}^2_+ \to \mathbb{R}$, defined by

$$U(c^1, c^2) \;:=\; u(c^1) + v(c^2)$$

that satisfies the following standard assumptions.

Assumption 7 (Preferences)**.**

(i) The utility functions u, v are twice continuously differentiable, strictly increasing, strictly concave, and satisfy the Inada conditions. We normalize $v(0) = 0$.

(ii) The Arrow-Pratt coefficients of relative risk aversion

$$\alpha_u \;:=\; -\frac{u''(c^1)c^1}{u'(c^1)}, \quad \text{and} \quad \alpha_v \;:=\; -\frac{v''(c^2)c^2}{v'(c^2)}$$

are constants that satisfy $0 < \alpha_u \leq \alpha_v < 1$.

Constant relative risk aversion is a widespread assumption in the finance literature as it helps to keep the analytical complexity manageable. Assumption 7 (ii) implies that agents' optimal investments are non-decreasing in the return on investment, thus ensuring that youthful and old-age consumption are gross substitutes.[3]

3. Note that a time discount factor can be incorporated in the utility function v. The CES utility function is a simple example that satisfies Assumption 7, cf. Example 13 below.

Production. The production sector of the economy is perfectly competitive. The technology of the representative firm is neoclassical and uses capital $K \geq 0$ and labor $N \geq 0$ with constant returns to scale. Capital is provided to firms by both agents and banks, paid its marginal product, and depreciates fully during production. We denote by $k := K/N$ the capital-labor ratio and by $f : \mathbb{R}_+ \to \mathbb{R}_+$ the production function, so that $y = f(k)$ is the economy's GDP per capita. For notational convenience, we define the two elasticities

$$\epsilon_1(k) \;:=\; \frac{f'(k)k}{f(k)}, \quad \text{and} \quad \epsilon_2(k) \;:=\; \frac{f''(k)k}{f'(k)}$$

and denote the marginal product of labor by

$$w(k) \;:=\; f(k) - f'(k)k.$$

The production technology satisfies the following assumptions.

Assumption 8 (Technology).

(i) The production function f is twice continuously differentiable, strictly increasing, strictly concave, and satisfies the Inada conditions.

(ii) The marginal product of labor w is strictly concave and the elasticity ϵ_2 satisfies

$$-1 \;<\; \epsilon_2(k) \;<\; 0 \quad \textit{for all} \quad k \;\geq\; 0.$$

The elasticity ϵ_1 is of central importance for our analysis as it captures the capital-dependence of the production sector. An increase in ϵ_1 implies that capital becomes more important relative to labor as a production factor. From an output perspective, ϵ_1 is the share of capital income to total income in the economy. Thus, for a larger ϵ_1, the production sector becomes more capital-intensive and less labor-intensive.[4] Moreover, it should be stressed that while Assumption 8 (i) is standard in the literature on OLG models, Assumption 8 (ii) is also satisfied by standard production functions, e.g., the Cobb-Douglas and the CES production function. It implies that capital and labor income are well behaved.

The young generation constitutes the workforce of the economy, whereas the old generation is retired. We normalize that each young agent supplies one unit of labor inelastically to a perfectly competitive labor market, implying that the market-clearing wage rate w_t in period t is determined by the marginal product of labor, $w_t = w(k_t)$.

4. Note that $0 < \epsilon_1 < 1$ by the strict concavity of the production function f.

Example 12 (Cobb-Douglas Technology).
For the Cobb-Douglas production function $f(k) = Ak^{\alpha}$, where $A > 0$ and $0 < \alpha < 1$, the elasticities $\epsilon_1 = \alpha$ and $\epsilon_2 = \alpha - 1$ are constants.

Projects & Expectations. Agents and banks can invest in projects (firms) in the production sector. Projects have a stochastic binary outcome, i.e., they can either be successful or fail. The likelihood of a successful project is exogenously given by a success rate $p \in (0,1)$, which is common knowledge . The uncertainty about the outcome of a project resolves one period after capital has been invested. If a project turns out to be successful, then a gross rate of return $\rho > 0$ realizes. The gross return on a failed project is zero. In each period t, agents form an expectation $\rho_t^e > 0$ with respect to the return on a successful project ρ_{t+1} realized in $t+1$. Expectations are assumed to be homogeneous and, unless stated otherwise, rational. Nonetheless, the modeling approach is more generic as it allows for subjective (erroneous) beliefs of agents, cf. Section 4.6.

4.2.2 Idiosyncratic Investments

In the absence of banks, a young agent may transfer resources to the second period of his life solely by investing part of his wage income in a risky production project, which exposes him to the idiosyncratic risk of an old-age income shock.[5] The decision problem of a young agent in this case is as follows. The agent's objective is to maximize his expected utility of lifetime consumption. Given the wage income w_t and the anticipated return ρ_t^e, the optimal idiosyncratic investment I_t solves

$$I_t \;=\; I(w_t, \rho_t^e) \;:=\; \underset{0 \le I \le w_t}{\operatorname{argmax}}\; u(w_t - I) + p\,v(\rho_t^e I) + (1-p)\,v(0). \tag{4.1}$$

By Assumption 7 (i), the optimal solution satisfies $0 < I_t < w_t$ and it is uniquely determined by the first-order condition

$$\frac{u'(w_t - I_t)}{v'(\rho_t^e I_t)} \;=\; p\rho_t^e, \tag{4.2}$$

which states that the marginal rate of intertemporal substitution equals the expected return on investment. The expected utility of investing with idiosyncratic risk is denoted by

$$U_{\text{res}}(w_t, \rho_t^e) \;:=\; u(w_t - I(w_t, \rho_t^e)) + p\,v(\rho_t^e I(w_t, \rho_t^e)) \tag{4.3}$$

5. Recall that there is no pure storage technology in this model, as is the case, for example, in Diamond and Dybvig (1983).

and, as will become clear shortly, constitutes the agent's reservation utility in the presence of banks. Figure 4.1 illustrates the timing associated with an idiosyncratic investment.

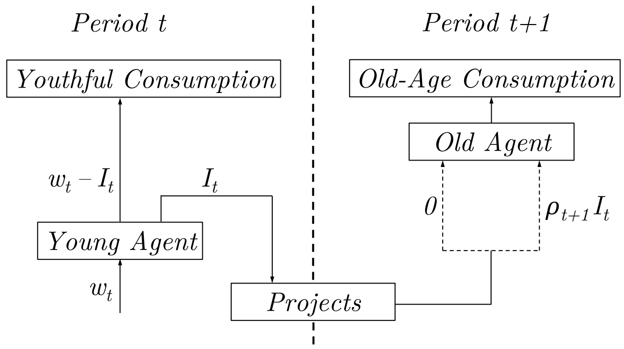


Figure 4.1. Idiosyncratic investment in a production project.

The case of agents endowed with CES preferences is our primary example throughout the article.

Example 13 (CES Preferences).
For the CES utility functions $u(c^1) = \frac{1}{\sigma}(c^1)^\sigma$ *and* $v(c^2) = \frac{\beta}{\sigma}(c^2)^\sigma$, *where* $\beta > 0$ *and* $0 < \sigma < 1$, *the optimal idiosyncratic investment (4.1) is*

$$I_t = I(w_t, \rho_t^e) = \left(1 + \left[\beta\, p\, (\rho_t^e)^\sigma\right]^{\frac{1}{\sigma-1}}\right)^{-1} w_t$$

such that the agent's reservation utility (4.3) becomes

$$U_{\text{res}}(w_t, \rho_t^e) = I(w_t, \rho_t^e)^{\sigma-1} \beta\, p\, (\rho_t^e)^\sigma \sigma^{-1} w_t.$$

4.2.3 Financial Intermediation

From an institutional perspective, a bank can be seen as a coalition of (old) agents that arises endogenously as a vehicle for risk sharing and rent extraction (see, e.g., Freixas and Rochet, 2008). A bank's objective is to maximize the expected profit generated by taking deposits from the public and investing them into the production projects.[6] To finance its investments, the bank offers young agents a gross rate

6. Investing in production projects may be considered the lending business of the banks. By the law of large numbers, the loan default rate is then $1 - p$.

$r_t > 0$ on deposits $D_t \geq 0$, which are at the agents' discretion.[7] Given that the bank has offered a deposit rate r_t on savings in period t, a young agent's optimal supply of deposits solves

$$D_t \;=\; D(w_t, r_t) \;:=\; \operatorname*{argmax}_{0 \leq D \leq w_t} \; u(w_t - D) + v(r_t D). \tag{4.4}$$

It follows from Assumption 7 (i) that the solution $0 < D_t < w_t$ is uniquely determined by the first-order condition[8]

$$\frac{u'(w_t - D_t)}{v'(r_t D_t)} \;=\; r_t. \tag{4.5}$$

However, since investing with idiosyncratic risk constitutes the agent's outside option, the utility of saving must be weakly larger than the reservation utility (4.3). Formally, an agent accepts the deposit contract if and only if r_t satisfies the *participation constraint*

$$u(w_t - D(w_t, r_t)) + v(r_t D(w_t, r_t)) \;\geq\; U_{\text{res}}(w_t, \rho_t^e). \tag{PC}$$

It remains to formalize the objective function of a bank. Exploiting the law of large numbers, the bank correctly anticipates that, on aggregate, the share p of all projects will be successful. Therefore, the *anticipated* intermediation margin per unit of deposits amounts to $p\rho_t^e - r_t$. With rational expectations, the intermediation margin $p\rho_{t+1} - r_t$ *realized* in $t+1$ is non-negative at all times, implying that banks realize non-negative profits. These profits are either retained and add to a bank's equity or released as dividend payments to agents instead. Note that if profits are retained, then the bank optimally reinvests them into the production projects since there is no storage technology (which is a standard assumption in the OLG literature). We denote a bank's equity in period t by $e_t \in \mathbb{R}$ and normalize

7. A short remark the design of our deposit contracts is in order. First, note that since the deposit contract is of the simple structure r_t, the contractual repayment $D_t \mapsto r_t D_t$ is linear. We thereby abstract from stochastic contracts, contracts specifying fixed fees, and from those stipulating both the deposit rate and the deposits. Second, note that young agents either save in the form of deposits at the bank or invest with idiosyncratic risk. Simultaneous saving and investing will neither occur in the case of perfect interbank competition nor in the monopoly case. With perfect competition, agents have no incentive to invest at all because the contract gives agents the entire available surplus. A monopolistic bank, on the other hand, will naturally exploit her market power and prohibit idiosyncratic investments in order to maximize the rent extracted.

8. The second-order condition is satisfied by the strict concavity of u and v.

$e_0 = 0$.[9] It should be noted that since the old agents own the bank, they would, in theory, release and consume the entire banking surplus. However, in a banking context, equity is subject to strict regulatory requirements (e.g., capital-adequacy ratios) that cap the amount of dividends that can be paid out.[10] To simplify the exposition and keep the complexity manageable, we will first consider a stylized environment in which the surplus is fully retained. Subsequently, in Section 4.7, we argue that our central findings are robust even if banks, subject to regulation, distribute part of the surplus to the owners. Finally, since deposit management costs are irrelevant to the point of this paper, we assume them to be negligible. Summarizing these considerations, the profit expected by a bank in period t, given w_t, ρ_t^e, and e_t, is then formally defined by

$$\pi_t^e(r_t) \; := \; D(w_t, r_t)(p\rho_t^e - r_t) + p\rho_t^e e_t.$$

Figure 4.2 illustrates the timing in the case of financial intermediation.

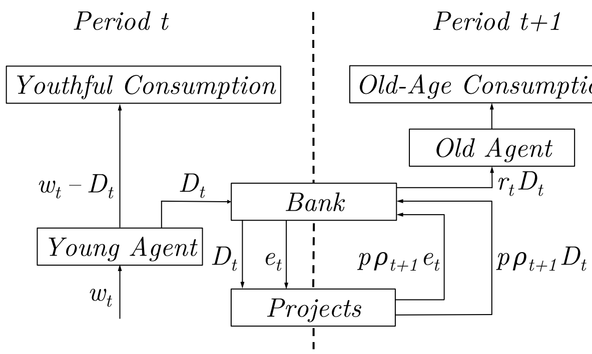


Figure 4.2. Investment in a project through financial intermediation.

Remark 3 (Notation).
Henceforth, the superscript $i = \mathrm{C}$ *will indicate the case of perfect competition between banks,* $i = \mathrm{M}$ *the case of a banking monopoly, and* $i = \mathrm{N}$ *the case without financial intermediation.*

9. Observe that the model allows for indebtedness of banks since e_t can take negative values. This case may obtain if the realized intermediation margins are negative owing to erroneous beliefs.

10. See, for instance, Gersbach and Wenzelburger (2003, 2008, 2012), who investigate the need for and the effectiveness of banking regulation in two-period lived OLG models.

4.2.4 Deposit Contracts

If there is perfect competition between banks, the anticipated intermediation margin is zero and agents are awarded the entire expected surplus. In this case, deposit contracts are actuarially fair and provide complete insurance without a premium for the bank. The deposit contract thereby offers a better risk-return characteristic than an idiosyncratic investment. We summarize these results in Proposition 13.

Proposition 13 (Perfect Competition).
The deposit rate in case of perfect competition between banks is

$$r_t^{\mathrm{C}} = r^{\mathrm{C}}(w_t, \rho_t^e) := p\rho_t^e.$$

The opposite pole of the competitive continuum is considered next. Given w_t, ρ_t^e, and e_t, the problem of a monopolistic bank consists in offering a deposit rate that solves

$$\max_{0 \le r \le p\rho_t^e} \pi_t^e(r) \quad \text{s.t.} \quad u(w_t - D(w_t, r)) + v(rD(w_t, r)) \ge U_{\mathrm{res}}(w_t, \rho_t^e). \tag{4.6}$$

The solution r_t^{M} to Problem (4.6) is either determined by the first-order condition

$$\frac{p\rho_t^e - r_t^{\mathrm{M}}}{r_t^{\mathrm{M}}} = \frac{1}{\eta\left(w_t, r_t^{\mathrm{M}}\right)} \tag{4.7}$$

corresponding to the relaxed problem without the participation constraint, where

$$\eta(w, r) := \frac{\frac{\partial D}{\partial r}(w, r)\, r}{D(w, r)} > 0, \tag{4.8}$$

or it is determined by the binding participation constraint

$$u(w_t - D(w_t, r_t^{\mathrm{M}})) + v(r_t^{\mathrm{M}} D(w_t, r_t^{\mathrm{M}})) = U_{\mathrm{res}}(w_t, \rho_t^e). \tag{4.9}$$

Recognize that (4.7) is the standard optimality condition for a monopoly, stating that the Lerner-index equals the inverse elasticity.[11] Intuitively, if (4.7) specifies

11. Since $1/\eta$ is a measure of the monopolist's market power, the competitive limit $r_t^{\mathrm{M}} = p\rho_t^e$ obtains if $\eta \to \infty$, cf. Proposition 13. Moreover, it can readily be shown that our model also covers oligopolistic banking. The sole difference is that η in (4.7) is multiplied by the number of competing banks. For a formal treatment of a related problem, we refer to Freixas and Rochet (2008).

a sufficiently high deposit rate that is accepted by agents, then it is optimal for the monopolist to implement this contract. Otherwise, the participation constraint (PC) is binding and the optimal contract is determined by (4.9). In the latter case, the solution to (4.7) is not implementable because agents would reject the bank's offer and invest with idiosyncratic risk instead. Indeed, the solution to (4.9) is the lowest deposit rate that still ensures an agent's participation. These considerations lead to the following proposition.

Proposition 14 (Banking Monopoly).
For any given $w_t > 0$ and $\rho_t^e > 0$, there exists a unique deposit rate $0 < r_t^{\mathrm{M}} < p\rho_t^e$ that solves the monopolist's decision problem (4.6). It is given by

$$r_t^{\mathrm{M}} = r^{\mathrm{M}}(w_t, \rho_t^e) := \max\left\{s(w_t, \rho_t^e), b(w_t, \rho_t^e)\right\},$$

where $s(w_t, \rho_t^e)$ and $b(w_t, \rho_t^e)$ are defined by (4.7) and (4.9), respectively.[12]

Proposition 14 shows that market power reduces the deposit rates offered, $r_t^{\mathrm{M}} < r_t^{\mathrm{C}}$. It is noteworthy, however, that if r_t^{M} is determined by (4.7), then the utility attained by the deposit contract exceeds the agent's reservation utility. In this case, the monopolist has to pay the depositors a rent to receive the desired amount of funds. Lemma 12 demonstrates that this case occurs if the production projects are sufficiently risky.

Lemma 12. *For any given $w_t > 0$ and $\rho_t^e > 0$, there exist thresholds $0 < \underline{p} \leq \overline{p} < 1$ for the success rate p such that $r_t^{\mathrm{M}} = s(w_t, \rho_t^e)$ if $p < \underline{p}$, whereas $r_t^{\mathrm{M}} = b(w_t, \rho_t^e)$ if $p \geq \overline{p}$.*

The intuition of Lemma 12 is straightforward. If an idiosyncratic investment is highly risky, then the monopolist can exploit agents' risk aversion and implement the profit-maximizing deposit rate $r_t^{\mathrm{M}} = s(w_t, \rho_t^e)$ determined by the relaxed problem. However, if the projects are rather safe, then this contract is not accepted by agents.[13]

12. For technical reasons, we claim that $r^{\mathrm{M}}(w_t, \rho_t^e) = b(w_t, \rho_t^e)$ whenever $b(w_t, \rho_t^e) = s(w_t, \rho_t^e)$.
13. Technically, it can be shown that the reservation utility U_{res} increases with p, so that the participation constraint becomes more demanding.

Example 14 (CES Preferences).
For the CES utility functions presented in Example 13, the deposit supply (4.4) is

$$D_t = D(w_t, r_t) = \left(1 + \left[\beta\,(r_t)^{\sigma}\right]^{\frac{1}{\sigma-1}}\right)^{-1} w_t.$$

For instance, let $\sigma = 0.5$ and $\beta = 1$, then

$$r_t^{\mathrm{M}} = r^{\mathrm{M}}(w_t, \rho_t^e) = \max\left\{p^2\rho_t^e, \sqrt{p\rho_t^e + 1} - 1\right\}.$$

4.3 Capital Accumulation

The aggregate capital stock of the economy is determined by the total capital endowment of successful projects. Without banks, the projects are financed by idiosyncratic investments of agents only. In the presence of banks, however, such investments do not occur since, in equilibrium, agents will always accept the deposit contract. Instead, capital accumulation is fed from deposits of private households (which are channeled into the production sector by the financial intermediaries) and investments from banks. The degree of competition among banks has a twofold impact on these two funding sources. While, on the one hand, a higher deposit rate implies higher-powered savings incentives for agents, fierce competition reduces the intermediation margins of banks on the other hand. The former mechanism is derived in the following corollary.

Corollary 5 (Capital Supply).
For any given $w_t > 0$ and $\rho_t^e > 0$, it holds that

$$I(w_t, \rho_t^e) \le D(w_t, r^{\mathrm{M}}(w_t, \rho_t^e)) < D(w_t, r^{\mathrm{C}}(w_t, \rho_t^e)),$$

where the first inequality holds strict if and only if $r^{\mathrm{M}}(w_t, \rho_t^e) = s(w_t, \rho_t^e)$.

The intuition of Corollary 5 is that the provision of risk sharing incentivizes agents to release additional funds into the production sector. As we will see later, it is exactly this feature of banks that is responsible for promoting long-run economic growth in our model.

Remark 4 (Expectations).
In the following, it is convenient to express the expectations in terms of the capital-labor ratio by letting $k_t^e = f'^{-1}(\rho_t^e)$ denote the expectation formed in period t with respect to the k_{t+1} realized in $t+1$.

