Communication, Delegation and Performance Evaluation in Organizations

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2013

vorgelegt von

Miriam Margarete Schütte

Referent:	Prof. Dr. Klaus M. Schmidt
Korreferent:	Prof. Dr. Philipp C. Wichardt
Promotionsabschlussberatung:	6. November 2013

Datum der mündlichen Prüfung: 24. Oktober 2013

Namen der Berichterstatter: Klaus Schmidt, Philipp Wichardt, Fabian Herweg

Für Ingo.

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

In erster Linie möchte ich meinem Doktorvater Klaus M. Schmidt für seine Unterstützung und seinen fachlichen Rat während meiner Promotion herzlich danken. Die Diskussionen mit ihm waren immer sehr wertvoll und herausfordernd. Meinem Zweitbetreuer, Philipp C. Wichardt, gilt ein ganz besonderes Dankeschön. Neben der inspirierenden Zusammenarbeit am ersten Kapitel meiner Dissertation danke ich ihm auch für die stete Ermutigung während meiner Dissertationszeit. Schließlich bedanke ich mich bei Fabian Herweg dafür, dass er sich bereit erklärt hat, die Rolle des Drittgutachters zu übernehmen.

Ich danke Alvaro Sandroni und Niko Matouschek für ihre Gastfreundschaft während meines Forschungsaufenthalts an der Kellogg School of Management, Evanston, USA. Für die finanzielle Unterstützung dieses Aufenthalts danke ich LMUMentoring.

Meinen Kollegen am Seminar für Wirtschaftstheorie und an der Munich Graduate School of Economics möchte ich ganz besonders danken, für viele wertvolle Diskussionen und für die gemeinsam verbrachte Zeit. Insbesondere bedanke ich mich bei Carmen Thoma für die gute und freundschaftliche Zusammenarbeit. Katrin Peters, Ines Pelger und Sandra Ludwig danke ich für die persönliche Unterstützung und Motivation.

Ich bin meinen Eltern sehr dankbar, dass ich mich bei all meinen Entscheidungen immer auf ihren Rückhalt verlassen kann.

Mein größter Dank gilt Dir, Ingo, für die großartige Unterstützung und den unermüdlichen Zuspruch, für Dein Vertrauen und Deine Liebe.

Miriam Schütte, Dezember 2013

Contents

List of Figures

List of Tables

Preface

In an organization, individuals can cooperate in order to achieve a joint surplus. Yet, individual objectives are often not aligned with the collective goal, which potentially yields opportunistic behavior. Arguments that conflicting interests are widespread and potentially detrimental to the performance of an organization go back to Knight (1921), who stated that “the internal problems of the corporation, the protection of its various types of members and adherents against each other’s predatory propensities, are quite as vital as the external problem of safeguarding the public interests against exploitation by the corporation as a unit.” This problem of opportunism is easily solved if it is feasible to write and enforce a complete contingent contract, specifying an action for all potential states of the world. However, this is often impossible since the optimal action depends on some state of the world, which is private information of one party (*hidden information*), or because the action itself cannot be observed (*hidden action*). Both situations induce organizational inefficiencies which potentially result in a monetary loss to the firm.

One challenge organizations face is that decision makers often lack relevant information in order to take optimal decisions. As already Hayek (1945) emphasized, “[...] the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess.” Thus, information has to be transmitted in order to optimize decision making in organizations and to alleviate inefficiencies. However, information processing might be costly because the decision maker is not able to master all relevant information or due to his limited capacity to communicate with all parties. In addition, and more importantly, parties might be reluctant to share information because they follow different objectives.

One way to mitigate this inefficiency due to asymmetric information in the presence of conflicting interests is to involve the informed party in the decision process by granting her certain decision rights. While private information is then fully taken into account, it might be risky to delegate the decision if the informed party's interest is imperfectly aligned with the organizational goal. Hence, a fundamental question in the design of organizations is how to optimally allocate decision rights in order to limit agency costs. The idea that authority plays a central role in the nature and function of a firm has a long tradition among economists, including Knight (1921), Coase (1937), Simon (1951), Arrow (1974) and Williamson (1975).

Instead of delegating decision rights to an informed agent, the principal – who represents the firm's interests – can simply ask the agent for a recommendation. Yet, if the principal is not able to verify the correctness of the agent's statement, the agent has an incentive to lie in order to manipulate the principal's decision in a favorable direction. As Crawford and Sobel (1982) state in their seminal paper, "sharing information makes available better potential agreements, but it also has strategic effects that make one suspect that revealing all to an opponent is not usually the most advantageous policy. Nevertheless, it seems clear that even a completely self-interested agent will frequently find it advantageous to reveal some information." This strategic transmission of information thus alleviates informational asymmetries in the organization.

A second source of inefficiency in organizations stems from the fact that actions or efforts of subordinate agents are often not observable to the principal or cannot be contracted upon. If individual efforts influence the success of some project, but agents cannot be rewarded contingent on this success, the agent's rational and self-interested reaction is to shirk by exerting inefficiently low effort. Anticipating this, the principal might refuse to hire the agent in the first place and the project is not implemented, although a successful realization would benefit both the principal and the agent.

In such a setting, efficiency can only be achieved if the principal trusts that the agent exerts high effort and the agent cooperates in turn – which is not possible assuming individuals to behave self-interestedly. However, behavioral economists and psychologists have shown that non-binding communication, such as informal agreements, statements of intent or promises, can be an effective tool to increase trust and to enhance cooperation, thereby amending organizational inefficiencies.[1]

This dissertation deals with both contractual and informal remedies to mitigate organizational inefficiencies due to asymmetric information in the presence of conflicting interests (*hidden information*; chapters 1 and 2) or due to shirking incentives when the agent's action is unobservable (*hidden action*; chapter 3). The first two chapters theoretically deal with the optimal allocation of formal authority if the agent is better informed than the principal and potentially exhibits diverging preferences. In chapter 1, we consider the effect of an opportunity to evaluate the agent on the principal's decision to delegate authority. Chapter 2, instead, investigates the merits of retaining authority if the principal can ask the agent for a non-verifiable recommendation. The third chapter experimentally investigates how non-binding communication can increase trust and enhance cooperation if subjects have an incentive to shirk.

For an illustration of the first two chapters, consider a situation where a principal, for example the headquarters of a firm, has to decide about the size of a certain project, such as the investment into a new plant or the allocation of human resources to a marketing campaign. Moreover, assume that the first best decision for the organization depends on some noisy variable, like market demand or consumer needs. While the headquarters is not familiar with these local conditions, a subordinate agent, for instance a division manager, is aware of these circumstances.

In this context, chapter 1 focuses on the role of interim performance evaluations when the principal is uncertain about the loyalty of the agent.[2] The agent can either be aligned with the principal's preferences or biased in the sense that the agent prefers a different project than the principal. We analyze a two-period interaction where the principal can

[1]Research on the effect of non-binding communication on cooperation goes back to Loomis (1959). Kerr and Kaufman-Gilliland (1994) and Sally (1995) survey experimental results along these lines.

[2]This chapter is based on the article "Delegation and Interim Performance Evaluation", which is joint work with Philipp C. Wichardt from the University of Rostock.

observe the agent's performance, as reflected in his project choice, after delegating the decision right in the first period – which gives rise to a signaling game. Furthermore, we assume that the agent's wage is tied to the market belief about his type, and that the principal can disclose a performance evaluation to the public.

Given that decision rights are delegated to the agent in the first period, an interim performance evaluation allows the principal to condition his delegation decision in the second period on the agent's project choice if he learns about the agent's type. In turn, a biased agent has some incentive to align his behavior in the first period to that of an unbiased agent in order to regain authority in the second period. If the evaluation is public, this adjustment incentive is further increased by the prospect of a higher wage in the second period.

We find that evaluating the agent privately always benefits the principal as he possibly acquires information about the agent's type. Furthermore, the principal discloses the result of the performance evaluation if the corresponding expected wage increase is not too large. In any case, the principal is more likely to delegate authority to the agent in the first place if an interim performance evaluation takes place, as the evaluation effectively reduces the economic risk of delegation. Finally, we show that in view of aggregate welfare, it is optimal to disclose an evaluation to the public since this strengthens the incentives for the biased agent to act in line with the principal's preferences and thereby minimizes agency costs.

We contribute to the existing literature on the optimal delegation of authority by considering a repeated interaction between the principal and the agent, where the principal can revise his delegation decision after the first period.[3] In an extension, we show that our results are robust to a finite repetition of the one-period game, where the principal can observe the agent's project choice after each period. From a methodological perspective, we extend the beer-quiche game, first discussed by Cho and Kreps (1987), by allowing payoffs to be endogenous in that they may depend on posterior market beliefs.

[3]Related works either consider reputation concerns when the principal and the agent are equally uninformed about the agent's type (e.g. Blanes i Vidal, 2007; Englmaier et al., 2010) or analyze asymmetric information about the agent's type in the absence of reputation concerns (e.g. Armstrong, 1995; Frankel, 2010; Kovàč and Krähmer, 2012).

Most studies considering conflicts of interest in organizations assume that the agent exhibits a systematic bias, implying that he always prefers to choose a larger project than the principal, independent of the realized state of the world.[4] Alternatively, it could be the case that managers act in line with the headquarters' objective as long as the environment resides in some status quo, but tend to under- or overreact to changes in the state of the world relative to the first best decision from a corporate point of view. Imagine, for example, a manager who is delegated to build up a new branch in an emerging market. While he acts in line with the organizational goal in case the market is stable, he potentially overreacts to changes in consumer demand since he is not familiar with this market environment. Thus, the manager exhibits a state-dependent bias.

In chapter 2, we assume that the preference divergence between the principal and the agent can contain both a systematic *and* a state-dependent component.[5] In addition, the principal, who is unaware of the market environment, can ask the informed agent for advice if he keeps the decision right. We build on the seminal paper by Crawford and Sobel (1982) in modeling strategic information transmission if the principal and the agent potentially agree about the right course of action in some state of the world. In this framework, we first characterize communication equilibria for this more general set of preference divergences between the principal and the agent. Second, we consider the delegation of decision rights as an alternative to retaining authority and communicating with the agent and analyze the optimal allocation of authority, thereby extending the model by Dessein (2002).

Introducing a state-dependent component into the agent's bias changes the precision of strategic information transmission in a Bayesian Nash equilibrium, which is characterized by a partition of the state space such that the agent truthfully reveals the interval which contains the state of the world. If the agent reacts stronger to changes in the state of the world than the principal, his incentive to misrepresent his private information is largest

[4] Prominent examples for systematic biases include status concerns or empire building, that is, to grow the own division beyond the optimal size in order to strengthen one's position in the organization (see, for instance, Chen et al., 2008, or Du et al., 2013).

[5] This chapter is based on the article "Communicating with Extreme or Conservative Agents".

for extreme states. Hence, extreme messages are noisy in equilibrium in order to make exaggerations costly for the agent, while messages near the agreement state are rather precise. In this case, the transmission of information can be infinitesimally precise near the agreement state. On the other hand, if the agent reacts more conservatively to changes in the state of the world than the principal, he has an incentive to distort his reports towards the agreement state. Consequently, messages are noisy even if preferences are almost aligned.

Different from Crawford and Sobel (1982), we find that a compact state space can be partitioned into infinitely many intervals in equilibrium if and only if the agreement state realizes with positive probability – given that the agent reacts stronger to changes in the state of the world than the principal. However, if the agent's reaction to a changing environment is more conservative than the principal's, communication is noisier in the sense that the state space is partitioned into only finitely many intervals. If the principal and the agent even exhibit reversed preferences, in the sense that the principal's preferred project increases in the state of the world while the agent's preference decreases, communication does not transmit any information. With regard to the optimal allocation of authority, we find that the principal retains authority if the agent's bias is "large" enough, which is in line with Dessein (2002). However, while Dessein's model predicts that communication is uninformative if the agent's systematic bias is so large that the principal retains decision rights, we find that, in the more general case, communication can be informative *and* dominate delegation from the principal's point of view.

To the best of our knowledge, chapter 2 is the first study to provide a comprehensive model of strategic information transmission for a general linear preference divergence. Thereby, it embeds the models by Crawford and Sobel (1982) and Alonso et al. (2008), where the latter consider a purely state-dependent bias and overreacting agents. Furthermore, in terms of the optimal allocation of authority, the present analysis emphasizes the importance of considering subordinates' likely reaction to environmental changes when designing organizations.

In both of the above chapters we assume that individuals are rational and self-interested, and that information is asymmetric with regard to the state of the world (and the agent's type), hence communication fulfills the mere function of strategic information transmission. Alternatively, economic experiments have shown that communication can also take on the role of enhancing trust and cooperation, in contrast to any prediction based on rationality and self-interest. In this context, promises have been found to be particularly effective.[6]

In chapter 3, we experimentally investigate the effect of promises on trust and cooperation, given that individuals have an incentive to free ride on their team partner's decision to trust.[7] More precisely, we are interested in reasons *why* people stick to an informal commitment and analyze the effect of social image concerns on the decision to keep a promise. In a controlled laboratory experiment, participants are randomly matched in pairs and play a one-shot trust game. A first mover decides whether to enter the game or to opt out, the latter choice inducing a low outside option for both players. If she enters the game, a second mover chooses between a selfish option, yielding a payoff of zero for the first mover, and cooperation, in which case a chance move determines whether the first mover gets a positive payoff or zero. The second mover maximizes his payoff by choosing the selfish option. Prior to the strategic decisions, the second mover sends one out of three predefined messages to the first mover, including a promise to cooperate. In order to test for social image concerns, we vary the ex-post observability of the second mover's action.

We observe slightly more cooperation by the second mover if his action choice is revealed than if it is not, though the results are not significant. The fact that already 81% of second movers deliver on their promises even if their action is not observable speaks to an inherent preference for keeping one's word as a reason for why people stick to a promise. The so-called "promise keeping per se" (Vanberg, 2008) seems to play a more important role than social image concerns in our experiment. Interestingly, we find that the preference

[6] See Ellingsen and Johannesson (2004) or Bochet and Putterman (2009), amongst others.

[7] This chapter is based on the article "Promises and Image Concerns", which is joint work with Carmen Thoma from the University of Munich.

for sticking to one's word does only exist for promises and not for statements of intent, which we add to the set of eligible messages in a further treatment. While most of the promises are kept, statements of intent tend to be broken. In line with this result, the possibility to communicate increases cooperation by second movers if they can only choose between sending a promise or an empty message, while communication has no effect on second movers' behavior if they have the additional option of sending a statement of intent. However, the receivers of messages trust both a promise and a statement of intent to the same degree, which is anticipated by the communicating party. Hence, guilt aversion, as argued by Charness and Dufwenberg (2006), is unlikely to account for promise keeping in our setting. A guilt-averse individual feels guilt proportional to the amount that he disappoints other's expectations and should thus behave similarly upon communicating a statement of intent as after sending a promise in our setting.

Chapter 3 is one of the first economic studies analyzing the combined effect of communication and social image concerns on cooperation. While previous works focus on the effect of communication on cooperation, our experiment is designed to gain insight into the relative importance of different reasons *why* people stick to a promise. Furthermore, this experiment highlights how the set of messages available to experimental subjects influences the effect of communication.