In order to obtain a dynamical system that describes the evolution of the economy over time, we next define an economic law of capital accumulation and then add a forecasting rule that stipulates how expectations are formed. By the law of large numbers, the share p of the projects is successful, while the rest of them fails. Therefore, in the case without financial intermediation, the *economic law of capital accumulation* is given by the map

$$k_{t+1} = G^{\mathrm{N}}(k_t, k_t^e) := pI(w(k_t), f'(k_t^e)). \tag{4.10}$$

In the presence of banks, capital accumulation is driven by

$$k_{t+1} = p\big(D_t + e_t\big). \tag{4.11}$$

The bank's proceeds from the successful projects (the capital income) in period t are $f'(k_t)k_t$, while it has to pay depositors the contractual amount $r^i_{t-1}D_{t-1}$. Hence, a bank's equity e_t in period t is determined by the map[14]

$$e^i(k_t, k_{t-1}, k^e_{t-1}) := f'(k_t)k_t - r^i(w(k_{t-1}), f'(k^e_{t-1}))\, D(w(k_{t-1}), r^i(w(k_{t-1}), f'(k^e_{t-1}))), \tag{4.12}$$

where $i \in \{\mathrm{C}, \mathrm{M}\}$ depending on the degree of competition. From (4.11) and (4.12), we can infer the economic law of capital accumulation

$$k_{t+1} = G^i(k_t, k_{t-1}, k^e_t, k^e_{t-1}) := p\left(D(w(k_t), r^i(w(k_t), f'(k^e_t))) + e^i(k_t, k_{t-1}, k^e_{t-1})\right).$$

Following Grandmont (1985), expectations are formed on the basis of a forecasting rule ψ^i. Agents have *perfect foresight* if expectations are rational at all times, i.e., if $k^e_t = k_{t+1}$ for all $t \geq 0$. *Perfect* forecasting rules in the sense of Böhm and Wenzelburger (1999) are forecasting rules that generate perfect foresight along *all* possible growth paths of the economy. Formally, these are defined as follows.[15]

14. To be precise, we define $e^i(k_0, k_{-1}, k^e_{-1}) := e_0 = 0$. Also, note that for any given $(k_t, k_{t-1}, k^e_{t-1})$, a bank's balance sheet satisfies

$$f(k_t) = w(k_t) + e^i(k_t, k_{t-1}, k^e_{t-1}) + r^i(w(k_{t-1}), f'(k^e_{t-1}))\, D(w(k_{t-1}), r^i(w(k_{t-1}), f'(k^e_{t-1}))),$$

such that output is split into wage payments to the young generation, contractual payments to the old generation, and bank equity.

15. Observe from Definition 9 that the functional form of a perfect forecasting rule is determined by the economic law and thus depends on whether financial intermediation takes place.

Definition 9 (Perfect Forecasting Rules).

(i) *A forecasting rule $\psi^{\mathrm{N}} : \mathbb{R}_+ \to \mathbb{R}_+$, defined by $k_t^e = \psi^{\mathrm{N}}(k_t)$, is called a perfect forecasting rule for the economic law G^{N} if it obeys*

$$\psi^{\mathrm{N}}(x) \;=\; G^{\mathrm{N}}(x, \psi^{\mathrm{N}}(x)) \quad \text{for all } x \in \mathbb{R}_+.$$

(ii) *Let $i \in \{\mathrm{C}, \mathrm{M}\}$. A forecasting rule $\psi^i : \mathscr{D}^i \to \mathbb{R}_+$, defined by $k_t^e = \psi^i(k_t, k_{t-1})$, is called a perfect forecasting rule for the economic law G^i if it satisfies*

$$\psi^i(x,y) \;= G^i(x,y,\psi^i(x,y),x)$$
$$\text{for all } (x,y) \in \mathscr{D}^i := \Big\{(x,y) \in \mathbb{R}_+^2 \mid e^i(x,y,x) \geq 0\Big\}.$$

Existence and uniqueness of a perfect forecasting rule is established in Lemma 13.

Lemma 13. *There exist uniquely determined perfect forecasting rules ψ^{i}, $i \in \{\mathrm{N}, \mathrm{C}, \mathrm{M}\}$, in the sense of Definition 9. In particular, $\mathscr{D}^i$ is forward-invariant: for all $(k_t, k_{t-1}) \in \mathscr{D}^i$, it holds that $\big(\psi^i(k_t, k_{t-1}), k_t\big) \in \mathscr{D}^i$.*

The lemma implies that the possibility of perfect foresight is not lost over time and, in particular, that banks never go bankrupt. Technically, the lemma shows that the perfect-foresight dynamics obtained by substituting expectations with realizations is well defined.

The perfect-foresight dynamics without banks is governed by the implicit difference equation

$$k_{t+1} \;=\; G^{\mathrm{N}}(k_t, k_{t+1}) \;=\; pI(w(k_t), f'(k_{t+1})), \quad t \geq 0, \tag{PFD$^{\mathrm{N}}$}$$

while the dynamics in the presence of banks is driven by

$$k_{t+1} \;=\; G^i(k_t, k_{t-1}, k_{t+1}, k_t), \quad i \in \{\mathrm{C}, \mathrm{M}\}, \quad t \geq 1. \tag{4.13}$$

The dynamical system (4.13) deserves further attention. In a perfectly competitive environment, the realized intermediation margin is zero at all times. Accordingly, banks' equity along any growth path $\{k_t\}_{t=0}^{\infty}$ with perfect foresight is zero and (4.13) simplifies to

$$k_{t+1} \;=\; G^{\mathrm{C}}(k_t, k_{t+1}) \;=\; pD(w(k_t), pf'(k_{t+1})). \tag{PFD$^{\mathrm{C}}$}$$

In the monopolistic case, on the contrary, the realized intermediation margins are positive, implying that the perfect-foresight dynamics is governed by the second-order difference equation

$$G^{\mathrm{M}}(k_t, k_{t-1}, k_{t+1}, k_t) = p\Big(D(w(k_t), r^{\mathrm{M}}(w(k_t), f'(k_{t+1}))) + e^{\mathrm{M}}(k_t, k_{t-1}, k_t)\Big). \quad (\mathrm{PFD}^{\mathrm{M}})$$

In view of the lagged variable in ($\mathrm{PFD}^{\mathrm{M}}$), we will next set up a linearization and evoke the Hartman–Grobman Theorem for further analyses, cf. Grobman (1959) and Hartman (1960). To do so, we define

$$\Psi : \mathscr{D}^{\mathrm{M}} \to \mathscr{D}^{\mathrm{M}}, \quad \Psi : \begin{pmatrix} k_t \\ k_{t-1} \end{pmatrix} \mapsto \begin{pmatrix} \psi^{\mathrm{M}}(k_t, k_{t-1}) \\ k_t \end{pmatrix},$$

as well as the vector

$$\boldsymbol{k}_t := \begin{pmatrix} k_t \\ k_{t-1} \end{pmatrix}.$$

This technique enables us to transform the second-order difference equation ($\mathrm{PFD}^{\mathrm{M}}$) into the dynamical system

$$\boldsymbol{k}_{t+1} = \Psi(\boldsymbol{k}_t), \quad \boldsymbol{k}_t \in \mathscr{D}^{\mathrm{M}}, \quad t \geq 1.$$

Steady states $\boldsymbol{k}^{\mathrm{M}} = (k^{\mathrm{M}}, k^{\mathrm{M}})$ of Ψ are defined by $\boldsymbol{k}^{\mathrm{M}} = \Psi(\boldsymbol{k}^{\mathrm{M}})$.[16] Linearizing the map Ψ at $\boldsymbol{k}^{\mathrm{M}}$ yields

$$\boldsymbol{k}_{t+1} = D_{\Psi}(\boldsymbol{k}^{\mathrm{M}})\, \boldsymbol{k}_t, \quad t \geq 1, \tag{4.14}$$

where

$$D_{\Psi}(\boldsymbol{k}^{\mathrm{M}}) = \begin{pmatrix} \frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}^{\mathrm{M}}) & \frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k}^{\mathrm{M}}) \\ 1 & 0 \end{pmatrix} \tag{4.15}$$

is the Jacobian of Ψ. The stability properties of $\boldsymbol{k}^{\mathrm{M}}$ determine the qualitative dynamics of the economy. The stability properties can be deduced from the Eigenvalues of the Jacobian (4.15), which compute

$$\lambda_{1,2}(\boldsymbol{k}^{\mathrm{M}}) = \frac{1}{2}\frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}^{\mathrm{M}}) \pm \sqrt{\frac{1}{4}\left(\frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}^{\mathrm{M}})\right)^2 + \frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k}^{\mathrm{M}})}. \tag{4.16}$$

We complete this section with a technical lemma that will facilitate establishing our central results.

Lemma 14. *For any given steady state $\boldsymbol{k}^{\mathrm{M}} \in \mathscr{D}^{\mathrm{M}}$, the Eigenvalues $\lambda_{1,2}$ in (4.16) satisfy* $0 < \left|\lambda_2(\boldsymbol{k}^{\mathrm{M}})\right| \leq \left|\lambda_1(\boldsymbol{k}^{\mathrm{M}})\right|$.

16. The existence of a hyperbolic steady state $\boldsymbol{k}^{\mathrm{M}}$ is established in Section 4.4.

4.4 Economic Growth and Welfare

Equipped with dynamical systems, we can now investigate how financial intermediation and in particular the degree of interbank competition affect economic growth and welfare. The long-run development of the economy is determined by (perfect-foresight) steady states $k^i \geq 0$ of the dynamical systems (PFD^i) and their stability properties. Steady states are defined by solutions to the fixed-point condition

$$\frac{k}{p} = \begin{cases} I(w(k), f'(k)) & \text{for } i = \text{N} \\ D(w(k), pf'(k)) & \text{for } i = \text{C} \\ D(w(k), r^{\text{M}}(w(k), f'(k))) + e^{\text{M}}(k, k, k) & \text{for } i = \text{M}. \end{cases} \tag{4.17}$$

As a first step, we analyze whether poverty traps may occur.[17]

Proposition 15 (Poverty Traps)**.**
Let $i \in \{\text{N}, \text{C}, \text{M}\}$. The origin $k^i = 0$ is a steady state if and only if $f(0) = 0$. If $k^i = 0$ is a steady state, then it is unstable.

Proposition 15 rules out poverty traps, regardless of the banking industry. Under perfect foresight, the economy never goes bankrupt since no growth path $\{k_t\}_{t=0}^{\infty}$ with initial capital $k_0 > 0$ converges to zero. Against this background, the question regarding positive steady states arises naturally. Theorem 5 states the first central result of this paper, namely the effects of financial intermediation on long-run economic growth.

Theorem 5 (Economic Growth)**.**
Let $k_0 > 0$ and consider economies that are identical in all aspects except their banking industry. Then the following holds.

(i) Each dynamical system (PFD^i), $i \in \{\text{N}, \text{C}, \text{M}\}$, attains a uniquely determined positive steady state $k^i > 0$, which is asymptotically stable globally on $\mathbb{R}_{++}$. These steady states satisfy $0 < k^{\text{N}} < k^{\text{C}}, k^{\text{M}}$.

(ii) If the elasticity ϵ_1 of the production function is sufficiently large, then $k^{\text{N}} < k^{\text{C}} < k^{\text{M}}$. Conversely, if ϵ_1 is sufficiently small, then $k^{\text{N}} < k^{\text{M}} < k^{\text{C}}$.

From an economic perspective, Theorem 5 (i) implies that economies featuring banks will experience a greater level of long-run growth than economies without.

17. The term “poverty trap” is used ambiguously in the literature. We define a poverty trap as the origin $k^i = 0$ being an asymptotically stable steady state.

This result holds irrespective of the degree of interbank competition and independently of the initial capital stock k_0. An economy with banks that is initially poor will in the long run outgrow any prosperous economy without financial intermediation. The underlying mechanism is that the provision of risk sharing incentivizes agents to allocate additional funds to the production sector, which fosters capital accumulation and economic growth, cf. Corollary 5. This insight contributes to explaining why low-income countries lacking a well-functioning banking system might face developmental obstacles.

The second part of Theorem 5 concerns the interplay between the competition among banks and the capital dependency of the production sector. Theorem 5 (ii) demonstrates that for a capital-intensive production technology, a banking monopoly induces the greatest level of long-run growth. By contrast, competition is favorable if the technology is labor-intensive. The mechanism works as follows. As noted at the outset of Section 4.3, competition between banks has a twofold impact on capital accumulation. While an increase in competition encourages private savings, it also reduces banks' equity owing to smaller intermediation margins. Which effect is decisive depends on the production technology: for a labor-intensive manufacturing sector, households receive opulent wage income, whereas the capital income of banks is relatively low. In such environments, the provision of strong saving incentives for households is crucial, which is why perfect competition among banks leads to the highest level of economic growth. If the production sector is capital-intensive, however, then saving incentives are of minor importance as agents' wage income is comparatively low. Due to the plentiful capital income, a monopolistic bank is then best capable of providing the production sector with sufficient capital and, consequently, is favorable for economic growth.

However, the pure growth perspective pursued so far abstracts from all normative considerations. Indeed, it is natural to question whether a monopoly can also improve welfare. Taking a utilitarian perspective, we measure agents' welfare using the expected lifetime utility.[18] Formally, given k_t and k_t^e, welfare in period t is captured by

$$\mathscr{W}^i(k_t, k_t^e) := \begin{cases} U_{\mathrm{res}}(w(k_t), f'(k_t^e)) & \text{for } i = \mathrm{N} \\ u\big(w(k_t) - D(w(k_t), r_t^i)\big) + v\big(r_t^i D(w(k_t), r_t^i)\big) & \text{for } i = \mathrm{C}, \mathrm{M}, \end{cases}$$

18. Note that the agent's expected utility is equivalent to the level of utility attained by his generation on average. Moreover, for a discussion of our welfare measure against the background of the literature on OLG models, we refer to Appendix 4.A.3.

where $r_t^i = r^i(w(k_t), f'(k_t^e))$. From this definition, we can directly infer the following corollary. It states the welfare effects of financial intermediation in a comparative statics manner.

Corollary 6 (Welfare Effects).
For any given $k_t > 0$ and $k_t^e > 0$, it holds that

$$\mathscr{W}^{\mathrm{N}}(k_t, k_t^e) \;\leq\; \mathscr{W}^{\mathrm{M}}(k_t, k_t^e) \;<\; \mathscr{W}^{\mathrm{C}}(k_t, k_t^e),$$

where the first inequality holds strict if and only if $r^{\mathrm{M}}(w(k_t), f'(k_t^e)) = s(w(k_t), f'(k_t^e))$.

Corollary 6 is a risk sharing result. Banks increase agents' welfare by insuring them against the idiosyncratic risk in the production sector.

In order to take a dynamic perspective on welfare, however, it is necessary to establish a benchmark first. Independently of financial intermediation, the realized consumption levels in period t satisfy

$$c_t^1 + c_t^2 \;=\; f(k_t) - \frac{k_{t+1}}{p}. \tag{4.18}$$

From Equation (4.18), it follows that *stationary feasible allocations* $(\bar{k}, \bar{c}^1, \bar{c}^2)$ of this OLG economy must fulfill

$$\bar{c}^1 \;\geq\; 0, \quad \bar{c}^2 \;\geq\; 0, \quad \bar{k} \;\geq\; 0, \quad \text{and} \quad \bar{c}^1 + \bar{c}^2 \;=\; f(\bar{k}) - \frac{\bar{k}}{p} \;=:\; \phi(\bar{k}). \tag{4.19}$$

The notion of a stationary feasible allocation now enables us to derive the following social-planner result, which is due to Diamond (1965).

Lemma 15. *The stationary feasible allocation $(\bar{k}, \bar{c}^1, \bar{c}^2)$ with the highest possible level of welfare is the golden-rule allocation (k_G, c_G^1, c_G^2), where $k_G = f'^{-1}(1/p)$ and the consumption plan (c_G^1, c_G^2) is uniquely determined by*

$$\frac{u'(c_G^1)}{v'(c_G^2)} \;=\; 1, \quad \textit{and} \quad c_G^1 + c_G^2 \;=\; \phi(k_G).$$

Since steady states of an OLG economy define stationary feasible allocations, we may evaluate them using (k_G, c_G^1, c_G^2) as an efficient benchmark. It is important to note that the golden-rule capital-labor ratio k_G is solely determined by the production technology and the success rate. In particular, it is not affected by financial intermediation or agents' preferences. The steady states k^i, however, depend on preferences. For this reason, k_G will generally not be attained as a steady state of the economy and the stationary feasible allocations corresponding to the steady states will deviate from the golden-rule allocation. Indeed, k_G is a steady state if and only if it solves (4.17). For this reason, capital accumulation will generally be inefficient and the steady states will not be welfare-maximizing.

The preceding considerations reveal that from a societal perspective, agents will generally fail to achieve investment efficiency. However, since the economy's steady state is also affected by financial intermediation, the question arises whether banking can enhance the efficiency of the long-run outcome. To this end, note that the level of welfare which will be attained in the long run is $\mathscr{W}^i(k^i, k^i)$, where $k^i > 0$ is the asymptotically stable steady state identified in Theorem 5. We are now in a position to state our second key result, namely how the interplay between banking competition and the production sector affects welfare.

Theorem 6 (Welfare)**.**
Let the unique positive steady states $k^i > 0$ of the dynamical systems (PFD^i) be given. Then the following holds.

(i) If $0 < k^{\mathrm{C}} \leq k_G$, then the economy with a competitive banking industry experiences a higher level of steady-state welfare than the economy without banks,

$$\mathscr{W}^{\mathrm{N}}(k^{\mathrm{N}}, k^{\mathrm{N}}) \;<\; \mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}}, k^{\mathrm{C}}).$$

(ii) If the elasticity ϵ_1 is sufficiently large, then the economy with a monopolistic banking industry experiences the highest level of steady-state welfare,

$$\mathscr{W}^{\mathrm{N}}(k^{\mathrm{N}}, k^{\mathrm{N}}) \;<\; \mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}}, k^{\mathrm{C}}) \;<\; \mathscr{W}^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}).$$

Conversely, if ϵ_1 is sufficiently small, then

$$\mathscr{W}^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}) \;<\; \mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}}, k^{\mathrm{C}}).$$

Economically, Theorem 6 (i) implies that introducing financial intermediation to an economy will enhance steady-state welfare if it alleviates under-accumulation of capital. Technically speaking, by providing investment incentives for agents and propelling capital accumulation, financial intermediation moves the steady-state allocation of the economy closer towards the golden-rule allocation. For over-accumulation, the welfare effects of financial intermediation become generally ambiguous. While the provision of risk sharing is always beneficial for welfare, the resulting investment incentives exacerbate over-accumulation. Thus, if financial intermediation induces over-accumulation, then a trade-off between insurance and reduced consumption arises.[19] The resolution of this trade-off and whether financial intermediation improves steady-state welfare depends on the production technology and agents' risk preferences.

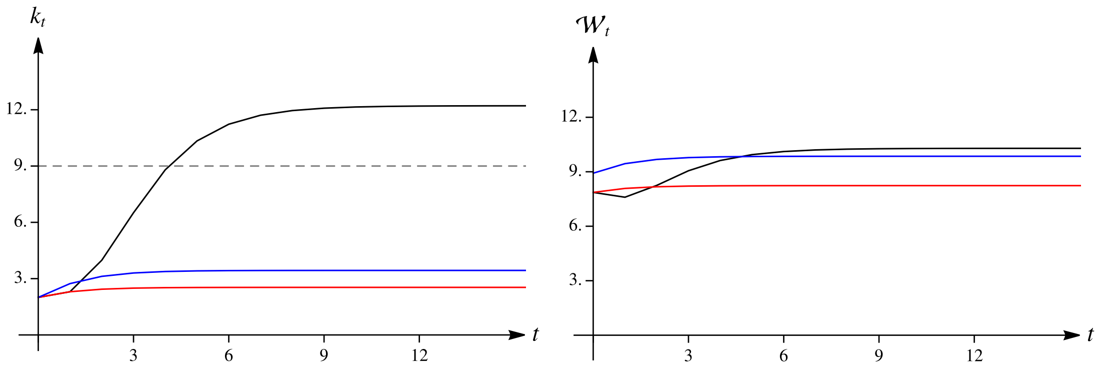
The second part of Theorem 6 reveals that the capital dependence of the production sector determines how competition among banks impacts the economy's investment efficiency. If the manufacturing sector is capital-intensive, then savings of private households are insufficient to implement efficient capital accumulation. A monopolistic bank's capital investments then improve efficiency by shifting the accumulation path towards a higher level of long-run output, consumption, and welfare. This effect can be explained by the *theory of the second best*, cf. Lipsey and Lancaster (1956). Since each cohort of agents is only concerned with maximizing their own (short-run) utility, capital accumulation is generally inefficient. An additional source of market failure, namely a banking monopoly, can then actually *improve* overall efficiency, provided that the production sector is sufficiently capital-intensive. In a labor-intensive production environment, however, private savings of agents are sufficient and the adverse effect of monopolistic rent extraction on agents' welfare outweighs. Overall, competition is then favorable for welfare.

At this point, it is worthwhile considering a standard parameterization from the literature in order to illustrate our central findings thus far.

Example 15 (CES Preferences and Cobb-Douglas Technology).
Consider the CES preferences and the Cobb-Douglas production function presented in Examples 12 – 14. Let $\sigma = 0.5$ and $\beta = 1$. Figure 4.3 portrays the perfect-foresight growth paths $\{k_t\}_{t=0}^{\infty}$ of the different economies and the corresponding welfare level, given a capital-intensive production technology. The dashed line represents the golden-rule value k_G. The following observations are immediate from Figure 4.3:

- *(i) The economies with financial intermediation display greater long-run growth than the economy without, cf. Theorem 5 (i).*
- *(ii) Since the economy without banks experiences under-accumulation, a competitive banking system has a positive impact on long-run welfare, cf. Theorem 6 (i).*
- *(iii) Since the technology is capital-intensive, the monopolistic economy displays the greatest level of long-run growth and welfare, cf. Theorems 5 (ii) and 6 (ii).*

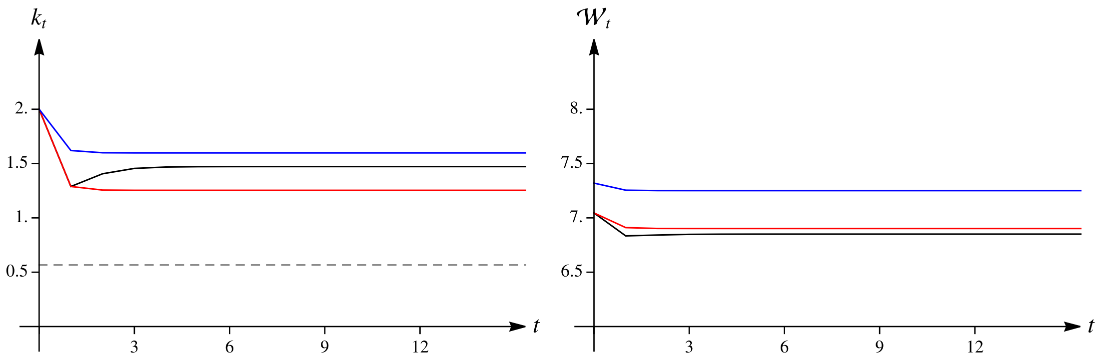
19. The observation that aggravated over-accumulation reduces consumption follows from the unimodal functional form of ϕ, cf. (4.19).



(a) Capital $\{k_t\}_{t=0}^{\infty}$ **(b)** Welfare

Figure 4.3. Monopolistic banking maximizes long-run growth and welfare.
Color code: — (PFD^N), — (PFD^C), — (PFD^M).
Parameters: $A = 10$, $\varepsilon_1 = \alpha = 0.5$, $p = 0.6$.

By contrast, Figure 4.4 depicts the growth paths for a labor-intensive technology. In this case, competition maximizes long-run growth and welfare. While, in the competitive economy, the insurance-consumption trade-off is resolved in favor of insurance, reduced consumption outweighs in the monopolistic economy, implying a reduction in steady-state welfare.



(a) Capital $\{k_t\}_{t=0}^{\infty}$ **(b)** Welfare

Figure 4.4. Perfect competition among banks maximizes long-run growth and welfare.
Color code: — (PFD^N), — (PFD^C), — (PFD^M).
Parameters: $A = 10$, $\varepsilon_1 = \alpha = 0.1$, $p = 0.6$.

4.5 Endogenous Fluctuations

The question of how financial intermediation affects the qualitative patterns of economic growth is addressed next. We investigate whether banking may cause endogenous output fluctuations, for example, in the form of oscillations, cycles, or chaos. Differentiating (PFD^N) directly shows that the dynamics without banks is monotonic. Indeed, in view of Assumption 7, we get

$$\frac{\partial k_{t+1}}{\partial k_t} = \frac{\frac{\partial I}{\partial w}(w(k_t), f'(k_{t+1}))\, w'(k_t)}{\frac{1}{p} - \frac{\partial I}{\partial \rho^e}(w(k_t), f'(k_{t+1}))\, f''(k_{t+1})} > 0$$

because investments $I(w, \rho^e)$ are strictly increasing in both w and ρ^e.[20] Hence, endogenous fluctuations are ruled out and, depending on k_0, the growth paths generated by (PFD^N) are either monotonically increasing or decreasing, cf. Figure 4.4a. Differentiating (PFD^C), we find that the exact same applies to the economy with a competitive banking industry,

$$\frac{\partial k_{t+1}}{\partial k_t} = \frac{\frac{\partial D}{\partial w}(w(k_t), pf'(k_{t+1}))\, w'(k_t)}{\frac{1}{p} - \frac{\partial D}{\partial r}(w(k_t), pf'(k_{t+1}))\, pf''(k_{t+1})} > 0.$$

Hence, in the competitive model, financial intermediation cannot generate output fluctuations since the dynamics is always monotonic. This result stands in contrast with the findings of Banerji et al. (2004), who argue that risk sharing may expose an economy to the full variety of complex dynamics. Finally, the case of a monopolistic banking is treated in the following proposition.

Proposition 16 (Monotonic Dynamics).
The perfect-foresight dynamics (PFD^M) is monotonic in the local neighborhood of the positive steady state $k^M > 0$.

Recall that by Theorem 5, all growth paths $\{k_t\}_{t=0}^{\infty}$ generated by (PFD^M) converge to $k^M > 0$. Proposition 16 thus rules out periodic cycles and chaotic dynamics as it implies that initial fluctuations will, sooner or later, vanish. The bottom line of the preceding analysis is that the perfect-foresight dynamics of this model is qualitatively equivalent to the dynamics of the standard OLG model without financial intermediation by Diamond (1965). In the monopolistic case, this insight applies to the local neighborhood of the steady state only because otherwise the Hartman-Grobman Theorem fails.

20. The comparative statics of $I(w, \rho^e)$ and $D(w, r)$ are established in technical Lemma 19 in Appendix 4.A.1.

4.6 Financial Stability

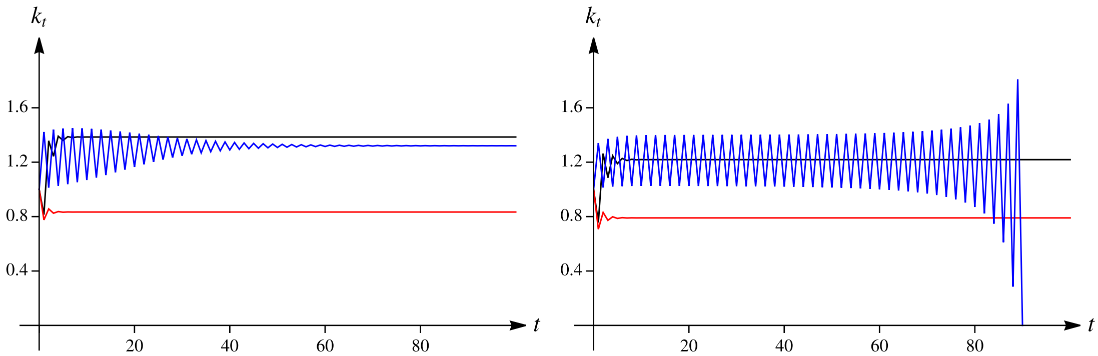
In real-world economies, banking crises severely affect economic growth and welfare. This section briefly discusses implications for the stability of the financial system. For this purpose, we consider price shocks caused by erroneous forecasts, i.e., we abstract from perfect foresight for a moment.

Overly optimistic expectations, $\rho_t^e > \rho_{t+1}$, can entail that banks accidentally go bankrupt. In a perfectly competitive banking industry, the law of large numbers implies that $r_t^C = p\rho_t^e$ is the maximal deposit rate a bank can offer in view of a given expectation ρ_t^e. Any inframarginal forecast error $\rho_t^e - \rho_{t+1} > 0$ will then result in a negative intermediation margin and negative profits. Once a bank has insufficient equity to fulfill its deposit contracts, it defaults. Monopolistic intermediation margins, however, are positive as long as the forecast error is sufficiently small. In addition, the monopolist holds a comparatively larger stock of equity, which can buffer losses caused by erroneous expectations. In fact, for any given $(k_t, k_{t-1}, k_{t-1}^e)$, it holds that $e^M(k_t, k_{t-1}, k_{t-1}^e) > e^C(k_t, k_{t-1}, k_{t-1}^e)$.

It becomes evident that fierce competition reduces the stability of the financial system by inducing banks to take more risk. The destabilization results from less conservative contracts on the one hand, and from less equity reserves on the other hand. Our model thus supports the *competition-fragility hypothesis* prevalent in the finance literature.[21]

Example 16 (Financial Stability).
We revisit the parameterization from Example 15. Figure 4.5 depicts the naive-expectations dynamics, that is, $k_t^e = k_t$ *for all* $t \geq 0$. *The growth path of the the monopolistic economy experiences significantly less volatility than that of the competitive economy. While the monopolistic economy attains a positive steady state, the destabilizing effect of competition becomes critical such that the financial system collapses.*

21. However, the competition-fragility hypothesis is not undisputed among economists (see, for instance, Boyd and De Nicolo, 2005).



(a) Damped oscillations ($\alpha = 0.2$) **(b)** Financial collapse ($\alpha = 0.169$)

Figure 4.5. Dynamic perspective on the competition-fragility hypothesis.
Color code: — no banks, — competition, — monopoly.
Parameters: $A = 10$, $p = 0.4$.

The result that erroneous beliefs may trigger endogenous fluctuations in an economy that otherwise displays monotonic perfect-foresight dynamics is consistent with previous research in the literature on OLG models (see, for example, De La Croix and Michel, 2002).

4.7 Dividend Payments

Thus far, we analyzed a rather restrictive setting without dividend payments of banks. We next conduct a robustness check for our previous results by analyzing a less restrictive regulatory environment and incorporating dividend payments into the model. It should be noted that the case of perfect competition is unaffected by dividend payments since, under perfect foresight, banks do not realize profits.

Consider a risk-neutral monopolistic bank that is collectively owned by agents. The bank maximizes its expected profit by raising deposits from the public. Young agents become shareholders of the bank by accepting a deposit contract. Dividends are distributed to old agents. Let $\vartheta_t^e \geq 0$ denote a young agent's expectation formed in period t with respect to the dividend payment $\vartheta_{t+1} \geq 0$ realized in $t+1$. The bank's dividend policy is given by the parameter $0 \leq \mu < 1$. It specifies the share of the bank's equity that is paid out to the shareholders. For simplicity, we assume that μ is exogenous and time-invariant.

Given the wage rate w_t, the deposit rate r_t, and the anticipated dividend payment ϑ_t^e, a young agent's supply of deposits solves

$$D_t \;=\; D(w_t, r_t, \vartheta_t^e) \;:=\; \operatorname*{argmax}_{0 \le D \le w_t} \; u(w_t - D) + v(r_t D + \vartheta_t^e). \tag{4.20}$$

Accordingly, the participation constraint takes the form

$$u(w_t - D(w_t, r_t, \vartheta_t^e)) + v(r_t D(w_t, r_t, \vartheta_t^e) + \vartheta_t^e) \;\ge\; U_{\text{res}}(w_t, \rho_t^e),$$

with the reservation utility U_{res} as defined in (4.3). Observe that dividend payments relax the participation constraint compared to the basic model. Since part of the banking surplus is distributed to the shareholders, agents are now willing to accept contracts which they would otherwise reject. Given w_t, e_t, and the expectations ρ_t^e and ϑ_t^e, the bank's decision problem reads

$$\begin{aligned} &\max_{0 \le r \le p\rho_t^e} \; D(w_t, r, \vartheta_t^e)\,(p\rho_t^e - r) + p\rho_t^e(1-\mu)e_t \\ &\quad \text{s.t.} \quad u(w_t - D(w_t, r, \vartheta_t^e)) + v(rD(w_t, r, \vartheta_t^e) + \vartheta_t^e) \;\ge\; U_{\text{res}}(w_t, \rho_t^e). \end{aligned} \tag{4.21}$$

Existence and uniqueness of a solution to Problem (4.21) is established in the following lemma.

Lemma 16. *For any given $w_t > 0$, $\rho_t^e > 0$, and $\vartheta_t^e \ge 0$, there exists a unique deposit rate $0 < r_t^{\mathrm{M}} = r^{\mathrm{M}}(w_t, \rho_t^e, \vartheta_t^e) < p\rho_t^e$ that solves Problem (4.21).*[22]

As before, equity e_t in period t is determined by a map of the form

$$e^{\mathrm{M}}(k_t, k_{t-1}, k_{t-1}^e, \vartheta_{t-1}^e) \;:=\; f'(k_t)k_t - r^{\mathrm{M}}(w(k_{t-1}), f'(k_{t-1}^e), \vartheta_{t-1}^e)\, D(w(k_{t-1}), r^{\mathrm{M}}(\cdot), \vartheta_{t-1}^e)$$

and the realized dividend payment ϑ_t to an old agent amounts to

$$\vartheta_t \;=\; \vartheta^{\mathrm{M}}\big(k_t, k_{t-1}, k_{t-1}^e, \vartheta_{t-1}^e\big) \;:=\; \mu\, e^{\mathrm{M}}(k_t, k_{t-1}, k_{t-1}^e, \vartheta_{t-1}^e).$$

The economic law of capital accumulation becomes

$$\begin{aligned} k_{t+1} = G(k_t, k_{t-1}, k_t^e, k_{t-1}^e, \vartheta_t^e, \vartheta_{t-1}^e) &:= pD(w(k_t), r^{\mathrm{M}}(w(k_t), f'(k_t^e), \vartheta_t^e), \vartheta_t^e) \\ &\quad + p(1-\mu)e^{\mathrm{M}}(k_t, k_{t-1}, k_{t-1}^e, \vartheta_{t-1}^e). \end{aligned}$$

We maintain the assumption that expectations are homogeneous and rational. However, agents must now form expectations regarding both, future capital *and*

22. A formal definition of the map $r^{\mathrm{M}}(w_t, \rho_t^e, \vartheta_t^e)$ is provided as part of the proof of Lemma 16 in Appendix 4.A.2.

dividend payments. Consequently, the expectations-feedback effect becomes more involved because the forecasts k_t^e and ϑ_t^e must be mutually consistent: if k_t^e is a correct forecast for k_{t+1}, then $\vartheta_t^e = \tilde{\psi}(k_t^e, k_t)$ must be a correct forecast for ϑ_{t+1} and vice versa.[23]

Lemma 17. *For any given* $(k_t, k_{t-1}) \in \mathscr{C}$*, where*

$$\mathscr{C} := \left\{(x,y) \in \mathbb{R}_+^2 \mid e^{\mathrm{M}}(x,y,x,\tilde{\psi}(x,y)) \geq 0\right\},$$

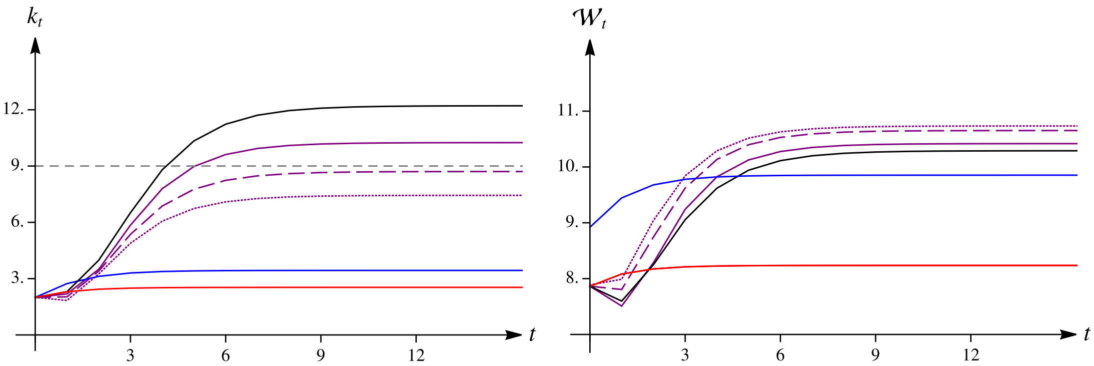
there exists a unique pair of perfect forecasts (k_t^e, ϑ_t^e) *in period t.*

As in the basic model, the set $\mathscr{C}$ is forward-invariant, implying that the perfect-foresight dynamics is well defined by the difference equation

$$k_{t+1} = G\big(k_t, k_{t-1}, k_{t+1}, k_t, \tilde{\psi}(k_{t+1}, k_t), \tilde{\psi}(k_t, k_{t-1})\big), \quad t \geq 1. \qquad (\mathrm{PFD}^{\vartheta})$$

An analytical investigation of the dynamics generated by (PFD^{ϑ}) is beyond the scope of this article. The important point is that the model without dividend payments reobtains if the parameter μ approaches zero. This is because agents with perfect foresight anticipate that no dividends will be paid out to them if $\mu \to 0$. Technically, $\mu \to 0$ causes that $\vartheta_t^e = \tilde{\psi}(k_t^e, k_t) \to 0$ for all $t \geq 0$ such that the growth path of the dynamical system (PFD^{ϑ}) coincides with that of ($\mathrm{PFD}^{\mathrm{M}}$). Figure 4.6 provides an illustration using a numerical example with CES preferences and a Cobb-Douglas production technology. We can thus conclude that the results in Theorem 5 and 6 also apply if the monopolistic bank distributes part of the surplus to agents, provided that the parameter μ is not too large. These considerations demonstrate that if a banking monopoly occurs in conjunction with a capital-intensive production sector and the dividend payments are constrained by regulation, then it may outperform its perfectly competitive counterpart. The model shows that from a dynamic viewpoint, more competition is not necessarily beneficial.

23. A formal definition of the forecasting rule $\tilde{\psi}$ and a proof of its existence and uniqueness is given in the proof of Lemma 17 in Appendix 4.A.2.



(a) Capital **(b)** Welfare

Figure 4.6. Effect of the dividend parameter μ on economic growth and welfare.
Color code: — (PFD^{N}), — (PFD^{C}), — (PFD^{M}), — (PFD$^{\theta}$).
Parameters: $A = 10$, $\alpha = 0.5$, $p = 0.6$, $\mu = 0.07$ (solid), 0.14 (dashed), 0.21 (dotted).

4.8 Conclusion

This article adopted a novel, micro-founded dynamical-systems approach to explore how competition among banks affects real economic growth, the occurrence of endogenous fluctuations, and agents' welfare. We have shown that the provision of risk sharing by banks mobilizes additional funds for productive activities, which fosters capital accumulation and leads to a higher level of long-term economic growth. In economies that experience under-accumulation of capital, financial intermediation improves welfare by providing insurance and enhancing the investment efficiency. This result implies that implementing a well-functioning banking system is of utmost importance in under-developed countries facing severe developmental obstacles.

The central finding of this paper is that a banking monopoly can induce greater long-run growth and welfare compared to its competitive analog. The prerequisite is that the production sector is highly financially dependent and that the bank's dividend payments are sufficiently restricted. Otherwise, competition between banks is favorable. For this reason, a growth and welfare-improving effect of reduced competition is solely conceivable in industrialized economies featuring capital-intensive branches such as the automotive industry, pharmaceuticals, or telecommunication. Our theoretical results are consistent with the empirical findings of Cetorelli and Gambera (2001) and Hoxha (2013), who demonstrate that banking concentration has a growth-promoting effect in financially-dependent industries. Moreover, Maudos and Fernandez de Guevara (2006) find that banks' exercise of market power promotes economic growth in a sample of particularly industrialized countries. On the other hand, our model revealed that imperfect competition among banks harms the growth of labor-intensive economies. Again, there exists empirical evidence. Beck et al. (2004) demonstrate that banking concentration implies financing obstacles for firms in developing countries only, i.e., those typically characterized by labor-intensive industries such as textiles, hospitality, or agriculture. For such countries, a negative impact of banking concentration on economic growth was also identified by Deidda and Fattouh (2005). Finally, Cetorelli and Gambera (2001) and Hoxha (2013) show that banking concentration curbs the growth of industries that have low levels of financial dependence.