The three chapters in this dissertation are concerned with different approaches to alleviate inefficiencies in organizations, which arise due to informational asymmetries or incentives to free ride. Chapter 1 shows that introducing interim performance evaluations fosters the delegation of decision rights to the agent and hence improves the use of private information. In chapter 2, we find that strategic information transmission can be more precise if the agent agrees with the principal in some state of the world than if he always dissents. Chapter 3 illustrates that non-binding communication can increase trust and enhance cooperation, mainly due to a preference to stick to a given promise.

All three chapters are self-contained and include their own introductions and appendices such that they can be read independently. The respective appendices contain all proofs, the experimental instructions and further details on some experimental results.

Chapter 1

Delegation and Interim Performance Evaluation*

A central question in the design of organizations, which has been much discussed in the literature over the last decades, is how to optimally allocate decision rights to subordinate agents (e.g. Holmström, 1977, 1984; Aghion and Tirole, 1997; Alonso and Matouschek, 2008). The general problem is that while agents often may have better information about the profitability of certain projects – or at least be able to obtain such information – this does not necessarily imply that they will always opt for the projects which are most preferred by the principal.[1] A possible reason for this is that agents may be biased and therefore disagree with the principal on which project to choose.[2]

In order to illustrate the problem, consider a situation where a principal, e.g. the headquarters of a firm, has to implement a certain project. Moreover, assume that the success of the project depends on its fit with some noisy variable, and that the realization of this variable is not observable to the principal but to some subordinate agent, e.g. a division

*This chapter is based on joint work with Philipp C. Wichardt.

[1]For a discussion of how the expected profit from delegation decreases due to too frequent choices of suboptimal projects see, for example, Jensen and Meckling (2012).

[2]Alternatively, agents may differ in their ability to interpret incoming signals about the state of the world so that delegation becomes very costly if the probability of unable agents letting opportunities pass, or choosing the wrong projects, is too high (e.g. Levy, 2005).

manager.[3] If the preferences of the principal and the agent are perfectly aligned, the problem can obviously be solved simply by either delegating the decision to the agent or asking him about his information. However, if there is a possibility that the agent's preferences differ from those of the principal, things change and the principal may rather prefer to take an uninformed decision himself.[4]

A possible way to improve matters for the principal in such a situation, which we focus on in the present chapter, is to introduce performance evaluations. In practice, such evaluations typically serve multiple goals, e.g. providing the agent with incentives (see MacLeod and Malcolmson, 1989, or Baker et al., 1994) or with performance feedback (see Zabojnik, 2011). What is more relevant in the present setting, however, is that performance appraisals are also used to determine assignments (cf. Cleveland et al., 1989), thereby taking an authority allocating function.

In fact, analyzing a two-period interaction with asymmetric information about the type of the agent, loyal or biased, we show that if the project can be subdivided into different stages, evaluating the agent on the way is always beneficial for the principal. First of all, the evaluation naturally grants the manager a possibility to learn the type of the agent and to condition at least the later delegation decision on this information. And, what is more, if the agent's wage is tied to the market's belief about his type, the principal may even want to make the result of the evaluation public. The reason for this is that the disclosure of the evaluation creates an incentive for the biased agent to align his behavior with that of the unbiased one in period one in order to gain from a higher wage in period two.[5] Thus, if the compensation of the agent in terms of an increased wage is not too high relative to the benefit of the principal from an aligned behavior in the first period, public evaluation is preferred by the principal.[6] In any case, with an interim performance

[3]Here one might think of certain aspects of market demand, the quality of some input, etc.

[4]There is, for example, anecdotal evidence that many managers engage in empire building, i.e. they invest more than would be optimal from the perspective of their principals (e.g. Jensen, 1986).

[5]For the case with more than two periods, which we consider as an extension, the effect persists for all but the last period (see Section 1.2.2 for details).

[6]Although the assumption of (partly) endogenous wages slightly complicates the analysis, the resulting overall two-period game is amenable to a common backward induction argument. Relying on standard equilibrium selection arguments (Cho and Kreps, 1987), we characterize the equilibria of the resulting

evaluation, there is more delegation in period one as the evaluation effectively reduces the economic risk of delegation for the principal.

Moreover, it is worth noting that the different outcomes in case of evaluation imply different effects on aggregate welfare compared to the situation without evaluation. As the exact nature of the (unique) outcomes depends the prior belief about the agent's type and the relation between wage schedule and utility, the details of the results are somewhat involved; see Section 1.1.5 for a discussion. Yet, at least two general observations can be made which are important from the perspective of a social planner. First of all, we find that, while wages have no direct impact on aggregate welfare,[7] they matter if the evaluation is public, namely through their (indirect) influence on outcomes. And, with high priors, giving all bargaining power to the agents maximizes aggregate welfare. Moreover, we show that, in terms of aggregate welfare, any evaluation conducted ought to be disclosed to the public. Thus, inasmuch as allowed by the stylized argument given in the sequel, strengthening the agent's rights regarding bargaining power and accessibility of evaluations is welfare enhancing in situations like the one considered here.

Related Literature.

As already indicated above, we are, of course, by no means the first to analyze delegation of authority in organizations. Since the seminal works by Holmström (1977, 1984), a stream of literature has been concerned with the question of how to optimally allocate decision rights to potentially biased but better informed subordinate agents. And optimal mechanisms to delegate decisions have been studied, for example, by Melumad and Shibano (1991) or Alonso and Matouschek (2007, 2008). Moreover, granting authority to the agent may not only serve as a means to optimally use available information. As demonstrated by Aghion and Tirole (1997), granting authority also provides an incentive to the agent to engage in information acquisition as formal control rights are vacuous without knowing the state of the world. In a similar vein, Bester and Krähmer (2008)

signaling game, depending on the observability of the evaluation. These equilibria are uniquely determined by the prior belief about the agent being loyal and the specific relation between wages and utilities.

[7] Note that wages affect only the redistribution of payoffs and, hence, have no direct effect on aggregate welfare.

show that delegating a project decision to the agent may also provide incentives to exert implementation effort. Adding to this literature, the present model focuses on the incentive effects of delegation in a context of asymmetric information about the type of the agent where delegation in early rounds induces reputation concerns due to the evaluation.

While reputation concerns themselves have already been studied in a delegation context – for example, Blanes i Vidal (2007) or Englmaier et al. (2010) consider situations with the principal and the agent being equally uninformed about the agent's ability and analyze a model of career concerns (see Holmström, 1999)[8] – the present analysis assumes that agents know their type ex ante in such a setting. However, inasmuch as optimal delegation mechanisms with asymmetric information about the agent's type have been studied before (e.g. Armstrong, 1995; Frankel, 2010; Kováč and Krähmer, 2012), reputation concerns are not central to the respective arguments.

Closer to the present discussion, Aghion et al. (2004) consider the optimal allocation of control rights contingent on announcing a type. They find that, if control is contractible, the loyal type will not be delegated any decision in order to induce truth-telling by the biased type. Different from the present setting, however, Aghion et al. consider a two-stage (but one-shot) interaction where the principal benefits from learning the agent's type but where there are no incentives for pooling derived from later periods.

Results that are more similar to the ones derived in this chapter, however, have been obtained for situations in which the principal cannot commit to a delegation mechanism such that transferring control is cheap talk (Crawford and Sobel, 1982). In particular, a common finding in these models is that agents shade their reports in order to gain the principal's trust in the future (e.g. Sobel, 1985, Benabou and Laroque, 1992, or Morris, 2001).[9] While considering a different environment, these results are similar in spirit to the

[8]Prat (2005) and Ottaviani and Sørensen (2006 a and b) analyze career concerns when the principal can not commit to delegate authority.

[9]While repeated games arguments do not lie at the heart of the present discussion, it is worth noting that similar arguments have been given in the literature on repeated games and reputation (see, for example, Kreps and Wilson, 1982, Milgrom and Roberts, 1982, and Mailath and Samuelson, 2001, for other models with incomplete information, or Fudenberg and Levine 1989, 1992 and Cripps et al. 2004, 2007, for examples where the asymmetry of information between players stems from a short-lived player playing against a long-lived one).

present ones, which show that, in a context with delegation, interim evaluations provide incentives for biased agents to align their choices with the principal's preferences.

A second important strand of literature, which is naturally related to the present discussion, is the literature on performance evaluations. In this area, most work is concerned with subjective performance evaluations and, thus, focuses on the problems caused by the non-verifiability of performance (e.g. Prendergast and Topel, 1996, or Fuchs, 2007).[10] In the present model, by contrast, the performance measure is perfectly observable (although not verifiable by courts), and the – strategic – disclosure of performance information together with the resulting incentives for the agent are central to our results.

Another related type of argument is provided by Goltsman and Mukherjee (2011), who consider a multi-stage tournament where the strategic disclosure of performance information in the first period can increase effort incentives in the second period. Different from our setting, however, they analyze information disclosure to peers in a tournament, who subsequently update winning probabilities. In contrast to this, we are concerned with the influence of signaling on outside option wages determined by market beliefs. Outside option wages, in turn, play a crucial role in Mukherjee (2008), who investigates strategic information disclosure to raiders in order to strengthen incentives for workers and shows that full disclosure is optimal. Although the main idea is related to the present one, Mukherjee assumes that principal and agent are equally uninformed about the agent's type, i.e. ex ante asymmetric information about types is not an issue.[11]

Finally, it is worth mentioning that the methodology applied in this analysis goes back to the seminal paper by Spence (1973), who analyzes signaling in a job market context. More specifically, the signaling game analyzed in the sequel can be viewed as a modified version of the beer-quiche game first discussed by Cho and Kreps (1987) except that we allow payoffs to be endogenous in that they may depend on posterior beliefs.

[10] To give one example, Baker et al. (1994) show that supervisors may be tempted to under report worker's performance and implicit incentives can be crowded out by explicit ones.

[11] Further works considering outside option wages in other contexts are Acemoglu and Pischke (1999) or Blanes i Vidal (2007).

The rest of the chapter is structured as follows. In Section 1.1, we introduce and analyze the model. In Section 1.2, we briefly discuss two possible extensions: a flexible timing of the evaluation and repeated evaluations. Section 1.3 concludes. All proofs are gathered in the appendix.

1.1 Model and Results

In the sequel, we describe the underlying delegation problem (Section 1.1.1). After briefly considering the benchmark case without evaluation in Section 1.1.2, we proceed to analyze the equilibria of the delegation game with private and with public intermediate evaluation (Section 1.1.3). As a next step, we compare the principal's optimal delegation decision in these three cases and analyze the principal's corresponding profit (Section 1.1.4). The analysis concludes with a discussion of welfare effects (Section 1.1.5).

1.1.1 The Underlying Delegation Problem

Consider the following standard delegation problem: A firm has to implement a project where the principal has the formal authority to decide which project is chosen, but he needs an agent to implement it.

Setup. The set of possible projects, X, is given by a subset of the real line, i.e. $X \subseteq \mathbb{R}$. Both the principal and the agent enjoy a personal benefit from a project choice $x \in X$.

However, by assumption, the principal himself is unable to implement his preferred project, e.g. because he lacks some important information about local conditions which only the agent can acquire. Thus, in order to circumvent his lack of information, the principal can delegate the project decision to the agent.[12] The agent can be either of two

[12]In line with the delegation literature, this lack of information can be modeled by the realization of some state of nature, θ, which is observed only by the agent. For the present purposes, we omit an explicit reference to θ as focus of the argument lies primarily on the agent's signaling motive (and not on the effects of different realizations of θ).

types, $\tau \in T := \{b, l\}$. If the agent is loyal, $\tau = l$, his preferences are aligned with those of the principal. If the agent is biased, $\tau = b$, the preferences of the principal and the agent do not match. A priori, the principal and the market share a common prior that the agent is loyal with probability $p \in (0, 1)$, i.e. $p = \Pr(\tau = l)$.

Remark 1. *For the purposes of the present discussion, we think of biases as reflecting differences in preferences. However, as should become clear from the subsequent argument, it is also possible to think of the agent's and the principal's preferences as being generally aligned. In that case, being biased would reflect a high cost when trying to satisfy these preferences for a given state of the world (and being unbiased would reflect a low cost of doing so). While some details of the modeling would of course change, the general thrust of the argument does not hinge on the interpretation.*

Utility. The players' utility is given by $u_i(x)$, $i \in \{P, b, l\}$. For the sake of argument, we assume that there always exists a unique optimal project x_i, $i \in \{P, b, l\}$, which maximizes utility for player i, $i \in \{P, b, l\}$,[13] i.e.

$$x_i := \arg\max_{x \in X} u_i(x) \ .$$

Note that, as the preferences of the biased agent differ from those of the principal, we have $x_b \neq x_P$, while the loyal agent prefers the same project, i.e. $x_l = x_P$.

The corresponding utility is referred to as u_i^+, i.e. for $i \in \{P, b, l\}$ we define

$$u_i^+ := u_i(x_i).$$

Similarly, utilities in case an undesired project is implemented are defined as follows:

$$\begin{aligned} u_b^- &:= u_b(x_P) \\ u_P^- &:= u_P(x_b) \\ u_l^- &:= u_l(x_b). \end{aligned}$$

[13] Here and in the following, we slightly abuse notation by sometimes referring to the agent's different types as player b and player l.

Furthermore, if the principal retains authority, he chooses a default project x_P^0, which can be interpreted as the best choice according to the principal's knowledge. The *default utilities* realized in this case are denoted by

$$u_i^0 := u_i(x_P^0) \ , \ \textit{for all } i \in \{P, b, l\} \, .$$

In order to make the problem interesting, we assume that delegating to a loyal agent is profitable for the principal while delegating to a biased agent is harmful, i.e.

$$u_P^- < u_P^0 < u_P^+ \, .$$

The biased (loyal) agent's default utility u_b^0 (u_l^0), in turn, is always smaller than his optimal choice, i.e. $u_b^0 < u_b^+$ ($u_l^0 < u_l^+$); whether u_b^0 (u_l^0) is smaller or larger than u_b^- (u_l^-) need not be specified here.

Contracts and Wages. As usual, the principal allocates authority, i.e. he determines whether he or the agent decides on the project x. The actual choice of x is not contractible, though. Only the allocation of authority can be contractually fixed. In either case, the principal pays a wage $w(\mu)$ to the agent, which depends on the market's belief about the agent being loyal, μ. Thus, while not restricting attention to common competitive wages, we still assume that the wage paid by the principal correctly matches the agent's expected outside option. In that sense, wages are considered endogenous and depend on the observability of the agent's project decision.

Note that the principal benefits from compensating the agent for potential upward adjustments in the market's belief whenever the payoff-surplus from the increased probability of employing the loyal agent outweighs the loss through increased wages. For high priors, this is generally the case if outside wages do not increase too fast. In order to ensure that this is also the case for small priors, we assume that the agent is needed for the implementation of the project and that employing a new or no agent is costly for the principal, e.g. because a new agent has to acquire firm-specific knowledge.[14]

[14]The point to note here is that, for small priors p and certain parameters, the only incentive for the

Moreover, although the project choice is observable to the principal (and to the market), we assume that it is not enforceable by the courts. Thus, the principal cannot reward the agent for a loyal decision, but instead pays a flat wage which can be adjusted after a public evaluation.[15]

As mentioned before, the principal and the market share a common prior that the agent is loyal with probability $p \in (0, 1)$. Accordingly, without any further (verifiable) information about the agent, i.e. if no evaluation takes place or evaluation is private, the common market wage for the agent is $w(p) \geq 0$.

Finally, we normalize $w(0) = 0$ and assume that w is (weakly) increasing in the market belief that the agent is loyal, μ.[16]

It is important to note that, if it is commonly known that the agent is loyal, the maximal profit the principal expects to earn, net of what he is able to generate himself, is given by $u_P^+ - u_P^0$. In the following, we therefore assume that the wage cannot exceed this threshold, i.e.

$$w(\mu) < u_P^+ - u_P^0 \qquad \forall \mu \in (0, 1)$$

and $w(1) \leq u_P^+ - u_P^0$.[17] Furthermore, we assume that w is continuous and weakly convex, i.e.[18]

$$w(tx + (1-t)y) \leq tw(x) + (1-t)w(y) \qquad \forall x, y, t \in [0, 1].$$

biased agent to choose the loyal action, x_l, with positive probability (if the decision is delegated) is the increased wage in Period 2 because the principal may centralize in Period 2 no matter which project the agent chooses. In that case, the principal would be better off paying only $w(p)$ – such that the agent leaves – and employing a new agent. However, assuming the agent to be needed for the implementation of the project and that employing a new or no agent is costly, with cost being larger than $w(\eta) - w(p)$ (where η is the delegation threshold without evaluation; see Section 1.1.2 for details), ensures that also these special cases remain tractable.

[15] In case of private evaluations, we assume that downward adjustments of wages require publicly verifiable information (public evaluation) and upward wage adjustments are never profitable for the principal as the agent's outside option remains fix.

[16] Assuming the agent to be liability constrained does not change the results of the analysis which essentially rely on the monotonicity of wages.

[17] Here we implicitly assume that all principals face the same constraints so that this threshold will not exclude any of them from the market.

[18] From a technical point of view, if wages are assumed to be competitive and thus reflecting the expected marginal product of a worker given the belief about his loyalty, one would have to pay negative wages in case $p \leq \eta$. If agents are protected by limited liability, principals are forced to pay a constant wage of zero for low prior beliefs, and a linearly increasing wage for $p > \eta$. This case provides an example for a weakly convex wage schedule. Besides lacking a real-world justification, concave wage functions do not provide any new results, which is why they are not considered.

In the sequel, we divide the agent's employment with the principal into two equal periods – before and after an evaluation takes place – and refer to these as Period 1 and Period 2, respectively. In the baseline case, where the agent is not evaluated, these two periods are equivalent and the respective payoffs are considered to be generated in each period.

Summing up, with the above specifications, total per period payoffs for the principal can be specified as

$$U_P(x, \mu) = u_P(x) - w(\mu) \ ,$$

and the corresponding expression for the agent is given by

$$U_\tau(x, \mu) = u_\tau(x) + w(\mu) \ .$$

1.1.2 No Evaluation

To begin with, consider the case where the principal does not evaluate the agent so that reputation is not an issue. In this case, the principal has to decide whether or not to delegate decision authority to the agent at the beginning of the employment phase.

Obviously, without interim evaluation, the principal cannot observe the agent's choice after Period 1 so that the agent always chooses his preferred project x_τ if the decision is delegated to him. Accordingly, the principal's expected per period benefit from project choice in case of delegation is given by

$$E[u_P(x_\tau)] = pu_P^+ + (1-p)u_P^-.$$

This immediately leads to Proposition 1.

Proposition 1. *Without intermediate evaluation, the principal prefers delegation to centralization if*

$$p \geq \eta := \frac{u_P^0 - u_P^-}{u_P^+ - u_P^-}.$$

1.1.3 Equilibria With Evaluation

As a next step, consider the case where the principal evaluates the agent in the middle of his employment phase.

Private Evaluation.

Recall that by assumption neither the project choice nor the allocation of authority in either period can be contractually fixed ex ante. However, we now assume that the principal privately observes the agent's project choice after the first half of the employment (given that it has been delegated to the agent). In this case, the agent's outside option remains unchanged as the market remains ignorant about the outcome of the evaluation so that the agent's wage in Period 2 is again given by $w(p)$. The principal, however, might reconsider his delegation decision at the beginning of Period 2.

More specifically, in Period 1, the principal decides whether or not to delegate the decision to the agent. The action space of the principal, thus, is given by $A_P = \{D, C\}$, consisting of delegation, D, and centralization, C. If the principal retains his decision authority, C, he always takes the default decision and his belief about the agent's loyalty is not updated. If the principal delegates his decision authority to the agent, D, the agent decides which project to implement. At the end of Period 1, the principal's profit is realized and the principal is informed about the agent's project choice.

In Period 2, the principal faces essentially the same delegation problem as in the first except that he now can utilize the agent's observed behavior from Period 1 to update his belief about the type of the agent. The updated beliefs are denoted as follows:

$$\begin{aligned} \mu_+ &:= \Pr(\tau = l | x = x_l) \\ \mu_- &:= \Pr(\tau = l | x = x_b). \end{aligned}$$

The resulting signaling game that arises after delegation in Period 1 is illustrated in Figure 1.1.

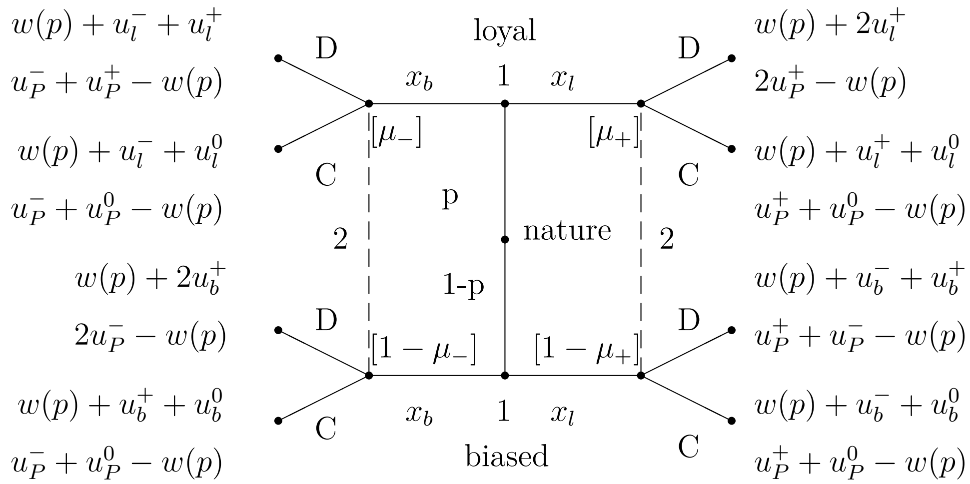

Figure 1.1: Graphical illustration of the strategic situation with private evaluation after a delegation decision of the principal in Period 1. Player 1, the sender, is the agent and player 2, the receiver, is the principal. First period wages are neglected as they are of no strategic relevance.

In the sequel, we analyze the subgame that arises after delegation by the principal in Period 1 while neglecting Period-1-wages (as these have no strategic impact). In doing so, we denote the action the principal chooses in Period 2 upon observing x by $a_P(x)$. A strategy of the principal in the subgame, thus, consists of a pair $s_P = (a_P(x_l), a_P(x_b))$. A strategy for the agent in the resulting subgame, in turn, comprises the choice of an action both in case of delegation in Period 1 (which has occurred by assumption) and in Period 2. As it is strictly dominant for the agent to choose his preferred action whenever the decision is delegated in Period 2, we focus the analysis on the agent's strategy in Period 1, denoted by s_τ, which is given by specifying $x \in \{x_b, x_l\}$.

Noting that it is dominant for the principal to choose D (C) upon observing x_l (x_b resp.) if $\mu_+ \geq \eta$ ($\mu_- \leq \eta$) and defining

$$\Delta_b := u_b^+ - u_b^-,$$

i.e. Δ_b is the biased agent's utility differential between choosing x_b and x_l, we obtain the following equilibria;[19] see Figure 1.2 for an illustration.[20]

[19] If $\mu = \eta$, the principal is indifferent which is why the dominance is not strict.

[20] The structure of these equilibria is robust to introducing some noise into the principal's observation

Lemma 1. *For the signaling game depicted in Figure 1.1, the following three types of equilibria are compatible with the Intuitive Criterion (Cho and Kreps, 1987):*

- *If* $\Delta_b < u_b^+ - u_b^0$ *and* $p > \eta$*:*
  Pooling (pool(D,C)): $s_P = (D, C),\ s_l = s_b = x_l$
- *If* $\Delta_b < u_b^+ - u_b^0$ *and* $p \leq \eta$*:*
  Principal and biased agent randomize (mix2(λ_P*,C)):*
  $s_P = \lambda_P D + (1 - \lambda_P)C,\ s_l = x_l,\ s_b = \lambda x_l + (1 - \lambda)x_b,\ (\lambda = 1 \text{ if } p = \eta)$
- *If* $\Delta_b = u_b^+ - u_b^0$*:*
  Biased agent randomizes (mix1(D,C)): $s_P = (D, C),\ s_l = x_l,\ s_b = \lambda x_l + (1 - \lambda)x_b$
- *If* $\Delta_b > u_b^+ - u_b^0$*:*
  Separating (sep(D,C)): $s_P = (D, C),\ s_l = x_l,\ s_b = x_b$

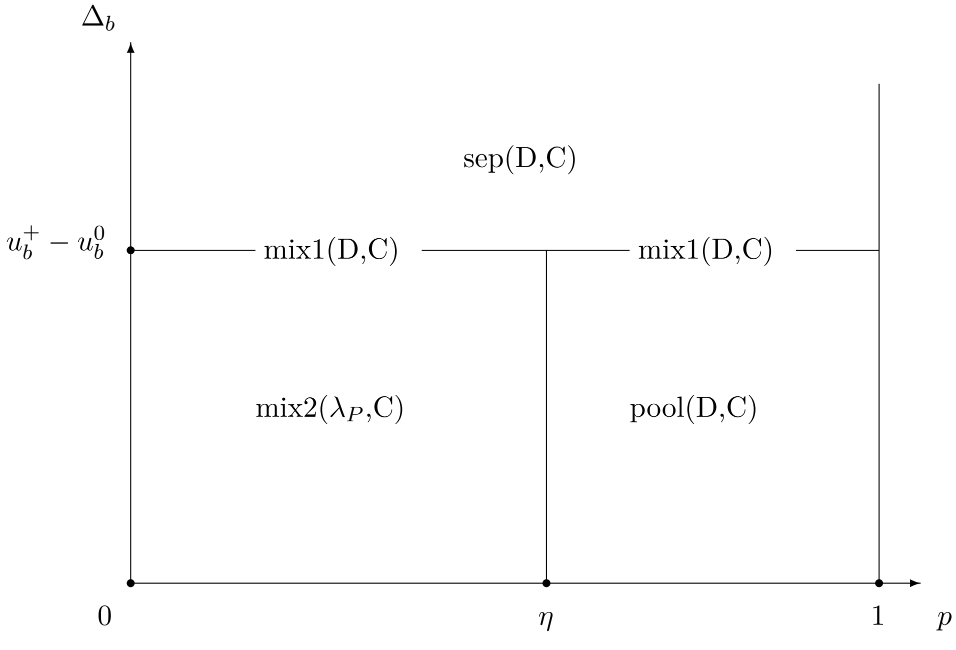

Figure 1.2: Equilibria of the signaling game with private evaluation depending on the prior belief p and the biased agent's utility differential, Δ_b. Note that outside of the null-set where the mix1(D,C) equilibrium or the mix2(λ_P,C) equilibrium with $p = \eta$ is played, all equilibria are unique for a given pair of parameters (p, Δ_b).

of the agent's project choice.

The main point to note here is that the biased agent's incentive to choose his preferred project in Period 1 is stronger the larger his utility from centralization in Period 2, i.e. u_b^0, is compared to his utility from taking an opportunistic choice in Period 1, u_b^-. Put differently, an equilibrium where the biased agent separates is more likely the larger the difference between the biased agent's utility differential Δ_b and $u_b^+ - u_b^0$. The principal on the other hand chooses to centralize in Period 2 whenever he observes a separating behavior. Upon observing the loyal project, though, he again delegates the decision in Period 2 – if his prior belief about the agent being loyal is high enough.

Public Evaluation.

Finally, we consider the case where the principal evaluates the agent and reveals this information to the market. In this case, the agent's outside option might change as a function of μ, the principal's and the market's posterior belief about the agent's loyalty after observing his project choice. More specifically, with public evaluation, the agent's wage depends on the updated beliefs, μ_+ and μ_-; the signaling game which arises after delegation in Period 1 is illustrated in Figure 1.3.

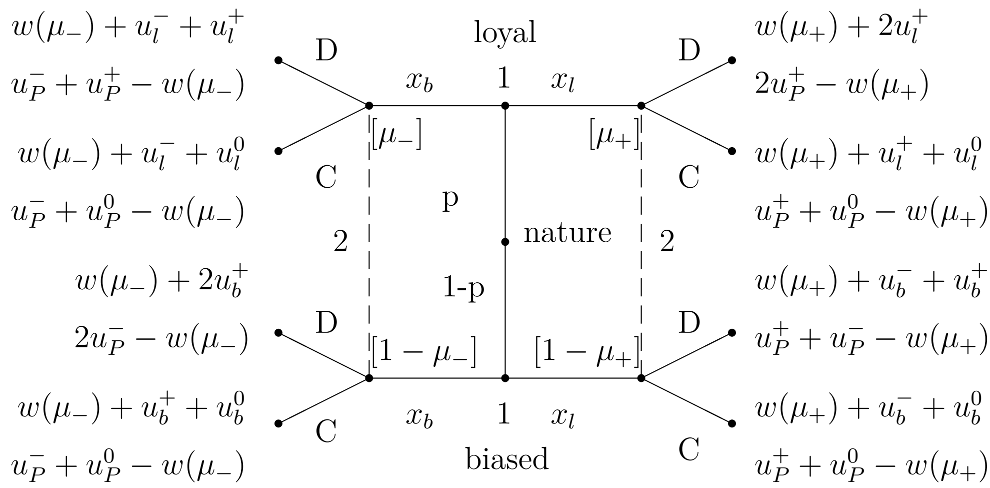

Figure 1.3: Graphical illustration of the strategic situation with public evaluation after a delegation decision of the principal in Period 1. Player 1, the sender, is the agent and player 2, the receiver, is the principal. First period wages are neglected as they are of no strategic relevance.

The analysis of the resulting signaling game is analogous to the one in case of private evaluation. The results are provided in the following lemmata.

Lemma 2. *If $p > \eta$, the following three types of equilibria of the signaling game depicted in Figure 1.3 are compatible with the Intuitive Criterion:*

- *If $\Delta_b \leq w(p) + u_b^+ - u_b^0$:*
  Pooling: (pool(D,C)) $s_P = (D, C)$, $s_l = s_b = x_l$
- *If $w(p) + u_b^+ - u_b^0 < \Delta_b < w(1) + u_b^+ - u_b^0$:*
  Biased agent randomizes (mix1(D,C)): $s_P = (D, C)$, $s_l = x_l$, $s_b = \lambda x_l + (1 - \lambda) x_b$
- *If $\Delta_b \geq w(1) + u_b^+ - u_b^0$:*
  Separating (sep(D,C)): $s_P = (D, C)$, $s_l = x_l$, $s_b = x_b$

Thus, similar to the case with private evaluation, we find that, for $p > \eta$, the biased agent is more likely to imitate the loyal agent if his utility differential, Δ_b, is small. However, with public evaluation the biased agent benefits from a loyal project choice after delegation in Period 1 not only through additional delegation in Period 2 but also through an increased wage in Period 2. Thus, pooling incentives and, hence, the area where the biased agent imitates the loyal one are increased. Moreover, there is a non-degenerate area where the biased agent trades off wage and project-choice based incentives and therefore randomizes between project choices (mix1(D,C)).