Apart from that, our analysis discovered two other interesting results. First, that financial intermediation cannot adversely affect the qualitative economic dynamics in two-period lived OLG models with rational expectations. In particular, the perfect-foresight dynamics with perfect competition is always monotonic. Second, once one allows for erroneous beliefs, however, evidence for the competition-fragility hypothesis obtains and the competitive banking system proves itself to be more susceptible to volatility and financial crises.

Like any other model, the one at hand does not come without limitations. While most of our assumptions on preferences and the production technology are standard in this strand of literature, a world with homogeneous two-period lived agents is, of course, stylized. The approach of initially disregarding dividend payments and incorporating them afterwards is limiting, but necessary to make the analytical complexity somewhat manageable. Nevertheless, the extension of the model has proven that our main findings are robust even if part of the banking surplus is redistributed to agents. Considering the provoking result that a monopoly may outperform its competitive counterpart, we think that a fruitful avenue for future research is to investigate the optimal regulation of endogenously determined dividend payments and how these may affect economic growth and welfare.

Appendix 4.A

Appendix 4.A.1 introduces three technical lemmas that facilitate the proofs of the main results, which are provided in Appendix 4.A.2. In Appendix 4.A.3, we discuss our welfare measure in light of the literature on OLG models.

4.A.1 Technical Lemmas

Lemma 18. *The elasticities of f and w satisfy*

$$0 < \frac{f'(k)k}{f(k)} < 1, \quad \text{and} \quad 0 < \frac{w'(k)k}{w(k)} < 1 \quad \text{for all} \quad k \geq 0.$$

Moreover, it holds that $\lim_{k\to 0} w'(k) = \infty$.

Proof of Lemma 18. By Assumption 8 (i) and (ii), the functions f and w are strictly concave, implying that their elasticities lie strictly between zero and one. The Inada condition $\lim_{k\to 0} f'(k) = \infty$ stated in Assumption 8 (i) together the property $-1 < \epsilon_2(k) < 0$ stated in Assumption 8 (ii) implies that $\lim_{k\to 0} w'(k) = \infty$. □

Lemma 19. *The following comparative statics hold.*

(i) *$I(w,\rho^e)$ defined in (4.1) is strictly increasing in both w and ρ^e.*
(ii) *$D(w,r)$ defined in (4.4) is strictly increasing in both w and r.*
(iii) *$\eta(w,r)$ defined in (4.8) is non-decreasing in w and strictly decreasing in r.*
(iv) *$r^{\mathrm{M}}(w,\rho^e)$ in Proposition 14 is strictly increasing in ρ^e and non-decreasing in w.*
(v) *$\eta(w,r,\vartheta^e) = \frac{\partial D}{\partial r}(w,r,\vartheta^e)r/D(w,r,\vartheta^e)$, with $D(w,r,\vartheta^e)$ as defined in (4.20), is strictly decreasing in r.*

Proof of Lemma 19.

Part (i). Since $0 < \alpha_u \leq \alpha_v < 1$ by Assumption 7 (ii), differentiation of (4.2) yields

$$\frac{\partial I}{\partial w}(w,\rho^e) = \frac{\alpha_u}{\alpha_u + \alpha_v \frac{w - I(w,\rho^e)}{I(w,\rho^e)}} > 0$$

and

$$\frac{\partial I}{\partial \rho^e}(w,\rho^e) = \frac{(1-\alpha_v)I(w,\rho^e)}{\rho^e\left(\alpha_u \frac{I(w,\rho^e)}{w - I(w,\rho^e)} + \alpha_v\right)} > 0.$$

Part (ii). Since $0 < \alpha_u \leq \alpha_v < 1$ by Assumption 7 (ii), differentiation of (4.5) yields

$$\frac{\partial D}{\partial w}(w,r) \;=\; \frac{\alpha_u}{\alpha_u + \alpha_v \frac{w-D(w,r)}{D(w,r)}} \;>\; 0 \tag{4.A.1}$$

and

$$\frac{\partial D}{\partial r}(w,r) \;=\; \frac{(1-\alpha_v)\,D(w,r)}{r\left(\alpha_u \frac{D(w,r)}{w-D(w,r)} + \alpha_v\right)} \;>\; 0. \tag{4.A.2}$$

Part (iii). Using (4.A.2), the elasticity $\eta(w,r)$ may be written as

$$\eta(w,r) \;=\; \frac{1-\alpha_v}{\alpha_u \frac{D(w,r)}{w-D(w,r)} + \alpha_v} \;>\; 0. \tag{4.A.3}$$

Differentiating (4.A.3) yields

$$\frac{\partial \eta}{\partial r}(w,r) \;=\; \frac{\alpha_u}{\alpha_v - 1}\,\frac{w\,\eta(w,r)^2\,\frac{\partial D}{\partial r}(w,r)}{(w-D(w,r))^2} \;<\; 0,$$

which is strictly negative in view of (4.A.2) and Assumption 7 (ii). Finally, set

$$\xi(w,r) \;:=\; \frac{\frac{\partial D}{\partial w}(w,r)\,w}{D(w,r)}$$

and obtain that

$$\frac{\partial \eta}{\partial w}(w,r) \;=\; \frac{D(w,r)(1-\xi(w,r))(1-\alpha_v)\alpha_u}{\left((\alpha_u \frac{D(w,r)}{w-D(w,r)} + \alpha_v)(w-D(w,r))\right)^2} \;\geq\; 0$$

since Assumption 7 (ii) implies $0 < \xi \leq 1$.

Part (iv). Since $v' > 0$ by Assumption 7 (i), differentiation of (4.9) yields

$$\frac{\partial b}{\partial \rho^e}(w,\rho^e) \;=\; \frac{v'(\rho^e I(w,\rho^e))\,p\,I(w,\rho^e)}{v'(b(w,\rho^e)D(w,b(w,\rho^e)))\,D(w,b(w,\rho^e))} \;>\; 0.$$

Differentiating (4.7) and using Part (iii) shows that

$$\frac{\partial s}{\partial \rho^e}(w,\rho^e) \;=\; p\left(\frac{p\rho^e}{s(w,\rho^e)} - \frac{s(w,\rho^e)}{\eta(w,s(w,\rho^e))}\,\frac{\partial \eta}{\partial r}(w,s(w,\rho^e))\right)^{-1} \;>\; 0.$$

Hence, we have that $\partial r^{\mathrm{M}}(w,\rho^e)/\partial \rho^e > 0$. Furthermore, differentiation of (4.9) yields

$$\frac{\partial b}{\partial w}(w,\rho^e) = \frac{u'(w - I(w,\rho^e)) - u'(w - D(w,b(w,\rho^e)))}{v'(b(w,\rho^e)D(w,b(w,\rho^e)))\,D(w,b(w,\rho^e))} = 0$$

because $I(w,\rho^e) = D(w,b(w,\rho^e)) = 0$ by Corollary 5. Since, by Part (iii),

$$\frac{\partial \eta(w,r)}{\partial w} \geq 0, \quad \text{and} \quad \frac{\partial \eta(w,r)}{\partial r} < 0,$$

differentiation of (4.7) yields

$$\frac{\partial s}{\partial w}(w,\rho^e) = \frac{\partial \eta}{\partial w}(w,s(w,\rho^e))\left(\frac{p\rho^e \eta(w,s(w,\rho^e))^2}{s(w,\rho^e)^2} - \frac{\partial \eta}{\partial r}(w,s(w,\rho^e))\right)^{-1} \geq 0.$$

Hence, it holds that $\partial r^{\mathrm{M}}(w,\rho^e)/\partial w \geq 0$.

Part (v). The elasticity $\eta(w,r,\vartheta^e)$ may be written as

$$\eta(w,r,\vartheta^e) = \frac{1 - \alpha_v + \frac{\vartheta^e}{rD(w,r,\vartheta^e)}}{\alpha_u \frac{D(w,r,\vartheta^e) + \frac{\vartheta^e}{r}}{w - D(w,r,\vartheta^e)} + \alpha_v} > 0.$$

Differentiation yields

$$\frac{\partial \eta}{\partial r}(w,r,\vartheta^e) = -\frac{(1+\eta(w,r,\vartheta^e))\,\vartheta^e}{r^2 D(w,r,\vartheta^e)\,T(w,r,\vartheta^e)} - \frac{\eta(w,r,\vartheta^e)\,\frac{\partial T}{\partial r}(w,r,\vartheta^e)}{T(w,r,\vartheta^e)},$$

where, for notational convenience, we set

$$T(w,r,\vartheta^e) := \alpha_u \frac{D(w,r,\vartheta^e) + \frac{\vartheta^e}{r}}{w - D(w,r,\vartheta^e)} + \alpha_v > 0.$$

Observe that $\partial \eta(w,r,\vartheta^e)/\partial r < 0$ if $\partial T(w,r,\vartheta^e)/\partial r > 0$, or equivalently, if

$$\eta(w,r,\vartheta^e) > \frac{w - D(w,r,\vartheta^e)}{D(w,r,\vartheta^e)\left(\frac{\vartheta^e + wr}{\vartheta^e}\right)}. \tag{4.A.4}$$

By inserting $\eta(w,r,\vartheta^e)$ into (4.A.4) and rearranging, we get

$$1 > \frac{rw\alpha_v + \vartheta^e \alpha_u}{rw + \vartheta^e}. \tag{4.A.5}$$

Observe that (4.A.5) is always satisfied since $0 < \alpha_u, \alpha_v < 1$ by Assumption 7 (ii). Consequently, $\partial \eta(w,r,\vartheta^e)/\partial r < 0$. □

Lemma 20. *The positive steady state $k^{\mathrm{M}} > 0$ of (PFD$^{\mathrm{M}}$) satisfies $k^{\mathrm{M}} > f'^{-1}(1/p)$.*

Proof of Lemma 20. Under perfect foresight, monopolistic equity $e^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}, k^{\mathrm{M}})$ in a steady state $k^{\mathrm{M}} > 0$ must be strictly positive because strictly positive intermediation margins realize at all times. From the strictly positive intermediation margin $pf'(k^{\mathrm{M}}) - r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})) > 0$ and the condition that

$$e^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}, k^{\mathrm{M}}) \;=\; \frac{pf'(k^{\mathrm{M}}) - r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))}{1 - pf'(k^{\mathrm{M}})} D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))) \;>\; 0,$$

we can then infer that any positive steady state $k^{\mathrm{M}} > 0$ must satisfy $1 - pf'(k^{\mathrm{M}}) > 0$. Since $f'' < 0$, it thus follows that $k^{\mathrm{M}} > f'^{-1}(1/p)$. □

4.A.2 Proofs of the Main Results

Proof of Proposition 13. Applying standard arguments from microeconomic theory, the expected intermediation margin in a perfectly competitive banking industry must be zero since the full expected surplus is with the consumers. Hence, $r_t^{\mathrm{C}} = p\rho_t^e$ must hold. By the strict concavity of v, $r_t^{\mathrm{C}} = p\rho_t^e$ satisfies the participation constraint (PC) because

$$\begin{aligned} U_{\mathrm{res}}(w_t,\rho_t^e) &< u(w_t - I(w_t,\rho_t^e)) + v(p\rho_t^e I(w_t,\rho_t^e)) \\ &= u(w_t - I(w_t,\rho_t^e)) + v(r_t^{\mathrm{C}} I(w_t,\rho_t^e)) \\ &\leq u(w_t - D(w_t,r_t^{\mathrm{C}})) + v(r_t^{\mathrm{C}} D(w_t,r_t^{\mathrm{C}})), \end{aligned}$$

where the last inequality is implied by the definition of $D(w,r)$ in (4.4). □

Proof of Proposition 14. The proof works in three steps. First, we prove existence and uniqueness of $s(w,\rho^e)$ and $b(w,\rho^e)$, using the INTERMEDIATE VALUE THEOREM. In Step 2, we prove the proposed functional form of $r^{\mathrm{M}}(w,\rho^e)$. Finally, in Step 3, we show that the sufficient conditions for a maximum are satisfied.

Step 1. Let $w_t > 0$ and $\rho_t^e > 0$ be arbitrary but fixed. The map $s(w_t,\rho_t^e)$ is defined by (4.7), which may be written as

$$\frac{p\rho_t^e}{r} - 1 - \frac{1}{\eta(w_t,r)} = 0. \tag{4.A.6}$$

Denote by $J(r)$ the l.h.s. of (4.A.6). The limits of $J(r)$ are

$$\lim_{r\to 0} J(r) = \infty - \frac{1}{\eta(w_t,0)} = \infty$$

because $\eta(w_t,0) = (1-\alpha_v)/\alpha_v > 0$ and

$$\lim_{r\to p\rho_t^e} J(r) = \frac{-1}{\eta(w_t,p\rho_t^e)} < 0.$$

Since $J(r)$ is strictly decreasing by Lemma 19 (iii),

$$J'(r) = -\frac{p\rho_t^e}{r^2} + \frac{1}{\eta(w_t,r)^2}\frac{\partial\eta}{\partial r}(w_t,r) < 0,$$

a uniquely determined solution $0 < s(w_t, \rho_t^e) < p\rho_t^e$ to (4.A.6) exists. The map $b(w_t, \rho_t^e)$ is defined by a solution to

$$B(r) \;=\; U_{\text{res}}(w_t, \rho_t^e), \tag{4.A.7}$$

where $B(r)$ denotes the l.h.s. of (4.9). For $r \to 0$, we obtain the limit

$$\lim_{r\to 0} B(r) \;=\; u(w_t - D(w_t, 0)) \;=\; u(w_t) \;<\; U_{\text{res}}(w_t, \rho_t^e).$$

On the other hand, it follows from the definition of $D(w, r)$ in (4.4) and the strict concavity of v that

$$\begin{aligned}
\lim_{r\to p\rho_t^e} B(r) \;&=\; u(w_t - D(w_t, p\rho_t^e)) + v(p\rho_t^e D(w_t, p\rho_t^e)) \\
&\geq\; u(w_t - I(w_t, \rho_t^e)) + v(p\rho_t^e I(w_t, \rho_t^e)) \\
&>\; u(w_t - I(w_t, \rho_t^e)) + p\, v(\rho_t^e I(w_t, \rho_t^e)) \\
&=\; U_{\text{res}}(w_t, \rho_t^e).
\end{aligned}$$

Since, by the Envelope Theorem, $B(r)$ is strictly increasing,

$$B'(r) \;=\; v'(rD(w_t, r))D(w_t, r) \;>\; 0,$$

a uniquely determined solution $0 < b(w_t, \rho_t^e) < p\rho_t^e$ to (4.A.7) exists.

Step 2. We prove that the deposit rate

$$r_t^{\text{M}} \;=\; r^{\text{M}}(w_t, \rho_t^e) \;=\; \max\Big\{s(w_t, \rho_t^e), b(w_t, \rho_t^e)\Big\}$$

solves Problem (4.6). By the Envelope Theorem, the l.h.s. of (PC) is strictly increasing in r,

$$\frac{\mathrm{d}}{\mathrm{d}r}\Big(u(w - D(w, r)) + v(rD(w, r))\Big) \;=\; v'(rD(w, r))\, D(w, r) \;>\; 0. \tag{4.A.8}$$

Since $r^{\text{M}}(w_t, \rho_t^e) \geq b(w_t, \rho_t^e)$, we can conclude that $r^{\text{M}}(w_t, \rho_t^e)$ is always accepted by agents, i.e., it satisfies (PC). To verify that this contract maximizes the monopolist's expected profit, note that two cases can occur. First, if $s(w_t, \rho_t^e) > b(w_t, \rho_t^e)$, then $r^{\text{M}}(w_t, \rho_t^e) = s(w_t, \rho_t^e)$ and, by construction of $s(w_t, \rho_t^e)$, the objective function attains its maximum. Second, if $s(w_t, \rho_t^e) \leq b(w_t, \rho_t^e)$, then (PC) is binding and the optimum is determined by (4.9), implying that $r^{\text{M}}(w_t, \rho_t^e) = b(w_t, \rho_t^e)$.

Step 3. We prove that $r^{\mathrm{M}}(w_t, \rho_t^e)$ satisfies the sufficient conditions for a maximum by showing that the participation constraint in Problem (4.6) is convex, while the objective function is strictly quasi-concave on $[0, p\rho_t^e]$. Since, by (4.A.8), the l.h.s. of (PC) is strictly increasing in r, we can conclude that the participation constraint indeed defines a convex set of deposit rates. The quasi-concavity of the objective function is seen as follows. First, note that

$$\lim_{r \to 0} \pi_t^e(r) = p\rho_t^e e_t, \qquad \text{and} \qquad \lim_{r \to p\rho_t^e} \pi_t^e(r) = p\rho_t^e e_t,$$

because either the deposit supply or the intermediation margin converges to zero. Since the first-order condition (4.7) has a unique solution, the objective function in (4.6) has a unique stationary point on $[0, p\rho_t^e]$ at $0 < s(w_t, \rho_t^e) < p\rho_t^e$. Since

$$\pi_t^e(s(w_t, \rho_t^e)) > p\rho_t^e e_t,$$

we can conclude that the objective function must indeed be strictly quasi-concave.

□

Proof of Lemma 12. Let $w_t > 0$ and $\rho_t^e > 0$ be arbitrary but fixed. With a slight abuse of notation, let $\mathscr{P}_t$ denote the set of solutions $0 < p < 1$ to

$$b(w_t, \rho_t^e, p) = s(w_t, \rho_t^e, p). \tag{4.A.9}$$

The proof works in two steps and its idea is depicted in Figure 4.A.1. In Step 1, we exploit the Intermediate Value Theorem to show that at least one solution to Condition (4.A.9) exists. In the second step, we define the thresholds for p, based on the set $\mathscr{P}_t$.

Step 1. Since, by Proposition 14, $0 < s(w_t, \rho_t^e, p), b(w_t, \rho_t^e, p) < p\rho_t^e$, it follows that

$$\lim_{p \to 0} b(w_t, \rho_t^e, p) = 0, \quad \text{and} \quad \lim_{p \to 0} s(w_t, \rho_t^e, p) = 0.$$

From Condition (4.9), it can be shown that $\lim_{p \to 1} b(w_t, \rho_t^e, p) = \rho_t^e$. Moreover, it can be deduced from Condition (4.7) that $\lim_{p \to 1} s(w_t, \rho_t^e, p) < \rho_t^e$. To apply the Intermediate Value Theorem, it remains to establish that

$$\lim_{p \to 0} \frac{\partial s}{\partial p}(w_t, \rho_t^e, p) > \lim_{p \to 0} \frac{\partial b}{\partial p}(w_t, \rho_t^e, p).$$

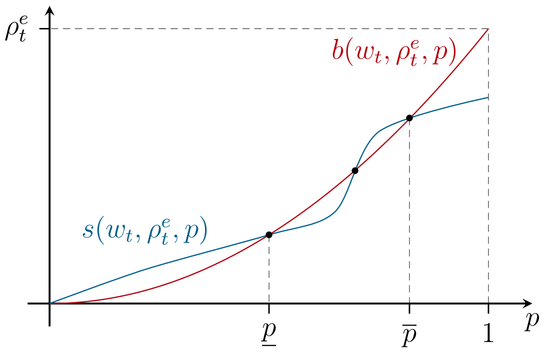


Figure 4.A.1. Elements of $\mathscr{P}_t$ are determined by solutions to (4.A.9) and define the thresholds for the success rate p.