For the case of low prior beliefs, considered in the lemma below, we also observe a larger area for the mix1(D,C) equilibrium. Furthermore, if the biased agent's utility differential, Δ_b, is sufficiently small, his incentive to opt for his preferred project is dominated by an increased wage resulting from the loyal project choice, even if the principal centralizes either way. The principal, in turn, is not willing to delegate in Period 2, even after observing his preferred project choice as his prior belief about the agent's loyalty is too small.

Lemma 3. *If $p \leq \eta$, the following three types of equilibria of the signaling game depicted in Figure 1.3 are compatible with the Intuitive Criterion:*

- *If $\Delta_b \leq w(p)$:*
  Pooling (pool(C,C)): $s_P = (C, C)$, $s_l = x_l$, $s_b = x_b$, (principal randomizes if $p = \eta$)

- *If $w(p) < \Delta_b \leq w(\eta)$:*
  Biased agent randomizes (mix1(C,C)): $s_P = (C, C)$, $s_l = x_l$, $s_b = \lambda x_l + (1-\lambda)x_b$

- *If $w(\eta) < \Delta_b < w(\eta) + u_b^+ - u_b^0$:*
  Principal and biased agent randomize (mix2(λ_P,C)):
  $s_P = \lambda_P D + (1-\lambda_P)C$, $s_l = x_l$, $s_b = \lambda x_l + (1-\lambda)x_b$, ($\lambda = 1$ if $p = \eta$)

- *If $w(\eta) + u_b^+ - u_b^0 \leq \Delta_b < w(1) + u_b^+ - u_b^0$:*
  Biased agent randomizes (mix1(D,C)): $s_P = (D, C)$, $s_l = x_l$, $s_b = \lambda x_l + (1-\lambda)x_b$

- *If $\Delta_b \geq w(1) + u_b^+ - u_b^0$:*
  Separating (sep(D,C)): $s_P = (D, C)$, $s_l = x_l$, $s_b = x_b$

Note that with public evaluation, all equilibria outside of the null-set $\{p = \eta\} \cap \left\{\Delta_b \leq w(\eta) + u_b^+ - u_b^0\right\}$ are unique for a given pair of parameters (p, Δ_b). Figure 1.4 illustrates the resulting equilibria and their relation to the underlying wage schedule.

In contrast to the situation with private evaluation, the biased agent now benefits from pooling with the loyal type not only because he may decide on the project in Period 2 but also because he receives a higher wage in the later period. Accordingly, the pooling equilibrium is played up to larger values of Δ_b if the prior belief is high. The mixed equilibrium, in turn, extends to a non-null set for $p \geq \eta$ because the biased agent's wage, given by $w(\mu_+) = u_b^0 - u_b^-$, endogenously adapts to changes in Δ_b, thereby keeping him indifferent for a continuum of values of Δ_b. Thus, if the prior belief is low, the principal might even choose to centralize in Period 2 independent of the agent's decision as the wage schedule entails some incentive to pool for the biased agent – in contrast to the case where evaluation is private.

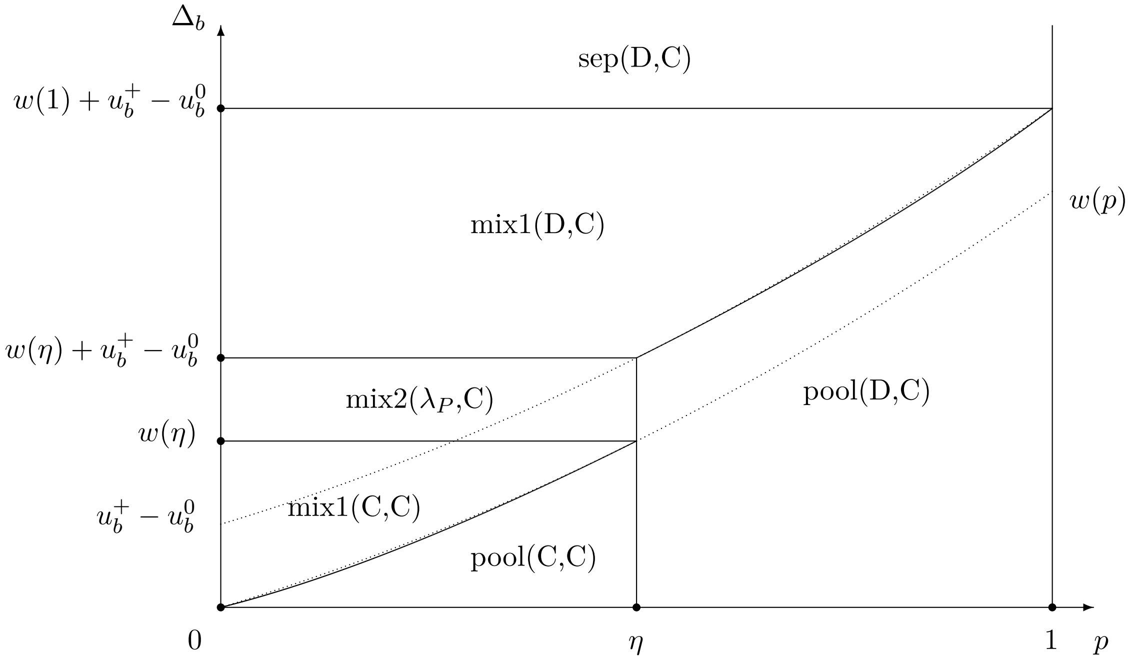

Figure 1.4: Equilibria of the signaling game with public evaluation depending on the prior belief p and the biased agent's utility differential, Δ_b – for a given wage schedule (indicated by the lower curve). The upper curve depicts the wage schedule increased by $u_b^+ - u_b^0$. Solid lines determine the borders of the respective equilibrium area.

1.1.4 The Principal's Behavior

We proceed to analyze whether the principal prefers to delegate the project decision to the agent in Period 1, depending on his prior belief about the agent's loyalty and the utility parameters. Having done so, we investigate if the principal benefits from an evaluation, given the policy of information disclosure, and whether he chooses to disclose an evaluation if he is allowed to do so.

The Principal's Delegation Decision

In case the principal retains authority in Period 1, the belief about the agent's loyalty cannot be updated and the principal delegates in Period 2 if $p \geq \eta$. Thus, the overall payoff of the principal in case of centralization, depending on p, is given by

$$E[U_P|C] = \begin{cases} 2u_P^0 - w(p) & \text{if } p < \eta, \\ u_P^0 + pu_P^+ + (1-p)u_P^- - w(p) & \text{else,} \end{cases}$$

where the first-period wage $w(p)$ is again left out.

In the sequel, we compare the principal's payoffs with and without evaluation depending on whether the evaluation is public or private.

Private Evaluation.

In case of a private evaluation, it is straightforward to show that the principal is more likely to delegate in the first period compared to the situation without evaluation.

Proposition 2. *If the principal can privately evaluate the agent, he delegates more often in Period 1 than in an interaction without evaluation. In particular, the principal delegates for all $p \geq p_U^*(\Delta_b)$, where p_U^* is given by*

$$p_U^*(\Delta_b) = \begin{cases} \eta^2 & \text{if } \Delta_b < u_b^+ - u_b^0, \\ \in \left[\eta^2, \frac{\eta}{2-\eta}\right] & \text{if } \Delta_b = u_b^+ - u_b^0, \\ \frac{\eta}{2-\eta} & \text{if } \Delta_b > u_b^+ - u_b^0. \end{cases}$$

Furthermore, $\eta^2 < \frac{\eta}{2-\eta} < \eta$, thus

- *the principal delegates for a larger range of prior beliefs than without evaluation, in which case he delegates only for $p \geq \eta$, and*
- *the principal delegates more the lower the biased agent's concern about the decision, Δ_b, is.*

Note that this result is rather intuitive. The biased agent has an incentive to deviate from his preferred option in Period 1 in order to signal loyalty if he does not suffer too much from an opportunistic choice in the first period.[21] By delegating the project decision to the agent, the principal benefits either from learning the agent's type (if Δ_b is high), or from more aligned decision making (if Δ_b is low). Thus, in either case the principal is more likely to delegate the decision to the agent than in the case without evaluation.

Moreover, aligning the biased agent's preferences is in fact more beneficial to the principal than learning his type. Therefore, the principal is more likely to delegate in the pooling equilibrium than in the separating equilibrium – indicated by $\eta^2 < \frac{\eta}{2-\eta}$.

Public Evaluation.

If the evaluation is observable to the market, the principal has to pay a higher wage to the biased agent in a pooling equilibrium than in a separating one. Still the principal is more likely to delegate the project choice than in the case without evaluation.

Proposition 3. *If evaluations are public, the principal delegates more often in Period 1 if he evaluates the agent than in an interaction without evaluation. In particular, he delegates if $p \geq p_O^*(\Delta_b)$, where the threshold $p_O^*(\Delta_b)$ is implicitly defined via*

$$p_O^* = \delta_{w,\mu}(p_O^*) = \frac{[u_P^0 - u_P^- - w(p_O^*)]\mu}{(u_P^+ - u_P^-)\max\{\mu,\eta\} + (u_P^+ - u_P^0) - w(\mu)}$$

and $\mu(\Delta_b) \in [0,1]$ is given by

$$\mu(\Delta_b) = \begin{cases} w^{-1}(\Delta_b) & \text{if } \Delta_b \leq w(\eta), \\ \eta & \text{if } w(\eta) < \Delta_b < w(\eta) + u_b^+ - u_b^0, \\ w^{-1}(\Delta_b - (u_b^+ - u_b^0)) & \text{if } w(\eta) + u_b^+ - u_b^0 \leq \Delta_b < w(1) + u_b^+ - u_b^0, \\ 1 & \text{if } \Delta_b \geq w(1) + u_b^+ - u_b^0, \end{cases}$$

Furthermore, $\delta_{w,\mu}(p) \leq \eta$ for all $p \in [0,1]$, so the principal delegates for a larger range of prior beliefs than without evaluation, in which case he delegates the decision only for $p \geq \eta$.

[21] Cast in the ability interpretation indicated in Remark 1, the biased agent would invest a high effort/cost, which is individually suboptimal in the setting without evaluation, in order to imitate the more able agent.

Note that the notion of the inverse wage function is valid in the above definition even for weakly increasing wages, as the domain of Δ_b is defined by means of the image of w.

The main point to note here is that the biased agent again has an incentive to deviate from his preferred option in Period 1 in order to signal loyalty – essentially as in the case with private evaluation. However, in exchange for this, he now not only is more likely to be granted the right to decide over the project but also receives a higher wage in Period 2. In combination, this induces more aligned decisions in the first period of the relationship and, hence, more delegation.

Furthermore, even if wages matter, the principal still benefits more from aligning the biased agent's preferences than from learning his type and paying him less in the later stage. Therefore, we obtain a similar result as in the case of private evaluation: The principal is more likely to delegate if the probability for the biased agent to choose the loyal project is higher.

Finally, we can show that the delegation threshold increases in Δ_b, that is, the principal benefits more from aligning the biased agent's preferences than from learning his type and paying him less in the later stage.

Proposition 4. *All other things equal, the principal delegates more the lower the biased agent's concerns about the decision are, i.e. the delegation threshold* $p_O^*(\Delta_b)$ *fulfilling* $p_O^* = \delta_{w,\mu}(p_O^*)$ *increases in* Δ_b*.*

Assuming wages to be linear in μ, the increasing delegation threshold can be determined explicitly, as stated in Corollary 1.

Corollary 1. *If wages are linear, i.e.* $w(\mu) = \mu W$*, the delegation threshold in Proposition 3 is given by*

$$p_O^*(\Delta_b) = \begin{cases} \frac{\eta}{W}\Delta_b & \text{if } \Delta_b \leq \eta W, \\ \eta^2 & \text{if } \eta W < \Delta_b < \eta W + u_b^+ - u_b^0, \\ \frac{(u_b^0 - u_b^-)\eta}{(u_b^0 - u_b^-) + (1-\eta)W} & \text{if } \eta W + u_b^+ - u_b^0 \leq \Delta_b < W + u_b^+ - u_b^0, \\ \frac{\eta}{2-\eta} & \text{if } W + u_b^+ - u_b^0 \leq \Delta_b. \end{cases}$$

A Comparison of Public and Private Evaluation.

A natural question that arises from the above analysis is how the observability of the evaluation influences the principal's delegation decision. At first sight, it might seem intuitive to assume that the principal is less likely to delegate if the evaluation is public in order not to reveal too much information to the market and thereby possibly increasing the agent's outside option. In the present setting, however, this may not be true. In fact, exposing the agent's behavior to the market can be beneficial to the principal if this changes the biased agent's behavior. However, as we will see below, this effect crucially depends on the shape of the wage schedule.

Proposition 5 shows that, if the biased agent's concern about the project decision, Δ_b, is high enough, the principal is more likely to delegate if the evaluation is not observed by the market.

Proposition 5. *If w is convex, there exists a threshold $\rho \in [0, w(1) + u_b^+ - u_b^0]$, such that for $\Delta_b \geq \rho$ delegation is more likely if the evaluation is private than if it is public.*

Note that the level of the threshold strongly depends on the shape of the wage function and decreases with its convexity. In fact, for "very convex" wage functions, ρ can be smaller than $w(\eta)$, such that public evaluation grants more discretion to the agent only if his concern about the project choice is negligible.

Figure 1.5 illustrates the results from Proposition 5; for the case of public evaluation, it distinguishes two cases:

- w is "not too convex", i.e. $\frac{\eta}{2-\eta} w(\eta) - \eta w(\frac{\eta}{2-\eta}) \leq (1-\eta)\frac{\eta}{2-\eta}(u_P^+ - u_P^0)$,
- w is "very convex", i.e. $\frac{\eta}{2-\eta} w(\eta) - \eta w(\frac{\eta}{2-\eta}) \geq (1-\eta)\frac{\eta}{2-\eta}(u_P^+ - u_P^0)$,

where p_{VC}^* refers to the "very convex" case and p_{NC}^* to the "not too convex" case. These two thresholds are shown as solid lines, whereas the delegation threshold in case of private evaluation is depicted as a dashed line. The benchmark threshold where no evaluation takes place is indicated by a dotted line.

Note that the shape of $w(\mu)$ does not allow for any conclusions about the shape of $p^*_O(\Delta_b)$ so that Figure 1.5 is just an illustration of the delegation thresholds.