Implicit differentiation of Condition (4.7) yields

$$\frac{\partial s}{\partial p}(w_t, \rho_t^e, p) \;=\; \frac{(1-\alpha_v)\,\rho_t^e}{1+\alpha_u \frac{D(w_t, s(w_t,\rho_t^e,p))\, w_t\, \left(1+\eta(w_t, s(w_t,\rho_t^e,p))\right) - D(w_t, s(w_t,\rho_t^e,p))^2}{\left(w_t - D(w_t, s(w_t,\rho_t^e,p))\right)^2}}.$$

Since $\lim_{p\to 0} D(w, s(w, \rho^e, p)) = 0$, it follows that

$$\lim_{p\to 0} \frac{\partial s}{\partial p}(w, \rho^e, p) \;=\; (1-\alpha_v)\,\rho_t^e \;>\; 0.$$

Consider Condition (4.9) now. The Envelope Theorem implies that the r.h.s. of (4.9) satisfies

$$\lim_{p\to 0} \frac{\mathrm{d}U_{\text{res}}}{\mathrm{d}p}(w_t, \rho_t^e, p) \;=\; v(\rho_t^e I(w_t, \rho_t^e, p)) \;=\; v(0) \;=\; 0.$$

By contrast, the l.h.s. of (4.9) is independent of p. Therefore, we can conclude that

$$\lim_{p\to 0} \frac{\partial b}{\partial p}(w, \rho^e, p) \;=\; 0.$$

Therefore, we indeed have that

$$\lim_{p\to 0} \frac{\partial s}{\partial p}(w_t, \rho_t^e, p) \;>\; \lim_{p\to 0} \frac{\partial b}{\partial p}(w_t, \rho_t^e, p).$$

The Intermediate Value Theorem now implies that for any given $w_t > 0$ and $\rho_t^e > 0$, there exists a solution $0 < p < 1$ to (4.A.9), implying that $\mathscr{P}_t$ is non-empty.

Step 2. From the limits of $s(w,\rho^e,p)$ and $b(w,\rho^e,p)$ in Step 1, it follows that

$$s(w_t,\rho_t^e,p) > b(w_t,\rho_t^e,p) \quad \text{for all} \quad p < \min \mathscr{P}_t.$$

Proposition 14 then implies that

$$r_t^{\mathrm{M}} = \max\left\{s(w_t,\rho_t^e,p), b(w_t,\rho_t^e,p)\right\} = s(w_t,\rho_t^e,p) \quad \text{for all} \quad p < \min \mathscr{P}_t.$$

By contrast, the limits obtained in Step 1 also imply that

$$s(w_t,\rho_t^e,p) < b(w_t,\rho_t^e,p) \quad \text{for all} \quad p > \max \mathscr{P}_t.$$

Therefore, by Proposition 14,

$$r_t^{\mathrm{M}} = \max\left\{s(w_t,\rho_t^e,p), b(w_t,\rho_t^e,p)\right\} = b(w_t,\rho_t^e,p) \quad \text{for all} \quad p > \max \mathscr{P}_t.$$

Hence, we define $\underline{p} := \min \mathscr{P}_t$ and $\overline{p} := \max \mathscr{P}_t$ and the proof is complete. □

Proof of Corollary 5. First, consider the second inequality. Since, by Lemma 19 (ii), $D(w,r)$ is strictly increasing in r and, by Propositions 13 and 14, $r^{\mathrm{M}}(w_t,\rho_t^e) < r^{\mathrm{C}}(w_t,\rho_t^e)$, it follows that $D(w_t, r^{\mathrm{M}}(w_t,\rho_t^e)) < D(w_t, r^{\mathrm{C}}(w_t,\rho_t^e))$. Next, consider the first inequality. Two cases can occur. First, if $r^{\mathrm{M}}(w_t,\rho_t^e) = b(w_t,\rho_t^e) \geq s(w_t,\rho_t^e)$, then the monopolist's rent is determined by the risk premium $RP > 0$, which solves

$$\begin{aligned} U_{\text{res}}(w_t,\rho_t^e) &= u(w_t - I(w_t,\rho_t^e)) + v(p\rho_t^e I(w_t,\rho_t^e) - RP) \\ &= u(w_t - I(w_t,\rho_t^e)) \\ &\quad + v(p\rho_t^e[I(w_t,\rho_t^e) - D(w_t,b(w_t,\rho_t^e))] + b(w_t,\rho_t^e)D(w_t,b(w_t,\rho_t^e))). \end{aligned}$$

Substituting $U_{\text{res}}(w_t,\rho_t^e)$ using (4.9) yields

$$\begin{aligned} & u(w_t - D(w_t,b(w_t,\rho_t^e))) + v(b(w_t,\rho_t^e))D(w_t,b(w_t,\rho_t^e))) \\ = \;& u(w_t - I(w_t,\rho_t^e)) \\ + \;& v(p\rho_t^e[I(w_t,\rho_t^e) - D(w_t,b(w_t,\rho_t^e))] + b(w_t,\rho_t^e)D(w_t,b(w_t,\rho_t^e))). \end{aligned} \tag{4.A.10}$$

Observe that $D(w_t,b(w_t,\rho_t^e)) = I(w_t,\rho_t^e)$ solves (4.A.10).[24] Hence, if $r^{\mathrm{M}}(w_t,\rho_t^e) = b(w_t,\rho_t^e)$, then $D(w_t, r^{\mathrm{M}}(w_t,\rho_t^e)) = I(w_t,\rho_t^e)$.

24. There actually exists a second solution $D(w_t,b(w_t,\rho_t^e)) \neq I(w_t,\rho_t^e)$ which, however, is discarded since it violates (4.5).

The second case is that $r^{\mathrm{M}}(w_t, \rho_t^e) = s(w_t, \rho_t^e) > b(w_t, \rho_t^e)$. In this case, Lemma 19 (ii) implies $D(w_t, r^{\mathrm{M}}(w_t, \rho_t^e)) > D(w_t, b(w_t, \rho_t^e)) = I(w_t, \rho_t^e)$. Combining both of cases, we get $D(w_t, r^{\mathrm{M}}(w_t, \rho_t^e)) \geq I(w_t, \rho_t^e)$, where the strict inequality holds if and only if $s(w_t, \rho_t^e) > b(w_t, \rho_t^e)$. □

Proof of Lemma 13. The existence and uniqueness of the perfect forecasting rule ψ^{N} obtains as follows. Let $k_t > 0$ be arbitrary but fixed. A perfect forecast k_t^e is determined by a solution to

$$k^e - pI(w(k_t), f'(k^e)) \;=\; 0. \tag{4.A.11}$$

For $k^e \to 0$, the l.h.s. in (4.A.11) is strictly negative, whereas it converges to infinity for $k^e \to \infty$ because $pI(w(k_t), f'(\infty)) < w(k_t) < \infty$. Since the l.h.s. in (4.A.11) is strictly increasing by Lemma 19 (i) and Assumption 8 (i),

$$1 - p\,\frac{\partial I}{\partial \rho^e}(w(k_t), f'(k^e))f''(k^e) \;>\; 0,$$

it follows that there exists a uniquely determined solution $0 < k_t^e < \infty$ to (4.A.11). Hence, the forecasting rule ψ^{N} is well defined. Since $k_t > 0$ was arbitrary, it follows that ψ^{N} is perfect in the sense of Definition 9 (i).

Next, we prove existence and uniqueness of the perfect forecasting rules ψ^i, $i \in \{\mathrm{C}, \mathrm{M}\}$, in the sense of Definition 9 (ii). Let $(k_t, k_{t-1}) \in \mathscr{D}^i$ be given. A perfect forecast k_t^e is determined by a solution to

$$\frac{k^e}{p} - D(w(k_t), r^i(w(k_t), f'(k^e))) \;=\; e^i(k_t, k_{t-1}, k_t). \tag{4.A.12}$$

For $k^e \to 0$, the l.h.s. in (4.A.12) converges to $-D(w(k_t), r^i(w(k_t), f'(0))) < 0$, whereas it converges to infinity for $k^e \to \infty$. Since, by construction of $\mathscr{D}^i$,

$$0 \;\leq\; e^i(k_t, k_{t-1}, k_t) \;<\; \infty,$$

a solution $0 < k_t^e < \infty$ to (4.A.12) exists. Since the l.h.s. in (4.A.12) is strictly increasing by Assumption 8 (i) and Lemma 19,

$$\frac{1}{p} - \frac{\partial D}{\partial r}(w(k_t), r^i(w(k_t), f'(k^e)))\,\frac{\partial r^i}{\partial \rho^e}(w(k_t), f'(k^e))f''(k^e) \;>\; 0, \tag{4.A.13}$$

the solution is also unique. Hence, the perfect forecasting rule ψ^i, $i \in \{\mathrm{C}, \mathrm{M}\}$, is well defined.

It remains to prove that $\mathscr{D}^i$ is forward-invariant for ψ^i, $i \in \{\mathrm{C}, \mathrm{M}\}$. The proof works by induction.

Induction base case. Consider period $t = 0$, with initial capital $k_0 > 0$ given. Since $e^i(k_0, k_{-1}, k_0) = e_0 = 0$, we have, by a slight abuse of notation, $(k_0, k_{-1}) \in \mathscr{D}^i$.

Induction step. Consider period $t = n$. By the induction hypothesis, we claim that $(k_n, k_{n-1}) \in \mathscr{D}^i$ such that $e^i(k_n, k_{n-1}, k_n) \geq 0$. Since $(k_n, k_{n-1}) \in \mathscr{D}^i$, a unique perfect forecast $k_n^e > 0$ exists. Since, under perfect foresight, the realized intermediation margin is non-negative, we have $e^i(k_{n+1}, k_n, k_{n+1}) \geq 0$. Hence, $(k_{n+1}, k_n) \in \mathscr{D}^i$. The proof by induction is now complete. □

Proof of Lemma 14. We establish the partial derivatives of ψ^{M} first. Differentiating (PFD$^{\mathrm{M}}$) and evaluating in a steady state $\boldsymbol{k}^{\mathrm{M}}$ yields

$$\begin{aligned}\frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}^{\mathrm{M}}) =& \frac{w'(k^{\mathrm{M}})\,\frac{\mathrm{d}D}{\mathrm{d}w}(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))) + f'(k^{\mathrm{M}})\,(1+\epsilon_2(k^{\mathrm{M}}))}{\frac{1}{p} - \frac{\partial D}{\partial r}(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\,f''(k^{\mathrm{M}})} \\ &- \frac{f''(k^{\mathrm{M}})\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\,D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\,[1 + \eta(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))]}{\frac{1}{p} - \frac{\partial D}{\partial r}(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\,f''(k^{\mathrm{M}})},\end{aligned} \tag{4.A.14}$$

and

$$\begin{aligned}\frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k}^{\mathrm{M}}) =& -\frac{w'(k^{\mathrm{M}})\,\frac{\partial r^{\mathrm{M}}}{\partial w}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\,D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\,[1 + \eta(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))]}{\frac{1}{p} - \frac{\partial D}{\partial r}(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\,f''(k^{\mathrm{M}})} \\ &- \frac{w'(k^{\mathrm{M}})\,r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\,\frac{\partial D}{\partial w}(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))}{\frac{1}{p} - \frac{\partial D}{\partial r}(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\,f''(k^{\mathrm{M}})}.\end{aligned} \tag{4.A.15}$$

By Assumption 8, we have $f' > 0$, $f'' < 0$, $w' > 0$, and $-1 < \epsilon_2 < 0$. In view of Lemma 19 (ii) and (iv), we can infer from (4.A.14) and (4.A.15) that

$$\frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}^{\mathrm{M}}) > 0, \quad \text{and} \quad \frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k}^{\mathrm{M}}) < 0. \tag{4.A.16}$$

The functional form of the Eigenvalues in (4.16) together with the properties in (4.A.16) then imply that

$$0 < \left|\lambda_2(\boldsymbol{k}^{\mathrm{M}})\right| \leq \left|\lambda_1(\boldsymbol{k}^{\mathrm{M}})\right|.$$

□

Proof of Proposition 15. The proof works in two steps. In Step 1, we show that the origin is a steady state if and only if $f(0) = 0$. In Step 2, we prove its instability.

Step 1. The origin is a steady state if and only if zero solves (4.17). For $k \to 0$, the l.h.s. in (4.17) converges to zero. Note that the r.h.s. in (4.17) is bounded from above by the production function, that is, for all $k \geq 0$, it lies below $f(k)$. Therefore, if $f(0) = 0$, then the r.h.s. in (4.17) converges to zero for $k \to 0$. Hence, $k^i = 0$ solves (4.17) if $f(0) = 0$. On the other hand, if $f(0) > 0$, then Assumption 8 (i) implies $w(0) > 0$. By the Inada condition imposed on v in Assumption 7, the r.h.s. in (4.17) then satisfies

$$\begin{aligned}
\lim_{k\to 0} I(w(k),f'(k)) &> 0 \qquad \text{for } i = \mathrm{N},\\
\lim_{k\to 0} D(w(k),pf'(k)) &> 0 \qquad \text{for } i = \mathrm{C},\\
\lim_{k\to 0} D(w(k),r^{\mathrm{M}}(w(k),f'(k))) &> 0 \qquad \text{for } i = \mathrm{M},
\end{aligned}$$

showing that zero *cannot* solve (4.17). It follows that the origin is a steady state if and only if $f(0) = 0$.

Step 2. For reasons of clarity, we consider each case $i \in \{\mathrm{N}, \mathrm{C}, \mathrm{M}\}$ separately. Consider the case **without financial intermediation** first, $i = \mathrm{N}$. Then $k^{\mathrm{N}} = 0$ is unstable if

$$\lim_{k\to 0} \frac{\mathrm{d}\psi^{\mathrm{N}}}{\mathrm{d}k}(k) > 1 \qquad \Longleftrightarrow \qquad \lim_{k\to 0} \frac{\mathrm{d}I}{\mathrm{d}k}(w(k),f'(k)) > \frac{1}{p}.$$

We next show that

$$\lim_{k\to 0} \frac{\mathrm{d}I}{\mathrm{d}k}(w(k),f'(k)) = \lim_{k\to 0} w'(k) = \infty > \frac{1}{p}. \tag{4.A.17}$$

To this end, observe that the first-order condition (4.2) implies

$$\frac{u'(w(k) - I(w(k),f'(k)))}{v'(f'(k)I(w(k),f'(k)))} = pf'(k). \tag{4.A.18}$$

Considering (4.A.18), the Inada conditions imposed on u, v, and f imply

$$I(w(k),f'(k)) \overset{k\to 0}{\longrightarrow} w(k).$$

Consequently, we can conclude that

$$\lim_{k\to 0} \frac{\mathrm{d}I}{\mathrm{d}k}(w(k),f'(k)) = \lim_{k\to 0} w'(k) = \infty,$$

where the last equality follows from Lemma 18. Hence, if the origin $k^{\mathrm{N}} = 0$ is a steady state, then it is unstable.

Consider the case of a **competitive banking industry** now, $i = \mathrm{C}$. Repeating the arguments presented above, the first-order condition (4.5) together with the Inada conditions imposed on u, v, and f implies

$$\lim_{k\to 0} \frac{\mathrm{d}D}{\mathrm{d}k}(w(k), pf'(k)) = \lim_{k\to 0} w'(k) = \infty > \frac{1}{p}. \tag{4.A.19}$$

Hence, $k^{\mathrm{C}} = 0$ is unstable.

Finally, we analyze the **monopolistic case**, $i = \mathrm{M}$. The origin $k^{\mathrm{M}} = 0$ is unstable if the Eigenvalue λ_1 in (4.16) satisfies

$$\lim_{k\to 0} |\lambda_1(\boldsymbol{k})| > 1. \tag{4.A.20}$$

Observe that (4.A.20) is satisfied if $\lim_{k\to 0} \partial \psi^{\mathrm{M}}(\boldsymbol{k})/\partial k_t = \infty$ holds, which is what we show next. Since the denominator in (4.A.14) is strictly positive, two cases can occur.

Case 1. Suppose the denominator in (4.A.14) is finite for $k \to 0$. Since all summands in the numerator of (4.A.14) are non-negative, it follows that

$$\lim_{k\to 0} \frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}) \geq \lim_{k\to 0} \frac{f'(k)\,(1+\epsilon_2(k))}{\frac{1}{p} - \frac{\partial D}{\partial r}(w(k), r^{\mathrm{M}}(w(k), f'(k)))\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(w(k), f'(k))f''(k)}. \tag{4.A.21}$$

By the Inada condition $\lim_{k\to 0} f'(k) = \infty$ and the fact that $-1 < \epsilon_2 < 0$ by Assumption 8 (ii), we can conclude that the limit on the r.h.s. of (4.A.21) is infinity, implying that $\lim_{k\to 0} \partial \psi^{\mathrm{M}}(\boldsymbol{k})/\partial k_t = \infty$.

Case 2. Suppose that the denominator in (4.A.14) converges to infinity for $k \to 0$. Since all summands in the numerator of (4.A.14) are non-negative, the following sufficient condition obtains:

$$\begin{aligned}
\lim_{k\to 0} \frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}) &\geq \lim_{k\to 0} \frac{-\frac{\partial D}{\partial r}(w(k), r^{\mathrm{M}}(w(k), f'(k)))\, r^{\mathrm{M}}(w(k), f'(k))\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(w(k), f'(k))f''(k)}{\frac{1}{p} - \frac{\partial D}{\partial r}(w(k), r^{\mathrm{M}}(w(k), f'(k)))\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(w(k), f'(k))f''(k)} \\
&= \lim_{k\to 0} \frac{r^{\mathrm{M}}(w(k), f'(k))}{1 - \left(p\,\frac{\partial D}{\partial r}(w(k), r^{\mathrm{M}}(w(k), f'(k)))\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(w(k), f'(k))f''(k)\right)^{-1}} \\
&= \lim_{k\to 0} r^{\mathrm{M}}(w(k), f'(k)).
\end{aligned}$$

By definition, $r^{\mathrm{M}}(w(k),f'(k)) \geq s(w(k),f'(k))$ for all $k \geq 0$. In addition, it follows from (4.7) together with the Inada condition $\lim_{k\to 0} f'(k) = \infty$ that $\lim_{k\to 0} s(w(k),f'(k)) = \infty$. Consequently, $\lim_{k\to 0} r^{\mathrm{M}}(w(k),f'(k)) = \infty$ and, therefore, $\lim_{k\to 0} \partial\psi^{\mathrm{M}}(\boldsymbol{k})/\partial k_t = \infty$.

We can conclude that, regardless of which case occurs, $\lim_{k\to 0} \partial\psi^{\mathrm{M}}(\boldsymbol{k})/\partial k_t = \infty$. Hence, the origin $k^{\mathrm{M}} = 0$ is indeed unstable. □

Proof of Theorem 5.

Part (i). We analyze each case $i \in \{\mathrm{N},\mathrm{C},\mathrm{M}\}$ separately again. Consider the case **without financial intermediation** first, $i = \mathrm{N}$. It follows from (4.A.17) that in the local neighborhood of zero, we have $I(w(k),f'(k)) > k/p$. Since $0 \leq I(w(k),f'(k)) \leq f(k)$ for all $k \geq 0$ while f is strictly concave and satisfies the Inada condition $\lim_{k\to 0} f'(k) = 0$, we can conclude that there exists a solution $k^{\mathrm{N}} > 0$ to (4.17). Uniqueness and asymptotic stability of $k^{\mathrm{N}} > 0$ is established next. To do so, we show that

$$\frac{\mathrm{d}\psi^{\mathrm{N}}}{\mathrm{d}k}(k^{\mathrm{N}}) < 1 \quad\Longleftrightarrow\quad \frac{\mathrm{d}I}{\mathrm{d}k}(w(k^{\mathrm{N}}),f'(k^{\mathrm{N}})) < \frac{1}{p}$$

holds for *any* given $k^{\mathrm{N}} > 0$. This property does not only imply stability, but also uniqueness since it ensures that $I(w(k),f'(k))$ has a unique positive intersection with k/p. Calculating the total differential of I and exploiting (4.17) yields

$$\begin{aligned} &\frac{\mathrm{d}I}{\mathrm{d}k}(w(k^{\mathrm{N}}),f'(k^{\mathrm{N}})) < \frac{1}{p} \\ \Longleftrightarrow\quad &\alpha_u \frac{f'(k^{\mathrm{N}})}{w'(k^{\mathrm{N}})}\left(\frac{pw'(k^{\mathrm{N}})k^{\mathrm{N}} - k^{\mathrm{N}}}{pw(k^{\mathrm{N}}) - k^{\mathrm{N}}}\right) + \alpha_v\left(1 - \frac{f'(k^{\mathrm{N}})}{w'(k^{\mathrm{N}})}\right) < 1\,. \end{aligned} \tag{4.A.22}$$

Since $pw(k^{\mathrm{N}}) - k^{\mathrm{N}} > 0$ and, in addition, $0 < w'(k^{\mathrm{N}})k^{\mathrm{N}} < w(k^{\mathrm{N}})$ by Lemma 18, a sufficient condition for (4.A.22) is

$$\frac{f'(k^{\mathrm{N}})}{w'(k^{\mathrm{N}})}(\alpha_u - \alpha_v) < 1 - \alpha_v. \tag{4.A.23}$$

By Assumption 8 (i), we have $f' > 0$ and $w' > 0$. Since, additionally, $0 < \alpha_u \leq \alpha_v < 1$ by Assumption 7 (ii), it follows that (4.A.23) is satisfied for *any* given $k^{\mathrm{N}} > 0$. Hence, $k^{\mathrm{N}} > 0$ is uniquely determined and asymptotically stable, globally on $\mathbb{R}_{++}$.