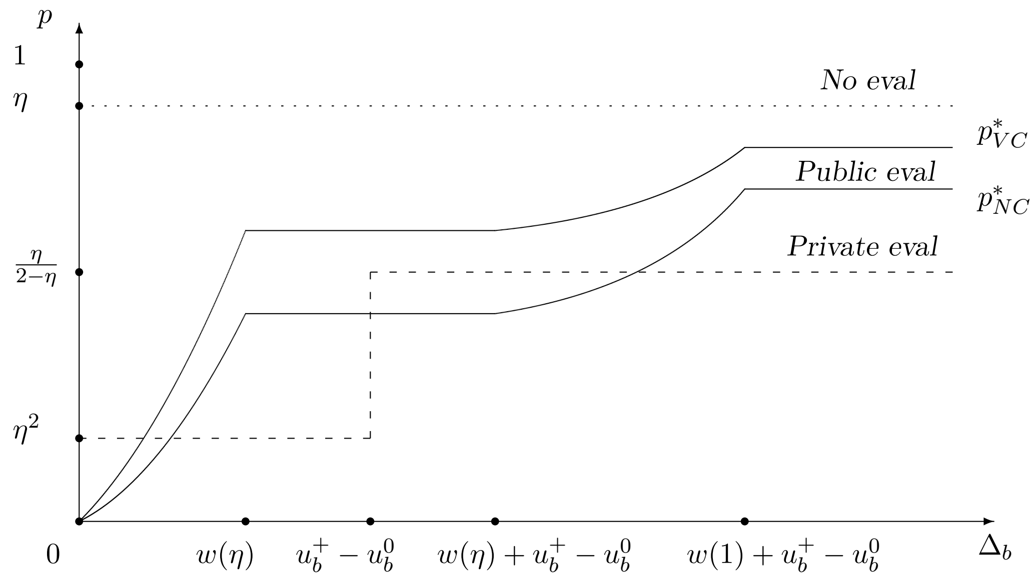

Figure 1.5: Illustration of the delegation threshold in case of public evaluation, p^* (solid lines). The upper graph, p^*_{VC}, refers to the "very convex" case, the lower graph, p^*_{NC}, to the "not too convex" case. The dashed (dotted) line illustrates the delegation threshold in case of private evaluation (no evaluation); here we assume $w(\eta) < u_b^+ - u_b^0$ (the opposite case looks similar). Except for the constant parts of p^*, the slope and curvature of the graph are only exemplary – the shape cannot be inferred from the shape of $w(\mu)$.

Moreover, considering linear wages, there is weakly more delegation with public evaluation.

Corollary 2. *If wages are linear, delegation is (weakly) more likely if the evaluation is public than if it is private.*

Figure 1.6 illustrates the results from Corollary 2. In particular, it shows the delegation thresholds depending on the evaluation regime.

The Principal's Profit

In the following, we analyze the rent effects for the principal resulting from an intermediate evaluation. The main point to note here is that a separation of agents as well as an alignment of the agents' behavior (in Period 1), which only occur when evaluation takes place, can be beneficial for the principal.

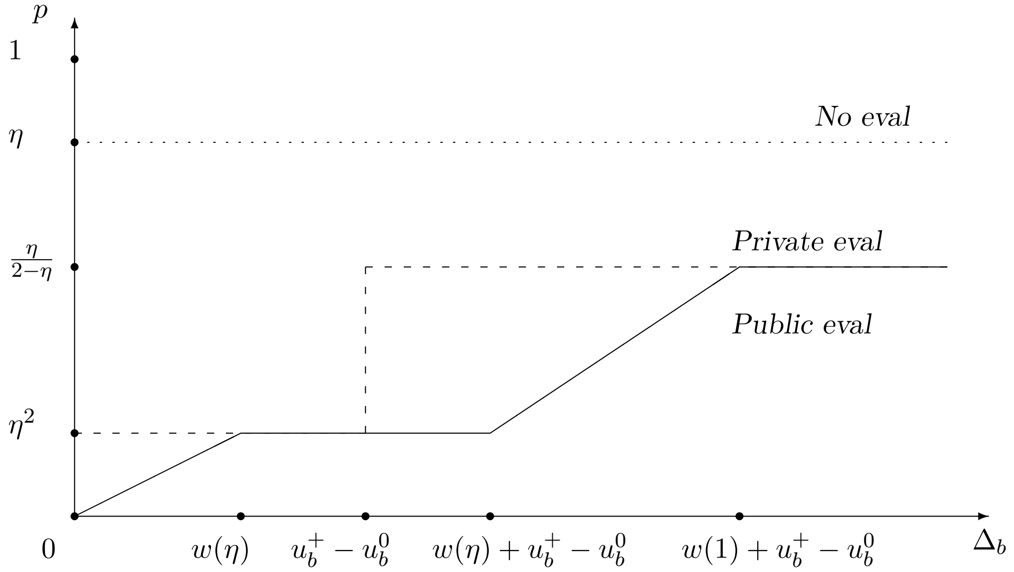

Figure 1.6: Illustration of the delegation threshold in the three cases of no evaluation (dotted line), private evaluation, p_U^* (dashed line), and public evaluation, p_O^* (solid line). Here, wages are linear and $w(\eta) < u_b^+ - u_b^0$ is assumed; the opposite case looks similar.

Private Evaluation.

For the principal, an intermediate evaluation can be profitable either because he learns the agent's type or because the biased agent imitates the loyal one by choosing the principal's preferred project in the first period. If the evaluation is not observable to the market, outside options do not change and the principal does not have to reimburse the agent for an opportunistic choice – as opposed to the case where the evaluation is public. Thus, the following proposition is immediate.

Proposition 6. *If the interim evaluation is private, the principal is always (weakly) better off by evaluating the agent.*

In order to see this, consider any equilibrium where $p > \eta$, the principal plays (D, C) and the biased agent plays a mixed strategy $(\lambda, 1-\lambda)$ with $\lambda \in [0, 1]$. Then the principal's expected total profits with (unobservable) private evaluation, Π_U, net of the profits with no evaluation, Π_N, are given by

$$E[\Pi_U - \Pi_N] = (1-p)\left[\lambda(u_P^+ - u_P^0) + (u_P^0 - u_P^-)\right].$$

As this term is non-negative for all $\lambda \in [0,1]$, it is immediate that the principal always prefers to evaluate for $p > \eta$ if the result is not observable to the market. A similar argument holds for intermediate prior beliefs where the principal still decides to delegate in Period 1 if a private evaluation takes place. If the prior belief is too low for the principal to delegate in Period 1, he is indifferent between evaluating or not.

Public Evaluation.

If the interim evaluation is observable to the market, the principal has to later compensate the biased agent for an earlier opportunistic behavior. Thus, it is no longer true in general that evaluating the agent benefits the principal. In fact, whether the biased agent's aligned behavior in Period 1 outweighs the possible losses in terms of higher Period-2-wages depends on which equilibrium is played in the respective case.

In particular, in case of high prior beliefs, the principal plays (D, C) and the biased agent plays a mixed strategy $(\lambda, 1-\lambda)$ with $\lambda \in [0,1]$ in any equilibrium. Thus, the principal's expected total profits with (observable) public evaluation, Π_O, net of the profits with no evaluation, Π_N, are given by

$$E[\Pi_O - \Pi_N] = (1-p)\left[\lambda(u_P^+ - u_P^0) + (u_P^0 - u_P^-)\right] - \left[\left[p + (1-p)\lambda\right] w(\mu_+) - w(p)\right],$$

where $\lambda \in [0,1]$ denotes the probability of the biased agent choosing x_l.

The above formula illustrates that the principal trades off his earnings from the agent's project choice in Period 1, which are increasing in the biased agent's probability to imitate the loyal one, against the wage payment to the agent, also increasing in λ.

In particular, in a pooling equilibrium, evaluating the agent publicly does not change his outside option, while the principal still profits from the biased agent imitating the loyal one. Thus, the principal decides to evaluate the agent in this case. However, as the probability of the biased agent choosing the principal's preferred project decreases, the expected wage payment to the agent increases faster than the corresponding gain from the agent's project choice due to convexity of wages. Thus, the principal prefers to evaluate the agent for $p > \eta$ whenever there is "enough pooling" and his scope of evaluation depends on the shape of the wage function.

For low prior beliefs, the principal centralizes authority in Period 1 if no evaluation takes place and the principal prefers to evaluate the agent if and only if he delegates the project choice in case of a public evaluation. Proposition 7 summarizes these findings.

Proposition 7. *If $p > \eta$ and the interim evaluation is observable to the market, the principal is better off by evaluating the agent if and only if*

$$pw(\mu_+) - \mu_+ w(p) \leq p(1-\mu_+)(u_P^+ - u_P^0) + (1-p)\mu_+(u_P^0 - u_P^-).$$

In particular, for all $p > \eta$, there exists a threshold $\rho(p) > 0$ such that an evaluation is preferred for all $\Delta_b \leq \rho(p)$, i.e. if the biased agent's concern about the project choice is low enough.
If $p \leq \eta$, the principal weakly prefers to evaluate the agent if the evaluation is public.

Remark 2. *The posterior belief $\mu_+ \in [0,1]$, in case a public evaluation takes place, is endogenously given by the model parameters. In particular, if $p \leq \eta$, we have*

$$\mu_+(\Delta_b) = \begin{cases} p & \text{if } \Delta_b \leq w(p), \\ w^{-1}(\Delta_b) & \text{if } w(p) < \Delta_b \leq w(\eta), \\ \eta & \text{if } w(\eta) < \Delta_b < w(\eta) + u_b^+ - u_b^0, \\ w^{-1}(\Delta_b - (u_b^+ - u_b^0)) & \text{if } w(\eta) + u_b^+ - u_b^0 \leq \Delta_b < w(1) + u_b^+ - u_b^0, \\ 1 & \text{if } \Delta_b \geq w(1) + u_b^+ - u_b^0, \end{cases}$$

and if $p > \eta$, this yields

$$\mu_+(\Delta_b) = \begin{cases} p & \text{if } \Delta_b \leq w(p) + u_b^+ - u_b^0, \\ w^{-1}(\Delta_b - (u_b^+ - u_b^0)) & \text{if } w(p) + u_b^+ - u_b^0 < \Delta_b < w(1) + u_b^+ - u_b^0, \\ 1 & \text{if } \Delta_b \geq w(1) + u_b^+ - u_b^0. \end{cases}$$

Moreover, again assuming wages to be linear, it is easy to see that $pw(\mu_+) - \mu_+ w(p) = 0$. Furthermore, as $p(1-\mu_+)(u_P^+ - u_P^0) + (1-p)\mu_+(u_P^0 - u_P^-) \geq 0$, the following corollary is immediate.

Corollary 3. *If wages are linear and the interim evaluation is observable to the market, the principal is (weakly) better off by evaluating the agent.*

A Comparison of Public and Private Evaluation.

Although we have seen that the principal always prefers to evaluate the agent for high prior beliefs if the evaluation is private – which is not necessarily the case for public evaluations – this does not allow us to conclude that the principal would always prefer to conceal the evaluation if he had the choice. In fact, for a given prior $p > \eta$ and a given utility differential Δ_b, the behavior of the biased agent might change if his evaluation is public because wage considerations then play a role. Thus, equilibria played in the two regimes might be different. Figure 1.7 illustrates this change of the biased agent's behavior for $p > \eta$.

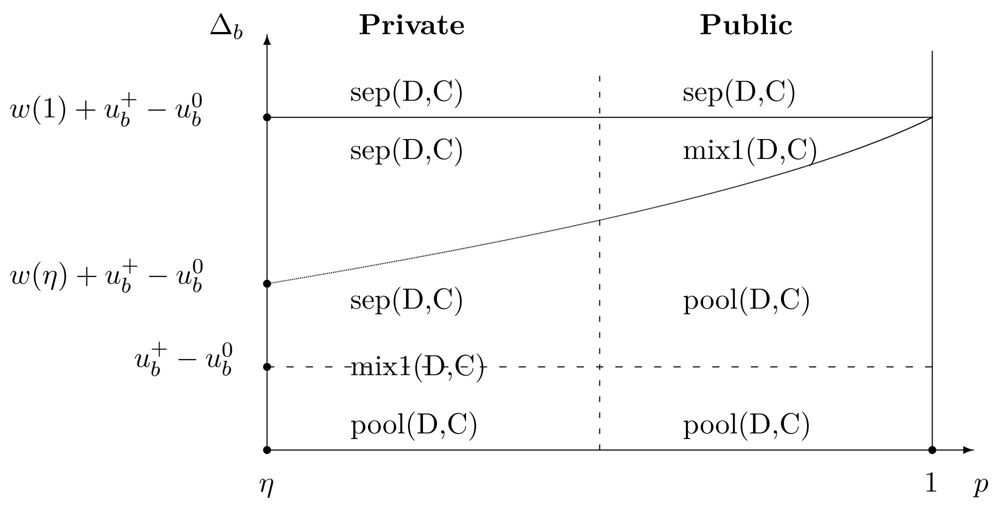

Figure 1.7: Illustration of the different equilibria played in case $p > \eta$. The left-hand side states the equilibria played in a private regime, whereas the right-hand side indicates which equilibrium is played in the same area of the graph if the evaluation is public.

For example, if the biased agent pools in the public regime in order not to lose his wage, he separates in a private regime for the same parameter values. In this case, the principal is better off in the public regime where he gets a profit of

$$E[U_P|O, pool(D,C)] = 2pu_P^+ + (1-p)(u_P^+ + u_P^-) - w(p),$$

compared to

$$E[U_P|U, sep(D,C)] = 2pu_P^+ + (1-p)(u_P^- + u_P^0) - w(p)$$

in the private case; here O and U again indicate public (observable) and private (unobservable) evaluation, respectively. Thus, whenever the separating equilibrium is played in the case of private evaluation, the principal is better off with public evaluation as long as this results in "enough pooling." And this occurs if the biased agent's utility differential is relatively low as stated in Proposition 8.

Proposition 8. *If $p > \eta$, there exists a threshold $\rho(p) \in [w(p) + u_b^+ - u_b^0, w(1) + u_b^+ - u_b^0]$, such that the principal reveals the evaluation for $\Delta_b < \rho$ and conceals it for $\Delta_b > \rho$.*
If $p \leq \eta$ and Δ_b is such that the decision is delegated both in case of private and public evaluation, then there exist thresholds $\rho_1(p) \in [w(p), w(\eta)]$ and $\rho_2(p) \in [\rho_1(p), w(1) + u_b^+ - u_b^0]$, such that the principal reveals the evaluation if $\Delta_b < \rho_1(p)$ and conceals it if $\Delta_b > \rho_2(p)$. Furthermore, if $w(\eta) \leq u_b^+ - u_b^0$ and w is "very convex", i.e.

$$pw(\eta) - \eta w(p) \geq p(1-\eta)(u_P^+ - u_P^0),$$

we have $\rho_1(p) = \rho_2(p)$.

However, if wages are assumed to be linear, the principal pays an expected wage of $w(p)$ to the agent in any case. Hence, the biased agent's increased incentive to imitate the loyal agent in case of public evaluation is no longer costly for the principal.

Corollary 4. *For any prior belief p, the principal weakly prefers to reveal the evaluation to the market if wages are linear.*

1.1.5 Welfare

In the following, we consider the players' welfare obtained with evaluation and compare the results to the interaction without evaluation. As the results differ depending on whether delegation occurs without evaluation, we split the subsequent discussion in two parts: to begin with, we analyze the case of high priors, $p > \eta$, where we see delegation if no evaluation takes place, and then move on to the case of low priors, $p \leq \eta$, where there is no delegation without evaluation.

The Case of High Priors: $p > \eta$

First, we analyze aggregate welfare, followed by a discussion of comparative statics with regard to utility parameters and the wage schedule. As the main analysis is essentially the same whether evaluation is public or not – as transfers from principal to agent cancel out – the following propositions hold for both regimes unless otherwise stated.