Consider the case of a **competitive banking industry** now, $i = \mathrm{C}$. Again, it follows from (4.A.19) that in the local neighborhood of zero, we have $D(w(k),pf'(k)) > k/p$. By the strict concavity of f and the Inada condition $\lim_{k\to 0} f'(k) = 0$, a solution

$k^{\mathrm{C}} > 0$ to (4.17) thus exists. We next prove that $k^{\mathrm{C}} > k^{\mathrm{N}}$ by contradiction. Assume for a moment that $k^{\mathrm{C}} \leq k^{\mathrm{N}}$. By the uniqueness of k^{N}, we must then have

$$pI(w(k^{\mathrm{C}}), f'(k^{\mathrm{C}})) \geq pD(w(k^{\mathrm{C}}), pf'(k^{\mathrm{C}})). \tag{4.A.24}$$

However, (4.A.24) contradicts Corollary 5. Hence, $k^{\mathrm{C}} > k^{\mathrm{N}} > 0$ must hold.

The uniqueness and asymptotic stability of $k^{\mathrm{C}} > 0$ obtains from the same arguments as presented above, namely that $0 < \alpha_u \leq \alpha_v < 1$ implies that

$$\frac{\mathrm{d}D}{\mathrm{d}k}(w(k^{\mathrm{C}}), pf'(k^{\mathrm{C}})) < \frac{1}{p}$$

holds for *any* given $k^{\mathrm{C}} > 0$.

Finally, consider the **monopolistic case**, $i = \mathrm{M}$. The existence of a positive steady state $k^{\mathrm{M}} > 0$ is implied by the instability of the origin, cf. Proposition 15. This is seen as follows. Formally, from (4.A.14), (4.A.15), and (4.16), it can be shown that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}k}\Big(D(w(k), r^{\mathrm{M}}(w(k), f'(k))) + e^{\mathrm{M}}(k,k,k)\Big) \gtreqless \frac{1}{p} \\ \iff \quad & \frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}) + \frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k}) \gtreqless 1 \\ \iff \quad & |\lambda_1(\boldsymbol{k})| \gtreqless 1, \end{aligned} \tag{4.A.25}$$

where the last relation in (4.A.25) is implied by

$$\begin{aligned} |\lambda_1(\boldsymbol{k})| &= \left| \tfrac{1}{2}\frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}) + \sqrt{\tfrac{1}{4}\left(\tfrac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k})\right)^2 + \tfrac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k})} \right| \\ &\gtreqless \left| \tfrac{1}{2}\left(1 - \frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k})\right) + \sqrt{\tfrac{1}{4}\left(1 - \frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k})\right)^2 + \tfrac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k})} \right| = |1| = 1. \end{aligned}$$

In the proof of Proposition 15, we have shown that $\lim_{k\to 0} |\lambda_1(\boldsymbol{k})| = \infty$. Considering (4.A.25), we can thus infer that

$$\lim_{k\to 0} |\lambda_1(\boldsymbol{k})| = \infty \quad \Longrightarrow \quad \lim_{k\to 0} \frac{\mathrm{d}}{\mathrm{d}k}\Big(D(w(k), r^{\mathrm{M}}(w(k), f'(k))) + e^{\mathrm{M}}(k,k,k)\Big) > \frac{1}{p}.$$

This shows that in the local neighborhood of zero, we have

$$D(w(k), r^{\mathrm{M}}(w(k), f'(k))) + e^{\mathrm{M}}(k,k,k) > \frac{k}{p}.$$

Since, in addition, $0 \le D(w(k), r^{\mathrm{M}}(w(k), f'(k))) + e^{\mathrm{M}}(k,k,k) \le f(k)$ for all $k \ge 0$ while f is strictly concave and satisfies $\lim_{k\to 0} f'(k) = 0$, there must exist a solution $k^{\mathrm{M}} > 0$ to (4.17).

Next, we prove that $k^{\mathrm{M}} > k^{\mathrm{N}}$ by contradiction. Suppose that $k^{\mathrm{N}} \ge k^{\mathrm{M}}$. However, by Corollary 5 and the fact that monopolistic steady-state equity must be strictly positive, this yields a contradiction:

$$p\,I(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})) \ \ge\ p\left(D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))) + e^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}, k^{\mathrm{M}})\right) \ >\ p\,D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(\cdot))$$

$$\iff \quad 0 \ \ge\ I(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})) - D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))) \ \ge\ e^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}, k^{\mathrm{M}}) \ >\ 0.$$

Hence, $k^{\mathrm{M}} > k^{\mathrm{N}} > 0$ must hold. The uniqueness and asymptotic stability of k^{M} are shown as follows. Again, the property $0 < \alpha_u \le \alpha_v < 1$ stated in Assumption 7 (ii) implies that

$$\frac{\mathrm{d}}{\mathrm{d}k}\left(D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))) + e^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}, k^{\mathrm{M}})\right) \ <\ \frac{1}{p} \tag{4.A.26}$$

holds for *any* given $k^{\mathrm{M}} > 0$. Hence, $k^{\mathrm{M}} > 0$ is uniquely determined. Moreover, (4.A.25) together with (4.A.26) implies that $\left|\lambda_1(\boldsymbol{k}^{\mathrm{M}})\right| < 1$. We can then conclude from Lemma 14 that

$$0 \ <\ \left|\lambda_2(\boldsymbol{k}^{\mathrm{M}})\right| \ \le\ \left|\lambda_1(\boldsymbol{k}^{\mathrm{M}})\right| \ <\ 1,$$

showing that $k^{\mathrm{M}} > 0$ is also asymptotically stable, globally on $\mathbb{R}_{++}$.

Part (ii). First of all, note that $\epsilon_1 \to 1$ implies $w(k) \to 0$ and $f'(k)k \to f(k)$. Since $0 \le I(w(k), f'(k)) < D(w(k), pf'(k)) \le w(k)$ for all $k \ge 0$, we can directly deduce from (4.17) that $\epsilon_1 \to 1$ implies $k^{\mathrm{N}} \to 0$ and $k^{\mathrm{C}} \to 0$. However, $k^{\mathrm{N}} < k^{\mathrm{C}}$ must still hold by Theorem 5 (i). Since $k^{\mathrm{M}} > f'^{-1}(1/p) > 0$ by Lemma 20, it follows that $0 < k^{\mathrm{N}} < k^{\mathrm{C}} < k^{\mathrm{M}}$ if ϵ_1 is sufficiently large.

On the other hand, $\epsilon_1 \to 0$ implies $w(k) \to f(k)$ and $f'(k)k \to 0$, such that $k^{\mathrm{M}} > 0$ is determined by

$$\frac{k^{\mathrm{M}}}{p} \ =\ D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\left(1 - r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\right). \tag{4.A.27}$$

Combining Corollary 5 and (4.A.27) implies the inequality

$$\begin{aligned} \frac{k^{\mathrm{M}}}{p} \ &=\ D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\left(1 - r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\right) \\ &<\ D(w(k^{\mathrm{M}}), pf'(k^{\mathrm{M}})) \ <\ f(k^{\mathrm{M}}), \end{aligned}$$

which demonstrates that k^{M} cannot be a steady state of ($\mathrm{PFD^C}$). Instead, the strict concavity of f together with the Inada condition $\lim_{k\to 0} f'(k) = 0$ implies that there must exist a steady state $k^{\mathrm{C}} > 0$ of ($\mathrm{PFD^C}$) that satisfies $k^{\mathrm{C}} > k^{\mathrm{M}}$. Hence, $0 < k^{\mathrm{N}} < k^{\mathrm{M}} < k^{\mathrm{C}}$ if ϵ_1 is sufficiently small. □

Proof of Corollary 6. Concerning the first inequality, two cases can occur. First, if $r^{\mathrm{M}}(w(k_t), f'(k_t^e)) = b(w(k_t), f'(k_t^e)) \geq s(w(k_t), f'(k_t^e))$, then (PC) is binding and the relation holds with equality by (4.9). Second, if $r^{\mathrm{M}}(w(k_t), f'(k_t^e)) = s(w(k_t), f'(k_t^e)) > b(w(k_t), f'(k_t^e))$, then (4.A.8) implies that the inequality is strict. The second inequality is implied by (4.A.8) and the fact that, for any given $k_t > 0$ and $k_t^e > 0$, $r^{\mathrm{C}}(w(k_t), f'(k_t^e)) > r^{\mathrm{M}}(w(k_t), f'(k_t^e))$. □

Proof of Lemma 15. In order to determine the stationary feasible allocation $(\bar{k}, \bar{c}^1, \bar{c}^2)$ with the highest possible level of welfare, we solve the social-planner problem

$$\max_{\bar{c}^1, \bar{c}^2, \bar{k} \geq 0} \; u(\bar{c}^1) + v(\bar{c}^2) \quad \text{s.t.} \quad \bar{c}^1 + \bar{c}^2 \;=\; \phi(\bar{k}). \tag{4.A.28}$$

Inserting the feasibility constraint in (4.A.28) into the objective function yields

$$\max_{\bar{c}^2, \bar{k} \geq 0} \; u(\phi(\bar{k}) - \bar{c}^2) + v(\bar{c}^2). \tag{4.A.29}$$

The first-order conditions for the unconstrained problem (4.A.29) are

$$u'(\phi(\bar{k}) - \bar{c}^2) \;=\; v'(\bar{c}^2) \tag{4.A.30}$$

$$u'(\phi(\bar{k}) - \bar{c}^2)\, \phi'(\bar{k}) \;=\; 0. \tag{4.A.31}$$

Since $u' > 0$, Condition (4.A.31) is satisfied if and only if $\phi'(\bar{k}) = 0$, which, in turn, is satisfied if and only if $\bar{k} = k_G$, with k_G as defined in Lemma 15. Given k_G, the optimal consumption plan (c_G^1, c_G^2) is then uniquely determined by (4.A.30) together with the constraint in (4.A.28). □

Proof of Theorem 6. *Part (i)*. We first show that $\mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}},k^{\mathrm{C}})$ is maximal if and only if $k^{\mathrm{C}}=k_G$, with k_G as defined in Lemma 15. This is seen as follows. The level of welfare corresponding to k^{C} amounts to

$$\mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}},k^{\mathrm{C}}) \;=\; u\big(w(k^{\mathrm{C}})-D(w(k^{\mathrm{C}}),pf'(k^{\mathrm{C}}))\big)+v\big(pf'(k^{\mathrm{C}})\,D(w(k^{\mathrm{C}}),pf'(k^{\mathrm{C}}))\big). \quad (4.A.32)$$

Using the Envelope Theorem, (4.17), and $w'(k^{\mathrm{C}})=-f''(k^{\mathrm{C}})k^{\mathrm{C}}$, the first-order condition for a maximum of (4.A.32) reads

$$\frac{u'\big(w(k^{\mathrm{C}})-D(w(k^{\mathrm{C}}),pf'(k^{\mathrm{C}}))\big)}{v'\big(pf'(k^{\mathrm{C}})\,D(w(k^{\mathrm{C}}),pf'(k^{\mathrm{C}}))\big)} \;=\; 1. \quad (4.A.33)$$

By Condition (4.5), we have

$$\frac{u'\big(w(k^{\mathrm{C}})-D(w(k^{\mathrm{C}}),pf'(k^{\mathrm{C}}))\big)}{v'\big(pf'(k^{\mathrm{C}})\,D(w(k^{\mathrm{C}}),pf'(k^{\mathrm{C}}))\big)} \;=\; pf'(k^{\mathrm{C}}). \quad (4.A.34)$$

Inserting (4.A.34) into the first-order condition (4.A.33) yields

$$pf'(k^{\mathrm{C}}) \;=\; 1.$$

Since, by Lemma 15, k_G solves $pf'(k_G)=1$ and $f''<0$, we can infer that the unique maximizer of $\mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}},k^{\mathrm{C}})$ is $k^{\mathrm{C}}=k_G$.

Next, let an arbitrary steady state $0<k^{\mathrm{C}}\le k_G$ be given. By Theorem 5, it holds that $0<k^{\mathrm{N}}<k^{\mathrm{C}}\le k_G$ and, by Corollary 6, that

$$\mathscr{W}^{\mathrm{N}}(k^{\mathrm{N}},k^{\mathrm{N}}) \;<\; \mathscr{W}^{\mathrm{C}}(k^{\mathrm{N}},k^{\mathrm{N}}). \quad (4.A.35)$$

As elaborated above, $\mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}},k^{\mathrm{C}})$ attains a unique stationary point at $k^{\mathrm{C}}=k_G$, implying that

$$\frac{\mathrm{d}\mathscr{W}^{\mathrm{C}}}{\mathrm{d}k}(k,k) \;>\; 0 \quad \text{for all } 0 \;<\; k \;<\; k_G. \quad (4.A.36)$$

It now follows from (4.A.35) together with (4.A.36) that

$$\mathscr{W}^{\mathrm{N}}(k^{\mathrm{N}},k^{\mathrm{N}}) \;<\; \mathscr{W}^{\mathrm{C}}(k^{\mathrm{N}},k^{\mathrm{N}}) \;<\; \mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}},k^{\mathrm{C}}).$$

Part (ii). In the proof of Theorem 5 (ii), we have shown that $\epsilon_1\to 1$ implies $w(k^{\mathrm{N}}),w(k^{\mathrm{C}})\to 0$ and, in addition, that $k^{\mathrm{N}},k^{\mathrm{C}}\to 0$. Therefore, we can deduce that $\epsilon_1\to 1$ implies

$$\mathscr{W}^{\mathrm{N}}(k^{\mathrm{N}},k^{\mathrm{N}}) \;\to\; u(0)+p\,v(0), \quad \text{and} \quad \mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}},k^{\mathrm{C}}) \;\to\; u(0)+v(0).$$

Since $k^{\mathrm{M}} > f'^{-1}(1/p)$ by Lemma 20, we have $w(k^{\mathrm{M}}) > w(k^{\mathrm{N}}), w(k^{\mathrm{C}})$. It then follows from the monotonicity of u and v that

$$\mathscr{W}^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}) \; > \; u(0) + v(0) \; > \; u(0) + p\,v(0).$$

Since, for $\epsilon \to 1$, we have $k^{\mathrm{C}} < k_G$, Theorem 5 (i) implies

$$\mathscr{W}^{\mathrm{N}}(k^{\mathrm{N}}, k^{\mathrm{N}}) \; < \; \mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}}, k^{\mathrm{C}}).$$

Summing up, we see that $\epsilon_1 \to 1$ yields

$$\mathscr{W}^{\mathrm{N}}(k^{\mathrm{N}}, k^{\mathrm{N}}) \; < \; \mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}}, k^{\mathrm{C}}) \; < \; \mathscr{W}^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}).$$

The converse case is considered next. Let $i \in \{\mathrm{C}, \mathrm{M}\}$. If $\epsilon_1 \to 0$, then $f'(k^i)k^i \to 0$ and $w(k^i) \to f(k^i)$. Since $e^i(k^i, k^i, k^i) \geq 0$, we can then deduce from the bank's balance sheet equation

$$f(k^i) \; = \; w(k) + e^i(k^i, k^i, k^i) + r^i(w(k^i), f'(k^i))\, D(w(k^i), r^i(w(k^i), f'(k^i)))$$

that $\epsilon_1 \to 0$ implies $r^i(w(k^i), f'(k^i))\, D(w(k^i), r^i(w(k^i), f'(k^i))) \to 0$. This observation, in turn, yields

$$\mathscr{W}^{\mathrm{C}}(k^{\mathrm{C}}, k^{\mathrm{C}}) \; \to \; u(w(k^{\mathrm{C}})) + v(0), \quad \text{and} \quad \mathscr{W}^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}) \; \to \; u(w(k^{\mathrm{M}})) + v(0). \quad (4.A.37)$$

By Theorem 5 (ii), $\epsilon_1 \to 0$ also causes that $k^{\mathrm{C}} > k^{\mathrm{M}}$. Accordingly, the wage incomes in (4.A.37) must satisfy $w(k^{\mathrm{C}}) > w(k^{\mathrm{M}})$ because $w' > 0$ by Assumption 8 (i). Since u is strictly increasing by Assumption 7 (i), we can infer from (4.A.37) that $u(w(k^{\mathrm{C}})) + v(0) > u(w(k^{\mathrm{M}})) + v(0)$. □

Proof of Proposition 16. Recall that $k^{\mathrm{M}} > 0$ is asymptotically stable by Theorem 5 and, therefore, hyperbolic. The HARTMAN-GROBMAN THEOREM thus implies that the linearized dynamics (4.14) is qualitatively equivalent to the non-linear dynamics (PFD$^{\mathrm{M}}$) in the local neighborhood of $k^{\mathrm{M}} > 0$. As is well-known, the dynamics (4.14) is monotonic if the Eigenvalues (4.16) are positive real numbers. We first show that the Eigenvalues are real. From the functional form in (4.16), it follows that the Eigenvalues are real if and only if

$$\frac{1}{4}\left(\frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}^{\mathrm{M}})\right)^2 + \frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k}^{\mathrm{M}}) \; \geq \; 0.$$

By the binomial formulas, a sufficient condition is that

$$\frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}^{\mathrm{M}}) > 1 - \frac{1}{pN(\boldsymbol{k}^{\mathrm{M}})} - \frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k}^{\mathrm{M}}) \quad \Longrightarrow \quad \tfrac{1}{4}\left(\tfrac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}^{\mathrm{M}})\right)^2 + \tfrac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k}^{\mathrm{M}}) \geq 0,$$

where $N(\boldsymbol{k}^{\mathrm{M}})$ denotes the numerator in (4.A.14). Moreover, it is easy to derive from (4.A.14) and (4.A.15) that

$$\begin{aligned} & \frac{\partial \psi^{\mathrm{M}}}{\partial k_t}(\boldsymbol{k}^{\mathrm{M}}) > 1 - \frac{1}{pN(k^{\mathrm{M}})} - \frac{\partial \psi^{\mathrm{M}}}{\partial k_{t-1}}(\boldsymbol{k}^{\mathrm{M}}) \\ \Longleftrightarrow \quad & \frac{\mathrm{d}}{\mathrm{d}k}\Big(D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))) + e^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}, k^{\mathrm{M}})\Big) > 0. \end{aligned} \tag{4.A.38}$$

Calculating the total differential in (4.A.38) yields

$$\begin{aligned} & \frac{\mathrm{d}}{\mathrm{d}k}\Big(D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))) + e^{\mathrm{M}}(k^{\mathrm{M}}, k^{\mathrm{M}}, k^{\mathrm{M}})\Big) \\ = & f'(k^{\mathrm{M}})\left(1 + \epsilon_2(k^{\mathrm{M}})\right) + \frac{\mathrm{d}}{\mathrm{d}k}\Big(D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\left(1 - r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\right)\Big). \end{aligned}$$

Since, by Assumption 8, $f' > 0$ and $-1 < \epsilon_2 < 0$, the first summand is strictly positive. We next show that the second summand is also strictly positive. By Lemma 20, the steady state $k^{\mathrm{M}} > 0$ satisfies $k^{\mathrm{M}} > k_G$. Since stationary total consumption per capita ϕ is a unimodal function which attains its maximum at k_G, we must have $\phi'(k^{\mathrm{M}}) < 0$. From the identity

$$\phi(k) = w(k) - D(w(k), r^{\mathrm{M}}(w(k), f'(k)))\left(1 - r^{\mathrm{M}}(w(k), f'(k))\right)$$

and $\phi'(k^{\mathrm{M}}) < 0$, we can then infer that

$$0 < w'(k^{\mathrm{M}}) < \frac{\mathrm{d}}{\mathrm{d}k}\Big(D(w(k^{\mathrm{M}}), r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}})))\left(1 - r^{\mathrm{M}}(w(k^{\mathrm{M}}), f'(k^{\mathrm{M}}))\right)\Big).$$

Hence, the second summand is also positive, implying that $k^{\mathrm{M}} > 0$ satisfies (4.A.38) and that the Eigenvalues in (4.16) are real. The fact that the Eigenvalues are positive follows directly from (4.A.16) and the functional form in (4.16). □

Proof of Lemma 16. Let $w_t > 0$, $\rho_t^e > 0$, and $\vartheta_t^e \geq 0$ be arbitrary but fixed. The solution to Problem (4.21) is either determined by a solution $s(w_t, \rho_t^e, \vartheta_t^e)$ to the first-order condition

$$\frac{p\rho_t^e}{r} - 1 - \frac{1}{\eta(w_t, r, \vartheta_t^e)} = 0 \tag{4.A.39}$$

corresponding to the relaxed maximization problem without the participation constraint or by a solution $b(w_t, \rho_t^e, \vartheta_t^e)$ to the binding participation constraint

$$u(w_t - D(w_t, r, \vartheta_t^e)) + v(rD(w_t, r, \vartheta_t^e) + \vartheta_t^e) = U_{\text{res}}(w_t, \rho_t^e). \tag{4.A.40}$$

Consider Condition (4.A.39) first. For $r \to 0$, the l.h.s. in (4.A.39) converges to ∞ because $\lim_{r\to 0} \eta(w_t, r, \vartheta_t^e) = \infty$. By contrast, for $r \to p\rho_t^e$, the l.h.s. in (4.A.39) approaches $-1/\eta(w_t, p\rho_t^e, \vartheta_t^e) < 0$. Moreover, the l.h.s. in (4.A.39) is strictly decreasing by Part (v) of Lemma 19,

$$-\frac{p\rho_t^e}{r^2} + \frac{1}{\eta(w_t, r, \vartheta_t^e)^2} \frac{\partial \eta}{\partial r}(w_t, r, \vartheta_t^e) < 0.$$

Hence, a uniquely determined solution $0 < s(w_t, \rho_t^e, \vartheta_t^e) < p\rho_t^e$ to (4.A.39) exists.

Next, consider Condition (4.A.40). For $r \to 0$, the l.h.s. in (4.A.40) converges to $u(w_t) + v(\vartheta_t^e)$. By contrast, for $r \to p\rho_t^e$, it approaches

$$\begin{aligned}
& u(w_t - D(w_t, p\rho_t^e, \vartheta_t^e)) + v(p\rho_t^e D(w_t, p\rho_t^e, \vartheta_t^e) + \vartheta_t^e) \\
\geq \; & u(w_t - I(w_t, \rho_t^e)) + v(p\rho_t^e I(w_t, \rho_t^e) + \vartheta_t^e) \\
\geq \; & u(w_t - I(w_t, \rho_t^e)) + v(p\rho_t^e I(w_t, \rho_t^e)) \\
> \; & U_{\text{res}}(w_t, \rho_t^e).
\end{aligned}$$

In view of the fact that the l.h.s. in (4.A.40) is strictly increasing in r, two cases must be distinguished.

Case 1. Suppose that $u(w_t) + v(\vartheta_t^e) < U_{\text{res}}(w_t, \rho_t^e)$. Then the Intermediate Value Theorem implies a uniquely determined solution $0 < b(w_t, \rho_t^e, \vartheta_t^e) < p\rho_t^e$ to (4.A.40).

Case 2. Suppose that $u(w_t) + v(\vartheta_t^e) \geq U_{\text{res}}(w_t, \rho_t^e)$. In this case, *no* interior solution to (4.A.40) exists. Instead, the participation constraint is slack for *any* positive deposit rate, and, in particular, for $s(w_t, \rho_t^e, \vartheta_t^e)$. Thus, we can conclude that the unique solution to Problem (4.21) is given by

$$r_t^{\text{M}} = r^{\text{M}}(w_t, \rho_t^e, \vartheta_t^e) = \begin{cases} \max\left\{b(w_t, \rho_t^e, \vartheta_t^e), s(w_t, \rho_t^e, \vartheta_t^e)\right\} & \text{if } u(w_t) + v(\vartheta_t^e) < U_{\text{res}}(w_t, \rho_t^e), \\ s(w_t, \rho_t^e, \vartheta_t^e) & \text{otherwise.} \end{cases}$$

□

Proof of Lemma 17. We analyze the forecast for the dividend payment first. To this end, let the correct forecast k_t^e for k_{t+1} be given. Given the correct forecasting k_t^e and the current capital-labor ratio k_t, a correct forecast $\vartheta_t^e \geq 0$ that is consistent with k_t^e is then determined by a solution to

$$\vartheta^e \;=\; \vartheta^{\mathrm{M}}(k_t^e, k_t, k_t^e, \vartheta^e) \;=\; \mu\, e^{\mathrm{M}}(k_t^e, k_t, k_t^e, \vartheta^e). \tag{4.A.41}$$

For $\vartheta^e \to 0$, the l.h.s. in Condition (4.A.41) converges to zero, whereas its r.h.s. converges to $\mu\, e^{\mathrm{M}}(k_t^e, k_t, k_t^e, 0)$. For $\vartheta^e \to \infty$, the l.h.s. in (4.A.41) converges to infinity, whereas the r.h.s. converges to $\mu\, e^{\mathrm{M}}(k_t^e, k_t, k_t^e, \infty) < f(k_t) < \infty$. Evoking the Intermediate Value Theorem, a solution $0 \leq \vartheta_t^e < \infty$ to (4.A.41) exists if $(k_t^e, k_t) \in \mathscr{B}$, where

$$\mathscr{B} \;:=\; \Big\{(x,y) \;\in\; \mathbb{R}_+^2 \mid e^{\mathrm{M}}(x,y,x,0) \;\geq\; 0\Big\}.$$

We next show that at all solutions ϑ_t^e to (4.A.41), the slope of the r.h.s. in (4.A.41) is strictly smaller than the slope of the l.h.s., which then implies that the solution ϑ_t^e is, in fact, uniquely determined. By straightforward differentiation, we can show if evaluated at ϑ_t^e, the slope of the r.h.s. is strictly smaller than the slope of the l.h.s. if and only if ϑ_t^e satisfies

$$1 \;>\; -\mu \frac{\mathrm{d}}{\mathrm{d}\vartheta^e}\Big(r^{\mathrm{M}}(w(k_t), f'(k_t^e), \vartheta_t^e)\, D(w(k_t), r^{\mathrm{M}}(w(k_t), f'(k_t^e), \vartheta_t^e), \vartheta_t^e)\Big). \tag{4.A.42}$$

Since $\mu \in [0,1)$, Condition (4.A.42) is satisfied if ϑ_t^e satisfies

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta^e}\Big(r^{\mathrm{M}}(w(k_t), f'(k_t^e), \vartheta_t^e)\, D(w(k_t), r^{\mathrm{M}}(w(k_t), f'(k_t^e), \vartheta_t^e), \vartheta_t^e)\Big) \;>\; -1. \tag{4.A.43}$$

We next show that (4.A.43) indeed holds. Given the two correct forecasts (k_t^e, ϑ_t^e), the bank's balance sheet satisfies

$$\begin{aligned} & r^{\mathrm{M}}(w(k_t), f'(k_t^e), \vartheta_t^e)\, D(w(k_t), r^{\mathrm{M}}(w(k_t), f'(k_t^e), \vartheta_t^e), \vartheta_t^e) \\ &= f'(k_{t+1})k_{t+1} - \vartheta_{t+1} - (1-\mu)e_{t+1} \\ &= f'(k_t^e)k_t^e - \vartheta_t^e - (1-\mu)e_{t+1}. \end{aligned}$$

Differentiation then yields

$$\begin{aligned} & \frac{\mathrm{d}}{\mathrm{d}\vartheta^e}\Big(r^{\mathrm{M}}(w(k_t), f'(k_t^e), \vartheta_t^e)\, D(w(k_t), r^{\mathrm{M}}(w(k_t), f'(k_t^e), \vartheta_t^e), \vartheta_t^e)\Big) \\ = & \frac{\mathrm{d}}{\mathrm{d}\vartheta^e}\Big(f'(k_t^e)k_t^e - \vartheta_t^e - (1-\mu)e_{t+1}\Big) \;=\; -1 - (1-\mu)\frac{\mathrm{d}e_{t+1}}{\mathrm{d}\vartheta^e}\Big|_{\vartheta^e = \vartheta_t^e} \;>\; -1. \end{aligned} \tag{4.A.44}$$

Hence, (4.A.43) holds, implying that ϑ_t^e satisfies (4.A.42), i.e., it is uniquely determined. Accordingly, we may define the function $\tilde{\psi} : \mathscr{B} \to \mathbb{R}_+$, such that $\vartheta_t^e = \tilde{\psi}(k_t^e, k_t)$ is a correct forecast for ϑ_{t+1}, given the correct forecast k_t^e for k_{t+1} and the current capital-labor ratio k_t.

As a next step, we show that the correct forecast k_t^e for k_{t+1} exists and is uniquely determined. Let $(k_t, k_{t-1}) \in \mathscr{C}$ be given and, importantly, observe that $\mathscr{C} \subset \mathscr{B}$. The correct forecast k_t^e is then determined by a solution to

$$\begin{aligned} &\frac{k^e}{p} - D(w(k_t), r^{\mathrm{M}}(w(k_t), f'(k^e), \tilde{\psi}(k^e, k_t)), \tilde{\psi}(k^e, k_t)) \\ &= (1-\mu)\, e^{\mathrm{M}}(k_t, k_{t-1}, k_t, \tilde{\psi}(k_t, k_{t-1})). \end{aligned} \tag{4.A.45}$$

By construction of $\mathscr{C}$, the Intermediate Value Theorem yields the existence of a solution $0 < k_t^e < \infty$ to (4.A.45). Next, we show that at any solution k_t^e to (4.A.45), the l.h.s. in (4.A.45) is strictly increasing, implying that k_t^e is uniquely determined. Differentiation of the l.h.s. in (4.A.45) with respect to k^e yields

$$\underbrace{\frac{1}{p} - \frac{\partial D}{\partial r}(\cdot)\,\frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(\cdot) f''(k^e)}_{>0} - \underbrace{\tfrac{\mathrm{d}D}{\mathrm{d}\vartheta^e}(\cdot)}_{<0}\,\frac{\partial \tilde{\psi}}{\partial k^e}(k^e, k_t). \tag{4.A.46}$$

Observe that the derivative (4.A.46) is strictly positive if k_t^e satisfies

$$\frac{\partial \tilde{\psi}}{\partial k^e}(k_t^e, k_t) \;>\; 0. \tag{4.A.47}$$

Implicit differentiation of (4.A.41) with respect to k^e and evaluating at k_t^e yields

$$\frac{\partial \tilde{\psi}}{\partial k^e}(k_t^e, k_t) = \frac{f'(k_t^e)\,(1+\epsilon_2(k_t^e)) - f''(k_t^e)\, D(\cdot)\, \frac{\partial r^{\mathrm{M}}}{\partial \rho^e}(\cdot)\,[1+\eta(\cdot)]}{\frac{1}{\mu} + \frac{\mathrm{d}}{\mathrm{d}\vartheta^e}\Big(r^{\mathrm{M}}(w(k_t), f'(k_t^e), \tilde{\psi}(k_t^e, k_t))\, D\Big(w(k_t), r^{\mathrm{M}}(w(k_t), f'(k_t^e), \tilde{\psi}(k_t^e, k_t)), \tilde{\psi}(k_t^e, k_t)\Big)\Big)}. \tag{4.A.48}$$

By Assumption 8, $f' > 0$, $f'' < 0$, and $-1 < \epsilon_2 < 0$. Hence, the numerator in (4.A.48) is strictly positive. It follows from (4.A.44) that the denominator in (4.A.48) is also strictly positive, implying that (4.A.47) holds. Consequently, the derivative (4.A.46) is strictly positive and k_t^e is uniquely determined.

Summing up, we have proven that for any given $(k_t, k_{t-1}) \in \mathscr{C}$, there exists a uniquely determined pair of perfect forecasts (k_t^e, ϑ_t^e) in period t. □

4.A.3 Remark on the Welfare Measure

In the literature on OLG models, it is common to conduct the welfare analysis along the entire growth path of an economy (De La Croix & Michel, 2002). The welfare measure is then the sum of the discounted lifetime utilities of all generations, which, in our notation, reads

$$\sum_{t=0}^{\infty} \gamma^t \mathscr{W}^i(k_t, k_{t+1}),$$

where $0 < \gamma < 1$ is a time-discount factor. There is a dispute in economic literature whether the well-being of future generations should be discounted (Mas-Colell et al., 1995). In this article, we follow the side of the debate that refrains from discounting the welfare of future generations and thus set $\gamma = 1$. As we will demonstrate next, a comparison of the steady-state welfare levels, as done by Theorem 6, then leads to the same result as when analyzing the welfare along an entire growth path. This is seen as follows. By Theorem 5, all growth paths $\{k_t\}_{t=0}^{\infty}$ converge to asymptotically stable steady states $k^i > 0$. As a consequence,

$$\sum_{t=0}^{\infty} \mathscr{W}^i(k_t, k_{t+1}) - \sum_{t=0}^{\infty} \mathscr{W}^j(k_t, k_{t+1}) \ > \ 0, \quad i,j \in \{\mathrm{N}, \mathrm{C}, \mathrm{M}\}, \, i \neq j$$

if and only if the asymptotically stable steady states satisfy

$$\mathscr{W}^i(k^i, k^i) - \mathscr{W}^j(k^j, k^j) \ > \ 0.$$

This observation also holds true in the presence of discounting if γ is sufficiently large. However, it should be noted that this article's focus is not on constructing an appropriate welfare measure, but to take a reasonable criterion and investigate the effect of financial intermediation and banking competition on welfare as defined by that criterion.

References

Allen, F., & Gale, D. (1997). Financial markets, intermediaries, and intertemporal smoothing. *Journal of Political Economy*, *105*(3), 523–546. [153]

Allen, F., & Gale, D. (2000). *Comparing financial systems*. MIT Press. [155]

Aziakpono, M. J. (2011). Financial development and economic growth: Theory and a survey of evidence. *Studies in Economics and Econometrics*, *35*(1), 15–44. [154]

Banerji, S., Bhattacharya, J., & Van Long, N. (2004). Can financial intermediation induce endogenous fluctuations. *Journal of Economic Dynamics and Control*, *28*(11), 2215–2238. [153, 174]

Beck, T., Demirgüç-Kunt, A., & Maksimovic, V. (2004). Bank competition and access to finance: International evidence. *Journal of Money, Credit and Banking*, *36*(3), 627–648. [153, 180]

Beck, T., Levine, R., & Loayza, N. (2000). Finance and the sources of growth. *Journal of Financial Economics*, *58*(1-2), 261–300. [154]

Bencivenga, V. R., & Smith, B. D. (1991). Financial intermediation and endogenous growth. *Review of Economic Studies*, *58*(2), 195–209. [154]

Böhm, V., & Wenzelburger, J. (1999). Expectations, forecasting, and perfect foresight: A dynamical systems approach. *Macroeconomic Dynamics*, *3*(2), 167–186. [154, 165]

Boyd, J. H., & De Nicolo, G. (2005). The theory of bank risk taking and competition revisited. *Journal of Finance*, *60*(3), 1329–1343. [155, 175]

Cetorelli, N. (1997). *The role of credit market competition on lending strategies and on capital accumulation* (tech. rep.). Citeseer. [152, 155]

Cetorelli, N., & Gambera, M. (2001). Banking market structure, financial dependence and growth: International evidence from industry data. *Journal of Finance*, *56*(2), 617–648. [153, 179, 180]

Cetorelli, N., & Peretto, P. F. (2000). Oligopoly banking and capital accumulation. *Working paper. Available at SSRN 254343*. [152, 155]

Cetorelli, N., & Peretto, P. F. (2012). Credit quantity and credit quality: Bank competition and capital accumulation. *Journal of Economic Theory*, *147*(3), 967–998. [152, 155]

Coccorese, P. (2017). Banking competition and economic growth. In J. A. Bikker & L. Spierdijk (Eds.), *Handbook of Competition in Banking and Finance* (pp. 230–263). Edward Elgar Publishing. [152]

De La Croix, D., & Michel, P. (2002). *A theory of economic growth: Dynamics and policy in overlapping generations*. Cambridge University Press. [176, 204]

Deidda, L., & Fattouh, B. (2005). Concentration in the banking industry and economic growth. *Macroeconomic Dynamics*, *9*(2), 198–219. [153, 180]

Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, *91*(3), 401–419. [154, 158]

Diamond, P. A. (1965). National debt in a neoclassical growth model. *American Economic Review*, *55*(5), 1126–1150. [170, 174]

Freixas, X., & Rochet, J.-C. (2008). *Microeconomics of banking*. MIT Press. [153, 159, 162]

Gersbach, H., & Wenzelburger, J. (2003). The workout of banking crises: A macroeconomic perspective. *CESifo Economic Studies*, *49*(2), 233–258. [161]

Gersbach, H., & Wenzelburger, J. (2008). Do risk premia protect against banking crises? *Macroeconomic Dynamics*, *12*(S1), 100–111. [161]

Gersbach, H., & Wenzelburger, J. (2012). Interest-rate policies and stability of banking systems. In J. Chadha & S. Holly (Eds.), *Lessons for monetary policy from the financial crisis* (pp. 71–107). Cambridge University Press. [161]

Goldsmith, R. W. (1969). *Financial structure and development.* Yale University Press. [154]

Grandmont, J.-M. (1985). On endogenous competitive business cycles. *Econometrica*, *53*(5), 995–1045. [165]

Greenwood, J., & Jovanovic, B. (1990). Financial development, growth, and the distribution of income. *Journal of Political Economy*, *98*(5, Part 1), 1076–1107. [154]

Greenwood, J., & Smith, B. D. (1997). Financial markets in development, and the development of financial markets. *Journal of Economic Dynamics and Control*, *21*(1), 145–181. [154]

Grobman, D. M. (1959). Homeomorphism of systems of differential equations. *Doklady Akademii Nauk SSSR*, *128*(5), 880–881. [167]

Gurley, J. G., & Shaw, E. S. (1955). Financial aspects of economic development. *American Economic Review*, *45*(4), 515–538. [154]

Guzman, M. G. (2000). Bank structure, capital accumulation and growth: A simple macroeconomic model. *Economic Theory*, *16*(2), 421–455. [152, 155]

Hamada, K., Kaneko, A., & Yanagihara, M. (2018). Oligopolistic competition in the banking market and economic growth. *Economic Modelling*, *68*, 239–248. [152, 155]

Hartman, P. (1960). A lemma in the theory of structural stability of differential equations. *Proceedings of the American Mathematical Society*, *11*(4), 610–620. [167]

Hellmann, T. F., Murdock, K. C., & Stiglitz, J. E. (2000). Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review*, *90*(1), 147–165. [155]

Hellwig, M. (1991). Banking, financial intermediation and corporate finance. In A. Giovanni & C. Mayer (Eds.), *European Financial Integration* (pp. 35–63). Cambridge University Press. [154]

Hoxha, I. (2013). The market structure of the banking sector and financially dependent manufacturing sectors. *International Review of Economics & Finance*, *27*, 432–444. [153, 179, 180]

Keeley, M. C. (1990). Deposit insurance, risk, and market power in banking. *American Economic Review*, *80*(5), 1183–1200. [155]

King, R. G., & Levine, R. (1993a). Finance and growth: Schumpeter might be right. *Quarterly Journal of Economics*, *108*(3), 717–737. [154]

King, R. G., & Levine, R. (1993b). Finance, entrepreneurship and growth. *Journal of Monetary Economics*, *32*(3), 513–542. [154]

King, R. G., & Levine, R. (1993c). *Financial intermediation and economic development*. Cambridge University Press. [154]

Levine, R. (1997). Financial development and economic growth: Views and agenda. *Journal of Economic Literature*, *35*(2), 688–726. [152]

Levine, R. (2005). Finance and growth: Theory and evidence. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of Economic Growth* (pp. 865–934). North-Holland. [154]

Levine, R., & Zervos, S. (1998). Stock markets, banks, and economic growth. *American Economic Review*, *88*(3), 537–558. [154]

Lipsey, R. G., & Lancaster, K. (1956). The general theory of second best. *Review of Economic Studies*, *24*(1), 11–32. [152, 172]

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. Oxford University Press. [204]

Maudos, J., & Fernandez de Guevara, J. (2006). Banking competition, financial dependence and economic growth. *MPRA Paper No. 15254*. [153, 179]

McKinnon, R. (1973). *Money and capital in economic development*. The Brookings Institutions. [154]

Obstfeld, M. (1994). Risk-taking, global diversification, and growth. *American Economic Review*, *84*(5), 1310–1329. [154]

Pagano, M. (1993). Financial markets and growth: An overview. *European Economic Review*, *37*(2-3), 613–622. [154]

Schumpeter, J. A. (1911). *Theorie der wirtschaftlichen Entwicklung: Eine Untersuchung über Unternehmergewinn, Kapital, Kredit, Zins und den Konjunkturzyklus* (Vol. 1). Duncker und Humblot. [154]

Shaw, E. S. (1973). *Financial deepening in economic development*. Oxford University Press. [154]

Smith, R. T. (1998). Banking competition and macroeconomic performance. *Journal of Money, Credit and Banking*, *30*(4), 793–815. [152, 154]

Conclusion

Presenting four articles, the dissertation at hand contributes to three independent subfields of economic theory by exploring different facets of incentives. As will be elaborated below, the incentive perspective adopted in the dissertation thereby not only expands our current knowledge of economic interaction but also offers valuable implications for practice and opens up promising avenues for further research. This dissertation thus underlines once more Aumann's (2006, p. 17075) famous conclusion that *"economics is all about incentives"*.