Aggregate Welfare.
As we have seen in Section 1.1.4, introducing an interim evaluation may be beneficial for the principal, because both a separation of agents and an alignment of the agents' behavior (in Period 1), which only occur with evaluation, can increase his profit.

Regarding the agent, the appreciation of the evaluation depends on his type. In particular, the loyal agent weakly prefers being evaluated: he always chooses the principal's preferred project in Period 1; and with public evaluation, he may not only gain discretion but also a higher wage in Period 2 (except a pooling equilibrium is played). The biased agent, by contrast, is always better off without evaluation: with evaluation, choosing his preferred project is always punished by centralization (and a wage drop to zero if the evaluation is public), while without evaluation the decision is always delegated.

Formally, accounting for both players and both types of the agent, expected aggregate welfare with evaluation, W_E, net of expected aggregate welfare with no evaluation, W_N, is given by

$$E[W_E - W_N] = (1-p)\lambda[(u_P^+ - u_P^0) - (u_b^0 - u_b^-)] + (1-p)[(u_P^0 - u_P^-) - (u_b^+ - u_b^0)],$$

with $\lambda = 0$ if $u_b^0 - u_b^- \geq w(1)$ (separating equilibrium) and $\lambda = 1$ if $u_b^0 - u_b^- \leq w(p)$ (pooling equilibrium).

Thus, assuming that the principal's benefits from an optimal project decision outweigh the respective benefits of the biased agent, i.e. assuming

$$u_P^+ - u_P^- =: \Delta_P \geq \Delta_b = u_b^+ - u_b^-,$$

welfare is increased by an evaluation whenever a pooling equilibrium is played. Whether evaluating an agent is efficient in general, though, depends on the relation of the principal's profit from centralization compared to a biased project choice, $u_P^0 - u_P^-$, to the biased agent's net gain from deciding over the project, $u_b^+ - u_b^0$.

Proposition 9. *If $p > \eta$, the comparison of expected welfare with and without evaluation yields the following results:*

- *If $u_P^0 - u_P^- \geq u_b^+ - u_b^0$, an interim evaluation increases aggregate welfare.*
- *If $u_P^0 - u_P^- < u_b^+ - u_b^0$, there exists a threshold $\rho \in [w(\eta) + u_b^+ - u_b^0, w(1) + u_b^+ - u_b^0]$ ($\rho = u_b^+ - u_b^0$) if the evaluation is public (private), such that an interim evaluation increases aggregate welfare if and only if $\Delta_b < \rho$.*

The intuition for the first result is straightforward. With evaluation, the principal has the opportunity to centralize in the second period when observing an inappropriate project choice whereas without evaluation the decision is always delegated if the prior is high. If this opportunity to centralize is more valuable to the principal than it harms the biased agent, evaluation increases the aggregate welfare.

Regarding the second result, if centralization harms the biased agent more than it benefits the principal, welfare is no longer increased if a separating equilibrium is played. Nevertheless, aggregate payoffs are larger if the biased agent imitates the loyal agent in Period 1 – compared to the situation where the principal delegates twice without evaluation. Thus, welfare is still increased if (and only if) the biased agent's utility differential, Δ_b, is low enough as this increases pooling incentives.

A Comparison of Public and Private Evaluation.

Considering each type of equilibrium separately, aggregate welfare is the same whether the evaluation is public or private. However, the highest aggregate welfare is achieved in the pooling equilibrium and it decreases in the probability that the biased agent separates. Furthermore, for any given prior $p > \eta$, the probability that the biased agent separates is at least as high with private evaluation than it is with public evaluation. Thus, Lemma 4 is immediate.

Lemma 4. *If $p > \eta$, the aggregate welfare generated in case of public evaluation is at least as high as if the evaluation is private.*

Welfare Effects of the Wage Schedule.

Another determinant of aggregate welfare is the wage schedule. Of course, total welfare never changes under a simple redistribution of payoffs. Thus, in a usual context, the wage schedule does not influence total welfare. In the present setting with public evaluation, however, the wage schedule, together with $u_b^0 - u_b^-$, determines the equilibrium played in the signaling game which, in turn, influences total welfare. In particular, high wages (weakly) increase the area where a pooling equilibrium is played and (weakly) decrease the area where a separating equilibrium is played, which has a positive effect on aggregate welfare. Hence, focusing on the effect of the wage schedule, paying competitive wages weakly increases welfare when intermediate evaluation takes place if prior beliefs are high.

Proposition 10. *If $p > \eta$, welfare in case of public evaluation is maximized if $w(\mu)$ is maximal within the principal's budget constraints for $\mu \in [p, 1]$.*

Intuitively, increasing wages for a given belief μ strengthens the incentive for the biased agent to mirror the behavior of the loyal agent in Period 1 (so as to obtain the high wage and the discretion to choose in Period 2). In particular, the higher the wages, the more likely is the biased agent to choose the project which is aligned with the principal's preferences. Accordingly, for higher wages, the region where a pooling equilibrium is played is increased (while the region for a separating equilibrium is reduced). As the pooling equilibrium is welfare enhancing if $u_b^+ - u_b^- = \Delta_b \leq \Delta_P = u_P^+ - u_P^-$, this improves aggregate welfare. If the evaluation is private, the wage schedule has no influence on aggregate welfare.

The Case of Low Priors: $p \leq \eta$**.**

Finally, we briefly turn to the case where the prior belief is such that the principal would not delegate without an evaluation, i.e. $p \leq \eta$.

Unfortunately, for such small priors, it is difficult to determine clear-cut welfare results. However, if the principal's benefit from centralization compared to letting the biased agent decide about the project, $u_P^0 - u_P^-$, is smaller than the biased agent's respective gain when choosing the project, $u_b^+ - u_b^0$, welfare is enhanced with evaluation if Δ_b is large enough.

Still assuming that the principal's benefit from an optimal project decision outweighs the respective benefits of the biased agent, i.e. $u_P^+ - u_P^- = \Delta_P \geq \Delta_b = u_b^+ - u_b^-$, we obtain the following result.

Proposition 11. *If $p \leq \eta$ and $\Delta_b \geq w(\eta) + u_b^+ - u_b^0$ if the evaluation is public ($\Delta_b \geq u_b^+ - u_b^0$ in case of private evaluation), total expected welfare is (weakly) increased with evaluation if*

$$u_b^+ - u_b^0 \geq u_P^0 - u_P^-.$$

Note that the intuition for this result is in line with the interpretation in case $p > \eta$. If the prior is low, the principal centralizes if no evaluation takes place. By contrast, if the agent is evaluated, he delegates in the first period and thereby induces a benefit of u_b^+ for the biased agent – instead of u_b^0 without evaluation. If this relative gain of the biased agent is larger than the principal's benefit from centralizing compared to letting the biased agent decide, aggregate welfare is increased.

Arguing along the same lines as in the case $p > \eta$, we conclude that public evaluation yields a higher aggregate welfare than private evaluation whenever the former regime yields more pooling by the biased agent.

Lemma 5. *If $p \leq \eta$ and $\Delta_b \geq w(\eta) + u_b^+ - u_b^0$, public evaluation generates a higher aggregate welfare than private evaluation.*

1.2 Extensions

In this section, we briefly consider two natural extensions of our model. First, in Section 1.2.1, we relax the assumption of equally long/important periods. Finally, Section 1.2.2 shows that our results are robust to adding more evaluations and inter-temporal discounting.

1.2.1 Flexible Timing of the Evaluation

Instead of presupposing that the principal might evaluate the agent in the middle of his employment period, we now consider a situation where the principal can choose to evaluate the agent after a fraction $\alpha \in [0, 1]$ of his employment period. In particular, we analyze at which point in time it is optimal for the principal to evaluate the agent depending on the observability of the evaluation.

For the sake of argument, we focus on the case of high prior beliefs, i.e. $p > \eta$. Moreover, we define

$$\alpha_b(\mu) := \frac{w(\mu) + (u_b^+ - u_b^0)}{w(\mu) + (u_b^+ - u_b^0) + (u_b^+ - u_b^-)}.$$

Note that α_b is increasing in μ.

Private Evaluation

Consider first the case of private evaluation. In this case, the biased agent chooses x_l in the first period if the importance of the first period is sufficiently low, i.e. $\alpha < \alpha_b(0) =: \alpha_b$.

Otherwise, a separating equilibrium is played and the biased agent randomizes between strategies if $\alpha = \alpha_b$. Furthermore, the principal's payoff in the separating equilibrium decreases in α, whereas it increases in α for the pooling equilibrium. Thus, Lemma 6 is straightforward.

Lemma 6. *For $p > \eta$, the principal maximizes his expected utility by choosing $\alpha = \alpha_b(0) =: \alpha_b$ for all p. His expected payoff in this case, given that the biased agent plays the mixed strategy $(\lambda, 1-\lambda)$, is then given by*

$$\begin{aligned} E[\Pi_U|\alpha = \alpha_b] &= pu_P^+ + (1-p)\lambda[\alpha_b u_P^+ + (1-\alpha_b)u_P^-] \\ &+ (1-p)(1-\lambda)[\alpha_b u_P^- + (1-\alpha_b)u_P^0] - w(p). \end{aligned}$$

Public Evaluation

Similarly, in case of public evaluation, a pooling equilibrium is played if the weight on the first period is low, i.e. $\alpha < \alpha_b(p)$, and the biased agent separates for $\alpha > \alpha_b(1)$.

However, different from the private case, a mixed equilibrium is played for $\alpha \in [\alpha_b(p), \alpha_b(1)]$. Moreover, assuming $pw(1) - w(p) \leq (1-p)(u_P^0 - u_P^-)$ for all $p \geq \eta$, the same dynamics as in the previous case apply. Yet, the optimal α in case the mix1(D,C) equilibrium is played cannot be derived without further assumptions on the wage schedule.

Lemma 7. *For $p > \eta$ and $pw(1) - w(p) \leq (1-p)(u_P^0 - u_P^-)$, the principal maximizes his expected utility by choosing some $\alpha \in [\alpha_b(p), \alpha_b(1)]$. In this case, the biased player mixes between strategies and the principal's expected payoff is given by*

$$\begin{aligned} E[\Pi_O|\alpha] &= pu_P^+ + (1-p)\lambda[\alpha u_P^+ + (1-\alpha)u_P^-] + (1-p)(1-\lambda)[\alpha u_P^- + (1-\alpha)u_P^0] \\ &- \alpha w(p) - (1-\alpha)[p + (1-p)\lambda]w(\mu_+). \end{aligned}$$

Further,

$$w(\mu_+) = \frac{\alpha}{1-\alpha}(u_b^+ - u_b^-) - (u_b^+ - u_b^0).$$

Thus, the principal decides to evaluate the agent at an earlier stage if the evaluation is private in order to increase the importance of the second period. If the evaluation is public, the second period already provides higher incentives because of wage concerns and the principal evaluates the agent at a later stage, thereby emphasizing the first period.[22]

1.2.2 More Evaluations

Finally, we consider the case where $N-1$ evaluations take place, thereby dividing the employment period into N intervals. For reasons of tractability, we focus on the case $p > \eta$. For small prior beliefs, however, the dynamics of the model are the same and similar results apply.

The game we analyze in the following is a repeated game where the principal is uncertain about the type of the agent. Both players are long-lived, as for example in Schmidt (1993) or Mailath and Samuelson (2006, ch. 16). However, our subsequent argument does not strictly follow these lines in the literature but rather gives an inductive argument for what happens if further periods are added.

As we consider a repeated game with incomplete information, the One-Shot-Deviation Principle does not hold. Nonetheless, we can show that for high prior beliefs an equilibrium strategy for the principal is to delegate upon observing the loyal outcome and to centralize up to the final period otherwise – regardless of the observability of the evaluations. This grim-trigger strategy turns out to be the principal's unique subgame perfect strategy consistent with the Intuitive Criterion (Cho and Kreps, 1987).

Private Evaluation

In order to reach a solution for the multi-period game, we apply a backward induction argument. If the evaluations are private, the biased agent chooses x_b in the penultimate period if $u_b^0 > u_b^-$, he pools if $u_b^0 < u_b^-$ and he plays a mixed strategy if $u_b^0 = u_b^-$.

[22] This result hinges on the assumption that wages are non-negative and not "too convex," i.e. $pw(1) - w(p) \leq (1-p)(u_P^0 - u_P^-)$.

The Case $u_b^0 \geq u_b^-$.

If $u_b^0 > u_b^-$, the biased agent chooses the biased project in the penultimate period, $N-1$, and the principal centralizes. Thus, if the biased agent chooses x_l in Period $N-2$, the principal's belief is unchanged and the separating equilibrium is played in the penultimate period.[23] However, if he chooses x_b, the principal's belief regarding the agent being loyal immediately drops to zero and, hence, the principal centralizes from then on. Thus, by choosing x_l instead of x_b, the biased agent receives only u_b^- in Period $N-2$ but avoids getting u_b^0 in Period $N-1$.

In the borderline case, if $u_b^0 = u_b^-$, the biased agent plays a mixed strategy in the penultimate period. However, if we analyze prior periods it turns out that this behavior cannot be sustained and the biased agent chooses x_b in prior periods, followed by centralization by the principal (see Appendix A1 for details).

Lemma 8 shows that these results hold even if we allow for arbitrary discounting by the biased agent, where $\delta \in [0,1]$ denotes the biased agent's discount factor.

Lemma 8. *If $u_b^0 \geq u_b^-$, there is an equilibrium in the finitely repeated game where, in all periods, the biased agent chooses x_b if the decision is delegated to him for all $\delta \in [0,1]$. The loyal agent chooses x_l, and the principal delegates unless he observes x_b, in which case he centralizes in all remaining periods.*

The Case $u_b^0 < u_b^-$.

If $u_b^0 < u_b^-$, a pooling equilibrium is played in Period $N-1$ and the principal delegates the decision to the agent. Thus, in Period $N-2$, the biased agent gets u_b^- by choosing x_l until in the last period he chooses his preferred project and gains a profit of u_b^+. On the other hand, if he chooses x_b at some point, he momentarily earns u_b^+ but is stuck with a payoff of u_b^0 for all remaining periods. Lemma 9 shows that it is therefore indeed optimal for the biased agent to pool up to the last period if he is patient enough.

[23] The belief does not change if the biased agent pools because the loyal agent always chooses x_l.

Lemma 9. *If* $u_b^0 < u_b^-$ *and* $\delta > \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0}$*, there is an equilibrium in the finitely repeated game where the biased agent chooses* x_l *in all but the last period if the decision is delegated to him, and* x_b *in the last period. The loyal agent chooses* x_l*, and the principal delegates unless he observes* x_b*, in which case he centralizes in all remaining periods.*

Public Evaluation

If the evaluations are public, the equilibria also depend on the wage schedule, and the area where a mixed equilibrium is played with one evaluation is not a null-set (as opposed to the case of private evaluation).

The Case $\Delta_b > w(p) + u_b^+ - u_b^0$.
If $\Delta_b \geq w(1) + u_b^+ - u_b^0$, that is, $u_b^0 - u_b^- \geq w(1)$, the biased agent chooses x_b in Period $N-1$ and the principal centralizes in Period N, the last period. Moving back to Period $N-2$, an argument similar to the one in the case of private evaluation can be applied. The difference is that by choosing x_l the biased agent not only gains discretion in the following period but also earns a higher wage. However, as wages cannot exceed $u_b^0 - u_b^-$ in the separating equilibrium, this effect is too small to make the biased agent pool in any prior period.

If $\Delta_b \in (w(p) + u_b^+ - u_b^0, w(1) + u_b^+ - u_b^0)$, the biased agent plays a mixed strategy in Period $N-1$. However, in this case it turns out that he chooses x_b in previous periods for all discount factors.

Lemma 10. *If* $\Delta_b > w(p) + u_b^+ - u_b^0$*, there is an equilibrium in the finitely repeated game where, in all periods, the biased agent chooses* x_b *if the decision is delegated to him for all* $\delta \in [0, 1]$*. The loyal agent chooses* x_l*, and the principal delegates unless he observes* x_b*, in which case he centralizes in all remaining periods.*

The Case $\Delta_b \leq w(p) + u_b^+ - u_b^0$.

Finally, if Δ_b is low, a pooling equilibrium is played in the penultimate period, $N-1$, and the decision is delegated to the agent in Period N. Therefore, the biased agent chooses x_l in any previous period if he is sufficiently patient.

Lemma 11. *If* $\Delta_b \leq w(p) + u_b^+ - u_b^0$ *and* $\delta > \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0 + w(p)}$*, there is an equilibrium in the finitely repeated game where the biased agent chooses* x_l *in all but the last period if the decision is delegated to him, and* x_b *in the last period. The loyal agent chooses* x_l*, and the principal delegates unless he observes* x_b*, in which case he centralizes in all remaining periods.*

To conclude, the pure strategy equilibria observed in the two-period version of the repeated game in case $p > \eta$ extend to an arbitrary finitely repeated game, given that the biased agent is patient enough. The equilibria where the biased agent mixes between strategies however can not be sustained with more periods. In these cases, the biased agent chooses the non-cooperative strategy irrespective of his patience.

1.3 Concluding Remarks

In this chapter, we have analyzed a model of delegation between an uninformed principal and an informed but potentially biased agent. In particular, assuming generic utilities and discrete choice options, we have explored the effects of an interim evaluation of the agent compared to a situation where no evaluation takes place.

As we have shown, in the relationship with interim evaluation, concerns for reputation may lead the biased agent to misrepresent his preferences and choose the project which is preferred by the principal at an early stage of the interaction, thereby inducing increased levels of delegation by the principal in the first period. Moreover, we have seen that this effect is particularly strong if the evaluation is observable to the market, in which case the monetary compensation for the agent's outside option provides an additional incentive.

Regarding the principal, it turned out that, somewhat contrary to what one might think at first glance, he may even benefit from a public evaluation compared to a situation without evaluation, if the biased agent's imitating behavior is not too costly in terms of wages. Thus, although the principal always benefits from a private (and costless) evaluation, revealing the evaluation to the market may even increase the profitability of the evaluation for the principal. This derives from the fact that, while the agent's expected wage may increase in case the evaluation is observable, the biased agent also has a stronger incentive to align his project choice with the principal's preferences which may overcompensate the wage effect.

In addition, we have argued that, from a welfare perspective, an interim evaluation may lead to an increase in aggregate welfare compared to the case without evaluation. Moreover, we have shown that, if the agent is evaluated, it is always welfare-enhancing to reveal this information to the market (as this increases the biased agent's incentive to imitate the loyal one). And, last but not least, considering the welfare effects of the wage schedule, we have shown that, if delegation occurs both with and without evaluation, paying competitive wages is optimal. This is due to the fact that increases in welfare are essentially achieved at the expense of the biased agent, and high wages provide the strongest incentives for this type to align his first period behavior with the principal's preferences.

Regarding possible policy implications, the analysis thus suggests that, in a setting as the one considered here, strengthening the agent's rights in terms of access to the results of his evaluation and wage claims improves welfare. However, while these results may indeed indicate some lines of thought for more general intuitions, we are fully aware that general claims are of course difficult to make based on the stylized theoretical analysis conducted here.

A1 Mathematical Appendix

Equilibrium Refinement

In this chapter, we use the Intuitive Criterion (Cho and Kreps, 1987) as a refinement for equilibria in the signaling game. In short, the Intuitive Criterion rules out all equilibria which are sustained by unreasonable beliefs off the equilibrium path. In particular, in a pooling equilibrium where both types choose x_b, the loyal agent is the first one to switch to x_l when payoffs are gradually changed. Thus, reasonable off-equilibrium beliefs have to satisfy $\mu_+ = 1$ in this case. But if $\mu_+ = 1$, the principal delegates after observing x_l and the agent's wage after choosing x_l is maximal. Therefore, choosing x_l is strictly dominant for the loyal agent if $\mu_+ = 1$ and pooling on x_b cannot be sustained in equilibrium. This leads us to the following lemma.

Lemma 12. *All pooling equilibria of the signaling game where both types choose x_b are ruled out by the intuitive criterion.*

Accordingly, for a pooling equilibrium where both types choose x_l, the Intuitive Criterion requires $\mu_- = 0$. This requirement is not restrictive for the considered equilibria. See Cho and Kreps (1987) or Mas-Colell et al. (1995, pp. 467ff.) for further details on the Intuitive Criterion.

Proofs of Section 1.1

Proof of Proposition 1. The principal prefers delegation to centralization when no evaluation takes place if $pu_P^+ + (1-p)u_P^- \geq u_P^0$, which is equivalent to

$$p \geq \frac{u_P^0 - u_P^-}{u_P^+ - u_P^-}.$$

■

Proof of Lemma 1. In order to analyze the equilibria of the signaling game in case of private evaluation, we have to conduct an analogous analysis to the one in the proof of Lemma 2 and Lemma 3. However, the difference with private evaluation is that wages are omitted in the analysis of best replies as they have no strategic impact. Hence, the equilibrium analysis in case of private evaluation yields the same result as assuming $w \equiv 0$ in the statement of Lemma 2 and Lemma 3. Indeed, by applying Lemmata 2 and 3 to $w \equiv 0$, the mix1(D,C) equilibrium reduces to a null-set and the equilibria where the principal centralizes in any case vanish. ■

Proof of Lemma 2 and Lemma 3. In order to analyze the equilibria of the signaling game in case of public evaluation, we consider all possible strategies of the principal separately. The pure strategies are analyzed in cases 1 to 4, while case 5 includes all possible mixed strategies by the principal. Throughout the proof, we omit the analysis of pooling equilibria where both types choose x_b because of Lemma 12.

Case 1: $s_P = (D, C) \Rightarrow \mu_- \leq \eta$**,** $\mu_+ \geq \eta$

In this case, the loyal agent chooses x_l if $s_P = (D, C)$ and $\mu_+ \geq \mu_-$, irrespective of the biased agent's strategy. In a separating equilibrium, we have $s_b = x_b$, $\mu_- = 0$ and $\mu_+ = 1$. Thus, $s_b = x_b$ is a best reply for the biased agent if $w(0) + u_b^+ + u_b^0 \geq w(1) + u_b^- + u_b^+$, or

$$\Delta_b \geq w(1) + u_b^+ - u_b^0.$$

In a pooling equilibrium, we have $s_b = x_l$ and $\mu_+ = p$. As $\mu_+ \geq \eta$ is needed for the principal to choose $s_P = (D, C)$, this equilibrium can only exist if $p \geq \eta$. Furthermore, this equilibrium can only be sustained if $w(p) + u_b^- + u_b^+ \geq w(\mu_-) + u_b^+ + u_b^0$, which is equivalent to

$$\Delta_b \leq w(p) + u_b^+ - u_b^0,$$

since we restrict the analysis to rational beliefs and hence $\mu_- = 0$. In an equilibrium where the biased agent mixes between x_l and x_b with probability λ, $1 - \lambda$, respectively, we have $\mu_- = 0$ as the loyal agent always chooses x_l. The biased agent's indifference condition is thus given by $w(\mu_+) + u_b^- + u_b^+ = w(\mu_-) + u_b^+ + u_b^0$, or

$$\Delta_b = w(\mu_+) + u_b^+ - u_b^0.$$

The randomization parameter λ is then implicitly given by $\mu_+ = \frac{p}{p+(1-p)\lambda}$. Furthermore, if $p \leq \eta$, we always have $\mu_+ \in [\eta, 1]$ or, equivalently, $\lambda \in [0, \frac{p(1-\eta)}{\eta(1-p)}]$. If instead $p \geq \eta$, $\mu_+ \in [p, 1]$ or, equivalently, $\lambda \in [0, 1]$. Thus, there exists some mixed strategy $(\lambda, 1-\lambda)$ by the biased agent fulfilling his indifference condition whenever

- $w(\eta) \leq u_b^0 - u_b^- \leq w(1)$ if $p \leq \eta$ or
- $w(p) \leq u_b^0 - u_b^- \leq w(1)$ if $p \geq \eta$.

Hence, if $s_P = (D, C)$, there are three types of equilibria depending on Δ_b and $u_b^+ - u_b^0$: separating, pooling on x_l and one where the biased agent randomizes.

Case 2: $s_P = (D, D) \Rightarrow \mu_- \geq \eta$**,** $\mu_+ \geq \eta$

In this case, we have $\mu_- \geq \eta$ and $\mu_+ \geq \eta$, so there cannot be a separating equilibrium. In a pooling equilibrium with $s_l = s_b = x_l$, rational beliefs are given if $\mu_- = 0$, which is not possible either. Finally, an equilibrium where one of the types mixes and the other type plays a pure strategy can only appear if the loyal agent mixes (in which case either $\mu_+ = 1$ or $\mu_- = 1$), which occurs if and only if $u_l^+ - u_l^- = w(\mu_-) - w(\mu_+)$, thus $\mu_- \geq \mu_+$. But in this case the biased agent chooses x_b, which yields $\mu_+ = 1$ and $\mu_- < 1$. Thus, there is no equilibrium where *one* of the types mixes. Accordingly, the only possible mixed equilibrium in this case is the one where both types mix between their pure strategies. This equilibrium requires $u_b^+ - u_b^- = -(u_l^+ - u_l^-) = w(\mu_+) - w(\mu_-)$, which is not possible as we assumed uniqueness of the preferred project, thus $u_i^+ > u_i^-$ for $i \in \{l, b\}$. Hence, there is no equilibrium with $s_P = (D, D)$.

Case 3: $s_P = (C, C) \Rightarrow \mu_- \leq \eta$**,** $\mu_+ \leq \eta$

In this case, $\mu_- \leq \eta$ and $\mu_+ \leq \eta$, so there is no separating equilibrium. Moreover, in any pooling equilibrium with $s_l = s_b = x_l$, we conclude $\mu_+ = p$ and hence $p \leq \eta$ is required.

Then, the biased agent chooses x_l if $w(p) + u_b^- + u_b^0 \geq w(\mu_-) + u_b^+ + u_b^0$, or

$$u_b^+ - u_b^- \leq w(p),$$

assuming rational beliefs, i.e. $\mu_- = 0$. The loyal type also prefers x_l in this case. Furthermore, an equilibrium where one of the types mixes and the other type plays a pure strategy can only occur if the biased agent mixes (in which case either $\mu_+ = 0$ or $\mu_- = 0$). The biased agent is indifferent in this case if and only if $u_b^+ - u_b^- = w(\mu_+) - w(\mu_-)$. Thus, $\mu_+ > \mu_-$, $\mu_- = 0$ and $s_l = x_l$. As $\lambda \leq 1$, we have that $\mu_+ \geq p$. On the other hand, $\mu_+ \leq \eta$ such that the principal centralizes when observing x_l. Thus, this mixed equilibrium is only possible if $p \leq \eta$, in which case we find a randomization λ whenever

$$w(p) \leq u_b^+ - u_b^- \leq w(\eta).$$

Finally, an equilibrium where both types play mixed strategies requires

$$u_b^+ - u_b^- = -(u_l^+ - u_l^-) = w(\mu_+) - w(\mu_-),$$

which is again not possible. Hence, if $p \leq \eta$, there are two types of equilibria in this case: A pooling equilibrium if $u_b^+ - u_b^- \leq w(p)$ and a mixed equilibrium if $w(p) \leq u_b^+ - u_b^- \leq w(\eta)$.

Case 4: $s_P = (C, D) \Rightarrow \mu_- \geq \eta$**,** $\mu_+ \leq \eta$

In this case, $\mu_- \geq \eta$ and $\mu_+ \leq \eta$. Thus, the only possible separating equilibrium is one in which $s_l = x_b$ and $s_b = x_l$ is played. However, if $\mu_- = 1$ and $\mu_+ = 0$, the biased agent strictly prefers x_b. So there is no separating equilibrium in this case. In any pooling equilibrium with $s_l = s_b = x_l$ and rational beliefs, we have $\mu_- = 0$, which is not possible. Hence, there is no pooling equilibrium. Furthermore, an equilibrium where the biased agent mixes is only possible if he is indifferent, i.e. $w(\mu_-) + 2u_b^+ = w(\mu_+) + u_b^- + u_b^0$, or $w(\mu_-) - w(\mu_+) = u_b^- + u_b^0 - 2u_b^+ < 0$, which is not possible if $\mu_- \geq \mu_+$. Thus, we can exclude such an equilibrium where the biased agent randomizes. On the other hand, if only the loyal agent mixes, the requirement $\mu_- \geq \eta$ can only be achieved if the biased agent chooses x_l, thus $\mu_- = 1$. But if $\mu_- = 1$, the biased agent strictly prefers x_b. Hence, there is no equilibrium in case $s_P = (C, D)$ where one of the types mixes. Finally, an equilibrium where both types mix is only possible if $u_i^+ = u_i^-$ for $i \in \{l, b\}$, which is ruled out by assumption. Thus, there is no equilibrium if $s_P = (C, D)$.

Case 5: $s_P = \lambda_P D + (1 - \lambda_P)C$

The principal randomizes between D and C with probabilities $(\lambda_P, 1-\lambda_P)$ after observing x_l if and only if $\eta = \mu_+$. Consider the case where the biased agent plays x_l with probability λ and x_b with probability $1 - \lambda$ and the loyal agent chooses x_l. Then, $\mu_- = 0$ and $s_P(x_b) = C$. The principal randomizes in order to make the biased agent indifferent, which occurs if $w(\mu_-) + u_b^+ + u_b^0 = w(\mu_+) + \lambda_P(u_b^- + u_b^+) + (1 - \lambda_P)(u_b^- + u_b^0)$, or

$$\lambda_P = \frac{u_b^+ - u_b^- - w(\eta)}{u_b^+ - u_b^0},$$

using $\eta = \mu_+$ and $\mu_- = 0$. Then, $\lambda_P \in [0, 1]$ if and only if $w(\eta) \leq \Delta_b \leq w(\eta) + u_b^+ - u_b^0$. By contrast, the principal is indifferent between C and D if and only if $\mu_+ = \eta$. However, if the biased agent randomizes with probabilities $(\lambda, 1 - \lambda)$, μ_+ is given by $\mu_+ = \frac{p}{p+(1-p)\lambda}$, which is equivalent to $\lambda = \frac{p(1-\eta)}{\eta(1-p)}$. Thus, $\lambda \leq 1$ if and only if $p \leq \eta$ and there is no equilibrium for $p > \eta$. Furthermore, $\lambda_P \in (0, 1)$ if and only if $\Delta_b \in (w(\eta), w(\eta) + u_b^+ - u_b^-)$. All other strategies by the biased and the loyal agent are ruled out in equilibrium by applying similar arguments as before. Hence, in an equilibrium where the principal randomizes between C and D, the loyal agent chooses x_l, while the biased agent randomizes between x_l and x_b. This equilibrium is possible only if $p \leq \eta$ and $\Delta_b \in (w(\eta), w(\eta) + u_b^+ - u_b^-)$. ■

Proof of Proposition 2. Here, we compare the principal's payoff in case he delegates in Period 1 with the payoff he could guarantee by centralization, as given in Section 1.1.4.