Due to their focus on contract theory, Articles 1 and 2 directly address our understanding of the effectiveness and design of incentives. They uncover intriguing and paradoxical phenomena that originate from dynamic interaction: a more lucrative environment may leave all interacting parties worse off and reduce the likelihood that the interaction will be concluded successfully. These phenomena may arise irrespective of whether incentives are exogenously given by the economic environment (cf. Chapter 1) or endogenously designed by an optimizing party, i.e., the principal (cf. Chapter 2).

Our insights bear crucial implications for both academic research and practical application. First, as it is shown that the effect of incentives and incentive mechanisms is prone to reverse in dynamic settings – i.e., to become counterproductive – our analysis offers an explanation for the lack of incentives in many real-world situations. The absence of high-powered incentives in firms has not only been considered by many theorists (see, e.g., Che and Yoo, 2001; Holmström and Milgrom, 1990, 1991; Williamson, 1985) but it has also been confirmed manifoldly in the empirical literature, which finds financial incentives for employees to be either completely absent or low in relation to total income (see, e.g., Baktash et al., 2022; Bell and Van Reenen, 2014; Hong et al., 2019; Lemieux et al., 2009). Second, since agents are shown to strategically withhold effort even to the extent that overall project success becomes less likely, our results help to understand the prevailing empirical patterns of project delays and failures. These are frequently reported for *public infrastructure* (e.g., Flyvbjerg et al., 2003; Mittal et al., 2020; Steininger et al., 2021), *research and development* (e.g., Gupta and Wilemon, 1990; Lhuillery and Pfister, 2009; Radas and Bozic, 2012), as well as *IT projects* (e.g., Al-Ahmad et al., 2009; Brown et al., 2007; Whitney and Daniels, 2013).

Articles 1 and 2 might also serve as the starting point for further research. In general, the presented time-discrete model with a finite horizon is highly plausible yet easily tractable, making it well suited for future research endeavors exploring incentives in dynamic interaction under moral hazard. A particularly fruitful avenue for further research might be unequal remuneration, i.e., allowing the principal to discriminate between agents. Although a thorough analysis of the equilibrium contracts is left for further research, some results can readily be predicted. While spot contracts and long-term contracts under limited commitment will induce equal compensation of all agents in a period, full commitment to long-term contracts may very well induce unequal bonus payments in certain periods.[1] According to this preliminary result, equal remuneration would constitute an interesting manifestation of the friction created by the lack of commitment. Put differently, since limited commitment is indeed extremely plausible in real-world applications, this result would provide a theoretical basis for why equal pay is so widespread in practice. This line of research would thus help to explain Hart and Holmström's (1987, p. 90) observation that *"in the real world incentive schemes do show variety, but not to the degree predicted by the basic theory."* It would also provide a novel contribution from a dynamic perspective to the recent debate on (in)equal remuneration in the principal-agent literature (see, e.g., Halac et al., 2021; Moroni, 2022; Weinschenk, 2021; Winter, 2004).

Article 3 contributes to environmental economics by exploring a simple mechanism to overcome free-riding incentives in greenhouse gas emissions. For sufficiently symmetric countries, allowing them to endogenously design a joint cap-and-trade system by negotiating the amount and allocation of permits suffices to implement efficient emissions levels and maximize welfare. The implications for policy-makers are evident. First, negotiations constitute, from a theoretical perspective, a valuable tool for designing cap-and-trade systems. Second, all obstacles – such as transac-

1. The underlying idea is as follows. Relying on spot contracts or long-term contracts under limited commitment, the principal will offer equal bonus payments in the last period since this is the cheapest way to induce *any* success probability in that period due to the concavity of efforts. Thus, in the second last period, all agents have the same continuation payoff. Given equal continuation payoffs, equal remuneration is again the most cost-efficient way for the principal to induce *any* success probability and so on. Technically speaking, equal remuneration is thus induced by backward induction as the solution concept in conjunction with the concavity of efforts. With full commitment, however, the principal faces a trade-off when designing optimal contracts. While providing equal incentives is still the cheapest way to induce a certain success probability within a certain period, it may render setting incentives in earlier periods more costly. The principal might thus be better off when incentivizing different agents in different periods by unequal bonuses to exploit "cheap incentives".

tion costs (cf. Montero, 1998), imperfections in the emissions market (cf. Stavins, 2008), and conflicting national regulations (cf. Hahn and Stavins, 2011) – that prevent establishing joint emissions trading systems or hinder linking existing schemes must be removed. This will then induce cooperative and efficient behavior: countries will agree to set up joint cap-and-trade systems, cap emissions at the efficient level, and share the efficiency gains through the allocation of permits.

A further research agenda building on our results could be threefold. First, it would be worthwhile to investigate whether allowing for explicit side payments would generalize our mechanism to settings with heterogeneous countries. Afterward, the question of whether this mechanism also applies to multilateral negotiations, i.e., for more than two countries, must then be clarified. Lastly, from a political economy viewpoint, how our theoretical results can be translated into political action needs to be analyzed, as implemented policies in the context of international environmental agreements often diverge from efficiency-oriented policy recommendations (Wangler et al., 2013).

Article 4 adds to microfounded macroeconomics: it unravels the relationship between banking competition on the one hand and long-run economic growth and welfare on the other. By determining the relative importance of savings incentives versus institutional investment, capital dependence of the production sector is shown to be the key determinant of whether a monopolistic or a competitive banking sector is more favorable. The former yields higher levels of long-term growth and welfare in a relatively capital-intensive production environment, whereas the latter leads to more favorable outcomes in a labor-intensive environment. Although quite simple, this theoretical condition is supported by ample empirical evidence (see, Beck et al., 2004; Cetorelli and Gambera, 2001; Deidda and Fattouh, 2005; Hoxha, 2013; Maudos and Fernandez de Guevara, 2006). Our results therefore suggest that establishing and maintaining a competitive banking sector is particularly important for developing countries, as they typically rely on labor-intensive industries such as textiles, hospitality, and agriculture.

We also find that the presence of banks unambiguously promotes economic growth compared to the absence of financial intermediation, regardless of the form of interbank competition. Accordingly, introducing a banking sector that enables risk-sharing provides a way to overcome growth obstacles in underdeveloped countries. Our findings furthermore support the competition-fragility hypothesis discussed in the finance literature: competition destabilizes the banking system (cf. Freixas and Rochet, 2008). Article 4 therefore suggests that the more competitive the banking system is, the more supervisory and regulatory measures are required to prevent banking crises.

The presented framework provides a fruitful basis for further empirical and theoretical research. Empirically, it would be interesting to directly examine the correlation between the combination of bank competition and capital dependence on the one hand and growth/welfare on the other for a large sample of economies. The results could then be contrasted against the predictions of our model. Further theoretical studies could proceed immediately from the extension of our model. Analyzing the qualitative dynamics induced by dividend payments could further shed light on the controversial question of whether or not banks can trigger complex dynamics and persistent business cycles in an otherwise calm economy (see, e.g., Banerji et al., 2004; Ritschel and Wenzelburger, 2024; Smith, 1998, as well as Chapter 4).

References

Al-Ahmad, W., Al-Fagih, K., Khanfar, K., Alsamara, K., Abuleil, S., & Abu-Salem, H. (2009). A taxonomy of an it project failure: Root causes. *International Management Review*, *5*(1), 93–104. [209]

Aumann, R. J. (2006). War and peace. *Proceedings of the National Academy of Sciences*, *103*(46), 17075–17078. [208]

Baktash, M. B., Heywood, J. S., & Jirjahn, U. (2022). Worker stress and performance pay: German survey evidence. *Journal of Economic Behavior & Organization*, *201*, 276–291. [208]

Banerji, S., Bhattacharya, J., & Van Long, N. (2004). Can financial intermediation induce endogenous fluctuations. *Journal of Economic Dynamics and Control*, *28*(11), 2215–2238. [211]

Beck, T., Demirgüç-Kunt, A., & Maksimovic, V. (2004). Bank competition and access to finance: International evidence. *Journal of Money, Credit and Banking*, *36*(3), 627–648. [210]

Bell, B., & Van Reenen, J. (2014). Bankers and their bonuses. *Economic Journal*, *124*(574), F1–F21. [208]

Brown, S. A., Chervany, N. L., & Reinicke, B. A. (2007). What matters when introducing new information technology. *Communications of the ACM*, *50*(9), 91–96. [209]

Cetorelli, N., & Gambera, M. (2001). Banking market structure, financial dependence and growth: International evidence from industry data. *Journal of Finance*, *56*(2), 617–648. [210]

Che, Y.-K., & Yoo, S.-W. (2001). Optimal incentives for teams. *American Economic Review*, *91*(3), 525–541. [208]

Deidda, L., & Fattouh, B. (2005). Concentration in the banking industry and economic growth. *Macroeconomic Dynamics*, *9*(2), 198–219. [210]

Flyvbjerg, B., Bruzelius, N., & Rothengatter, W. (2003). *Megaprojects and risk: An anatomy of ambition*. Cambridge University Press. [208]

Freixas, X., & Rochet, J.-C. (2008). *Microeconomics of banking*. MIT Press. [210]

Gupta, A. K., & Wilemon, D. L. (1990). Accelerating the development of technology-based new products. *California Management Review*, *32*(2), 24–44. [208]

Hahn, R. W., & Stavins, R. N. (2011). The effect of allowance allocations on cap-and-trade system performance. *Journal of Law and Economics*, *54*(4), S267–S294. [210]

Halac, M., Lipnowski, E., & Rappoport, D. (2021). Rank uncertainty in organizations. *American Economic Review*, *111*(3), 757–86. [209]

Hart, O., & Holmström, B. (1987). The theory of contracts. In T. F. Bewley (Ed.), *Advances in economic theory: Fifth world congress* (pp. 71–156). Cambridge University Press. [209]

Holmström, B., & Milgrom, P. (1990). Regulating trade among agents. *Journal of Institutional and Theoretical Economics*, *146*(1), 85–105. [208]

Holmström, B., & Milgrom, P. (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*, 24–52. [208]

Hong, B., Kueng, L., & Yang, M.-J. (2019). Complementarity of performance pay and task allocation. *Management Science*, *65*(11), 5152–5170. [208]

Hoxha, I. (2013). The market structure of the banking sector and financially dependent manufacturing sectors. *International Review of Economics & Finance*, *27*, 432–444. [210]

Lemieux, T., MacLeod, W. B., & Parent, D. (2009). Performance pay and wage inequality. *Quarterly Journal of Economics*, *124*(1), 1–49. [208]

Lhuillery, S., & Pfister, E. (2009). R&d cooperation and failures in innovation projects: Empirical evidence from french cis data. *Research Policy*, *38*(1), 45–57. [208]

Maudos, J., & Fernandez de Guevara, J. (2006). Banking competition, financial dependence and economic growth. *MPRA Paper No. 15254*. [210]

Mittal, Y. K., Paul, V. K., Rostami, A., Riley, M., & Sawhney, A. (2020). Delay factors in construction of healthcare infrastructure projects: A comparison amongst developing countries. *Asian Journal of Civil Engineering*, *21*, 649–661. [208]

Montero, J.-P. (1998). Marketable pollution permits with uncertainty and transaction costs. *Resource and Energy Economics*, *20*(1), 27–50. [210]

Moroni, S. (2022). Experimentation in organizations. *Theoretical Economics*, *17*(3), 1403–1450. [209]

Radas, S., & Bozic, L. (2012). Overcoming failure: Abandonments and delays of innovation projects in smes. *Industry and Innovation*, *19*(8), 649–669. [209]

Ritschel, P., & Wenzelburger, J. (2024). Financial intermediation and efficient risk sharing in two-period lived olg models. *Economic Theory Bulletin*, *12*, 57–78. [211]

Smith, R. T. (1998). Banking competition and macroeconomic performance. *Journal of Money, Credit and Banking*, *30*(4), 793–815. [211]

Stavins, R. N. (2008). A meaningful u.s. cap-and-trade system to address climate change. *Harvard Environmental Law Review*, *32*(2), 293–371. [210]

Steininger, B. I., Groth, M., & Weber, B. L. (2021). Cost overruns and delays in infrastructure projects: The case of stuttgart 21. *Journal of Property Investment & Finance*, *39*(3), 256–282. [208]

Wangler, L., Altamirano-Cabrera, J.-C., & Weikard, H.-P. (2013). The political economy of international environmental agreements: A survey. *International Environmental Agreements: Politics, Law and Economics*, *13*(3), 387–403. [210]

Weinschenk, P. (2021). Unequal remuneration in agency relationships. *Working Paper. Available at SSRN 3941136*. [209]

Whitney, K. M., & Daniels, C. B. (2013). The root cause of failure in complex it projects: Complexity itself. *Procedia Computer Science*, *20*, 325–330. [209]

Williamson, O. E. (1985). *The economic institutions of capitalism: Firms, markets, relational contracting*. Free Press. [208]

Winter, E. (2004). Incentives and discrimination. *American Economic Review*, *94*(3), 764–773. [209]

Tom Rauber

PERSONAL INFORMATION

Address:	Gottlieb-Daimler-Straße 42 D-67663 Kaiserslautern
E-Mail:	tom.rauber@wiwi.uni-kl.de
Phone:	+49 (631) 205 4806

EDUCATION

09/2017 – 12/2019	**Business Management & Engineering in the field of Mechanical Engineering (M.Sc.)** **University of Kaiserslautern (Germany)** **Grade:** Very good (1.1) **Economic specialisation block:** Industrial Organisation
01/2019 – 04/2019	**Visiting Student at the Department of Economics** **University of Oxford (United Kingdom)** **Included modules:** Public Economics, Labour Economics **Academic advisor:** Prof. Dr. Johannes Abeler
10/2013 – 08/2017	**Business Management & Engineering in the field of Mechanical Engineering (B.Sc.)** **University of Kaiserslautern (Germany)** **Grade:** Very good (1.2) **Structure:** 50% Economics and Management, 50% Engineering
08/2004 – 06/2013	**Allgemeine Hochschulreife (A-Level equivalent)** **Gesamtschule Marpingen (Germany)** **Grade:** Very good (1.1) **Advanced level subject:** Mathematics, English and Economics

HONOURS & AWARDS

08/2023	**Wipprecht Grant** for the research stay at the University of Cambridge
01/2022	**Erasmus+ Grant** for the research stay at the University of Luxembourg
05/2021	**Sparkassen Award** for outstanding academic achievements and extracurricular commitment during the graduate programme (top 3)
10/2017 – 12/2019	**Scholarship** of the German Academic Scholarship Foundation ('Studienstiftung des deutschen Volkes')
07/2018	**Hornbach Award** for outstanding academic achievements and extracurricular commitment during the undergraduate programme (top 4)
10/2013 – 03/2017	**Scholarship** of the German National Scholarship Programme ('Deutschlandstipendium')

PROFESSIONAL EXPERIENCE

02/2020 – present **Doctoral Researcher and PhD Candidate in Microeconomics**
University of Kaiserslautern (Germany)
Supervisor: Prof. Dr. Philipp Weinschenk
Research: 'Effects and Design of Incentives in Dynamic, Non-Repeated Principal-Agent Models' (Project funded by DFG)
Teaching: Behavioural Economics; Game Theory

10/2023 – 12/2023 **Visiting Doctoral Researcher**
University of Cambridge (United Kingdom)

02/2022 – 06/2022 **Visiting Doctoral Researcher**
University of Luxembourg (Luxembourg)

10/2019 – 12/2019 **Trainee in the Economic Division**
German Embassy Washington D.C. (USA)
Events: Participation in the International Monetary Fund (IMF) Annual Meetings; Attendance at think tank events
Reports: Trade, 5G, tariffs and agriculture

03/2015 – 10/2019 **Student Research Assistant in Production Management**
University of Kaiserslautern (Germany)
Position: Tutor in the undergraduate course on Production Management
Task: Preparation and discussion of exercises (group size: 50-100 students)

PUBLICATIONS

Rauber, T., & Ritschel, P. (2024). Banking Competition and Capital Dependence of the Production Sector: Growth and Welfare Implications, *International Review of Economics & Finance, 89,* 676-698.
Rauber, T., & Weinschenk, P. (2024). Dynamic Interaction and (In)effectiveness of Financial Incentives. *Working Paper. Available at SSRN 3987829.*
Rauber, T., & Weinschenk, P. (2023). Detrimental Incentive Mechanisms in Dynamic Principal-Agent Relationships. *Working Paper. Available at SSRN 4515328.*
Rauber, T., & Naumann, F. (2023). Designing Emissions Trading Schemes: Negotiations on the Amount & Allocation of Permits. *Available at SSRN 4636471.*

CONFERENCES & WORKSHOPS

2024 **Econometric Society Meeting** in New Delhi
Conference on Mechanism and Institution Design in Budapest

2023 **Econometric Society Meetings** in Los Angeles and Singapore
Microeconomic Theory Conference at University of Cambridge

2022 **IAAEU Workshop on Labour Economics** in Trier

ADDITIONAL QUALIFICATIONS

Language skills:

German (mother tongue)
English (fluent, IELTS Band 8)
French (basic)

Software skills:

Mathematica (advanced)
MS Office (advanced)

Programming languages:

TeX (advanced)
TikZ (advanced)
Java (basic)
SQL (basic)

EXTRACURRICULAR ACTIVITIES & INTERESTS

01/2021 – present	**Scholarship Reviewer** **German Academic Scholarship Foundation**
08/2015 – present	**Board Member (Secretary)** **Rotaract Club Kaiserslautern**
Interests	**Sports** (shooting, cycling, hiking, bouldering) and **travelling**

REFEREES

Prof. Dr. Philipp Weinschenk
Professor of Microeconomics
E-Mail: philipp.weinschenk@wiwi.uni-kl.de
Phone: +49 631 205 3382

Prof. Dr. Daniel Heyen
Professor of Environmental Economics
E-Mail: daniel.heyen@wiwi.uni-kl.de
Phone: +49 631 205 3117