If $p > \eta$ and the principal delegates in Period 1, the principal delegates in Period 2 upon observing x_l and centralizes otherwise. The biased agent plays a mixed strategy $(\lambda, 1 - \lambda)$ with $\lambda \in [0, 1]$, i.e. also accounting for pure strategies. The principal's payoff if he delegates in Period 1 is given by

$$E[U_P|D] = 2pu_P^+ + (1-p)[\lambda(u_P^+ + u_P^-) + (1-\lambda)(u_P^- + u_P^0)] - w(p),$$

where we omit first-period wages. Thus, the principal delegates in Period 1 if

$$E[U_P|D] \geq u_P^0 + pu_P^+ + (1-p)u_P^- - w(p) \quad \Leftrightarrow \quad [p + (1-p)\lambda](u_P^+ - u_P^0) \geq 0,$$

which is always the case. Thus, the principal delegates at least for all $p > \eta$.

If $p \leq \eta$ and $\Delta_b \geq u_b^+ - u_b^0$, the biased agent mixes between strategies with $\lambda \leq \frac{1-\eta}{\eta}\frac{p}{1-p}$ and the principal delegates upon observing x_l and centralizes otherwise (in case the decision was delegated in Period 1). This also covers the separating equilibrium, in which case $\lambda = 0$. The principal's payoff in case of delegation in Period 1 is then given by

$$E[U_P|D] = u_P^0 + u_P^- + p(u_P^+ - u_P^-) + \frac{p}{\mu_+}(u_P^+ - u_P^0) - w(p),$$

using $\lambda = \frac{1-\mu_+}{\mu_+}\frac{p}{1-p}$ and omitting first-period wages. The principal delegates in Period 1 if $E[U_P|D] \geq 2u_P^0 - w(p)$, or

$$p \geq \frac{\mu_+(u_P^0 - u_P^-)}{\mu_+(u_P^+ - u_P^-) + (u_P^+ - u_P^0)} \in [\eta^2, \frac{\eta}{2-\eta}],$$

since $\mu_+ \in [\eta, 1]$. In particular, if $\Delta_b > u_b^+ - u_b^0$, $\mu_+ = 1$ and the threshold is given by $\frac{\eta}{2-\eta}$.

If $p \leq \eta$ and $\Delta_b < u_b^+ - u_b^0$, both the biased agent and the principal randomize, where the biased agent's strategy $(\lambda, 1-\lambda)$ is given by $\lambda = \frac{1-\eta}{\eta}\frac{p}{1-p}$ and the principal's delegates with probability λ_P upon observing x_l, and centralizes if $x = x_b$. The principal's payoff in case of delegation in Period 1 is then given by

$$E[U_P|D] = u_P^0 + u_P^- + p\lambda_P(u_P^+ - u_P^-) + \frac{p}{\eta}[(u_P^+ - u_P^-) - \lambda_P(u_p^0 - u_P^-)] - w(p),$$

again omitting first-period wages. The principal delegates in Period 1 if $E[U_P|D] \geq 2u_P^0 - w(p)$, which is equivalent to $p \geq \eta^2$. ∎

Proof of Proposition 3. As in the Proof of Proposition 2, we compare the principal's payoff in case he delegates in Period 1 with the payoff he could guarantee by centralization, as given in Section 1.1.4.

If $p > \eta$ and the principal delegates in Period 1, he delegates in Period 2 upon observing x_l and centralizes otherwise. In turn, if he centralizes in Period 1, he does not update his belief and delegates in Period 2 as $p > \eta$. Thus, delegation in Period 1 is preferred by the principal if

$$\begin{aligned} & p[2u_P^+ - w(\mu_+)] + (1-p)[\lambda(u_P^+ + u_P^- - w(\mu_+)) + (1-\lambda)(u_P^- + u_P^0)] \\ \geq \;& u_P^0 + pu_P^+ + (1-p)u_P^- - w(p) \\ \Leftrightarrow \;& w(p) + [p + (1-p)\lambda][u_P^+ - u_P^0 - w(\mu_+)] \geq 0, \end{aligned}$$

which holds for all $\lambda \in [0,1]$ as $w(\mu_+) \leq u_P^+ - u_P^0$. Thus, the principal delegates at least for all $p > \eta$, as in the case without evaluation.

If $p \leq \eta$ and the principal centralizes in Period 1, he does not update his prior belief and centralizes again in Period 2 as $p \leq \eta$. Now we consider all possible equilibria of the signaling game for the case $p \leq \eta$ and determine the delegation decision of the principal in Period 1.[24]

Case 1: $\Delta_b \leq w(\eta)$
In this case, the biased agent chooses x_l or a mixed strategy, while the principal centralizes in any case. Thus, the principal delegates in this case if

$$\begin{aligned} & p[u_P^+ + u_P^0 - w(\mu_+)] + (1-p)[\lambda(u_P^+ + u_P^0 - w(\mu_+)) + (1-\lambda)(u_P^- + u_P^0)] \\ \geq \;& 2u_P^0 - w(p) \\ \Leftrightarrow \;& \frac{p}{\mu_+}[u_P^+ - u_P^- - w(\mu_+)] \geq u_P^0 - u_P^- - w(p), \end{aligned}$$

where we used $\lambda = \frac{p}{1-p}\frac{1-\mu_+}{\mu_+}$. As $u_P^+ - u_P^- \geq w(\mu_+)$, it follows that the principal delegates

[24]Here, we neglect the null-set $\{p = \eta\} \cap \{\Delta_b \leq w(\eta) + u_b^+ - u_b^0\}$.

if

$$p \geq \frac{[u_P^0 - u_P^- - w(p)]\mu_+}{u_P^+ - u_P^- - w(\mu_+)} = \frac{[u_P^0 - u_P^- - w(p)]\mu_+}{(u_P^+ - u_P^-)\eta + (u_P^+ - u_P^0) - w(\mu_+)}.$$

The latter equals $\delta_{w,\mu_+}(p)$ because $\max\{\mu_+, \eta\} = \eta$.

In the pooling equilibrium, $\mu_+ = p$, and the principal delegates if $p \geq \frac{[u_P^0 - u_P^- - w(p)]p}{u_P^+ - u_P^- - w(p)}$, which is always the case. In order not to have to distinguish the case of a pooling equilibrium from the other cases, it remains to show that $\delta_{w,\mu}(p) \leq p$ for all $\Delta_b \leq w(p)$, such that in this area of Δ_b we have $p_O^*(\Delta_b) = 0$. Then we can use the definition of $\delta_{w,\mu}(p)$ with $\mu(\Delta_b) := w^{-1}(\Delta_b)$ for all $\Delta_b \leq w(\eta)$. Indeed, for $\Delta_b \leq w(p)$ and hence $\mu(\Delta_b) \leq p$, we have

$$\delta_{w,\mu}(p) = \frac{[u_P^0 - u_P^- - w(p)]\mu}{u_P^+ - u_P^- - w(\mu)} \leq \frac{[u_P^0 - u_P^- - w(p)]p}{u_P^+ - u_P^- - w(p)} \leq p.$$

Case 2: $\Delta_b > w(\eta)$

In this case, three equilibria are possible: sep(D,C), mix1(D,C) or mix2(λ_P,C). In general, the principal delegates if

$$\begin{aligned}
& p[\lambda_P(2u_P^+ - w(\mu_+)) + (1-\lambda_P)(u_P^+ + u_P^0 - w(\mu_+))] + (1-p)(1-\lambda)(u_P^- + u_P^0) \\
+ \quad & (1-p)\lambda[\lambda_P(u_P^+ + u_P^- - w(\mu_+)) + (1-\lambda_P)(u_P^+ + u_P^0 - w(\mu_+))] \geq 2u_P^0 - w(p) \\
\Leftrightarrow \quad & p[u_P^+ - u_P^- - w(\mu_+)] + p\lambda_P[\mu_+(u_P^+ - u_P^-) - (u_P^0 - u_P^-)] \geq [u_P^0 - u_P^- - w(p)]\mu_+,
\end{aligned}$$

where we used $\lambda = \frac{p}{1-p}\frac{1-\mu_+}{\mu_+}$. If $\lambda_P \neq 1$, i.e. in the mix2 equilibrium, it has to hold that $\mu_+ = \eta$ in order to make the principal indifferent between his choices. Thus, in this case, $\mu_+(u_P^+ - u_P^-) - (u_P^0 - u_P^-) = 0$ and the principal delegates if

$$p \geq \frac{[u_P^0 - u_P^- - w(p)]\eta}{u_P^+ - u_P^- - w(\eta)} = \frac{[u_P^0 - u_P^- - w(p)]\eta}{(u_P^+ - u_P^-)\eta + (u_P^+ - u_P^0) - w(\eta)} = \delta_{w,\eta}(p)$$

because $\mu_+ = \eta$. In all other cases, $\lambda_P = 1$ and the principal delegates if

$$p \geq \frac{[u_P^0 - u_P^- - w(p)]\mu_+}{(u_P^+ - u_P^-)\mu_+ + (u_P^+ - u_P^0) - w(\mu_+)} = \delta_{w,\mu_+}(p)$$

because $\mu_+ \geq \eta$ in these cases. Note that, apart from the pooling equilibrium, $\mu(\Delta_b) = \mu_+$.

Furthermore, as $w(\mu_+) \leq u_P^+ - u_P^0$, we have

$$\delta_{w,\mu_+}(p) \leq \frac{[u_P^0 - u_P^- - w(p)]\mu_+}{(u_P^+ - u_P^-)\max\{\mu_+, \eta\}} \leq \frac{u_P^0 - u_P^- - w(p)}{u_P^+ - u_P^-} \leq \eta$$

for all $p \in [0,1]$, hence $p_O^* < \eta$ for all Δ_b. Thus, we have shown that the principal delegates more often with public evaluation than if no evaluation takes place. ■

Proof of Proposition 4. We define

$$f(\mu, p) := \delta_{w,\mu}(p) - p.$$

Then, f is continuously differentiable in $p \in [0,1]$ and piecewise continuously differentiable on $\{\mu < \eta\}$ and $\{\mu > \eta\}$. Furthermore, $f(\mu, 0) = \delta_{w,\mu}(0) > 0$ and f is strictly decreasing in p:

$$\frac{\partial f}{\partial p}(\mu, p) = -\frac{\mu w'(p)}{(u_P^+ - u_P^-)\max\{\mu, \eta\} + (u_P^+ - u_P^0) - w(\mu)} - 1 < 0$$

Thus, for every $\mu \geq 0$ there exists a unique $p^*(\mu) > 0$ with $f(\mu, p^*(\mu)) = 0$. Fix such a point (μ^*, p^*) with $f(\mu^*, p^*) = 0$. Then we know that $\frac{\partial f}{\partial p}(\mu^*, p^*) \neq 0$ for all (μ^*, p^*). Hence, by the Implicit Function Theorem, $p^*(\mu)$ can be locally represented by a continuously differentiable function. Accordingly, there exists a continuously differentiable function $g : U \to V$ from an environment U of μ^* to an environment V of p^* such that $g(\mu) = p^*(\mu)$ for all $\mu \in U$. Furthermore,

$$\frac{\partial g}{\partial \mu}(\mu^*) = -\left(\frac{\partial f}{\partial p}(\mu^*, p^*)\right)^{-1} \frac{\partial f}{\partial \mu}(\mu^*, p^*).$$

As we have seen, $\frac{\partial f}{\partial p}(\mu^*, p^*) < 0$ for all (μ^*, p^*), thus the sign of $\frac{\partial g}{\partial \mu}(\mu^*)$ is given by the sign of $\frac{\partial f}{\partial \mu}(\mu^*, p^*)$.

Let $\mu^* < \eta$. Then, $p^* = p(\mu^*)$ is given by $[u_P^0 - u_P^- - w(p^*)]\mu^* = p^*[u_P^+ - u_P^- - w(\mu^*)]$. As the right-hand side is strictly positive for $p^* > 0$, we have $u_P^0 - u_P^- - w(p^*) > 0$ for all $p^* > 0$ in this case. If $\mu^* > \eta$, $p^* = p(\mu^*)$ is given by $[u_P^0 - u_P^- - w(p^*)]\mu^* = p^*[(u_P^+ - u_P^-)\mu^* + (u_P^+ - u_P^0) - w(\mu^*)]$. Applying the same reasoning as before, we conclude

$u_P^0 - u_P^- - w(p^*) > 0$ for all $p^* > 0$. Thus, for $\mu^* < \eta$ we have

$$\frac{\partial f}{\partial \mu}(\mu^*, p^*) = \frac{[u_P^0 - u_P^- - w(p^*)][u_P^+ - u_P^- - w(\mu^*) + \mu^* w'(\mu^*)]}{[u_P^+ - u_P^- - w(\mu^*)]^2},$$

and for $\mu^* > \eta$

$$\frac{\partial f}{\partial \mu}(\mu^*, p^*) = \frac{[u_P^0 - u_P^- - w(p^*)][u_P^+ - u_P^0 - w(\mu^*) + \mu^* w'(\mu^*)]}{[(u_P^+ - u_P^-)\mu^* + (u_P^+ - u_P^0) - w(\mu^*)]^2}.$$

In both cases, the derivative is non-negative, thus $\frac{\partial g}{\partial \mu}(\mu^*) > 0$ for all $\mu^* \neq \eta$. As μ weakly increases in Δ_b, we conclude that the delegation threshold weakly increases in Δ_b. ■

Proof of Corollary 1. Corollary 1 represents a special case of Proposition 3. Hence, let's assume that the delegation threshold is implicitly defined by

$$p_O^* = \frac{[u_P^0 - u_P^- - w(p_O^*)]\mu}{(u_P^+ - u_P^-)\max\{\mu, \eta\} + (u_P^+ - u_P^0) - w(\mu)}.$$

Using $w(\mu) = \mu W$ for all $\mu \in [0, 1]$, the above expression is equivalent to

$$[(u_P^+ - u_P^-)\max\{\mu, \eta\} + (u_P^+ - u_P^0) - \mu W]p_O^* = (u_P^0 - u_P^-)\mu - p_O^* \mu W,$$

thus

$$p_O^*(\mu) = \frac{(u_P^0 - u_P^-)\mu}{(u_P^+ - u_P^-)\max\{\mu, \eta\} + (u_P^+ - u_P^0)}.$$

Inserting $\mu(\Delta_b)$ as defined in Proposition 3 yields the result. ■

Proof of Proposition 5. Now we compare the delegation decision for public and private evaluation. In order to do so, we determine conditions such that p_O^* lies above or below the step function p_U^* for those Δ_b where p_O^* is constant. In the following considerations, we use that for all $x \in [0, 1]$

$$p_O^*(\Delta_b) \leq x \quad \Leftrightarrow \quad \delta_{w,\mu(\Delta_b)}(x) \leq x.$$

This equivalence is illustrated in Figure A1.1.

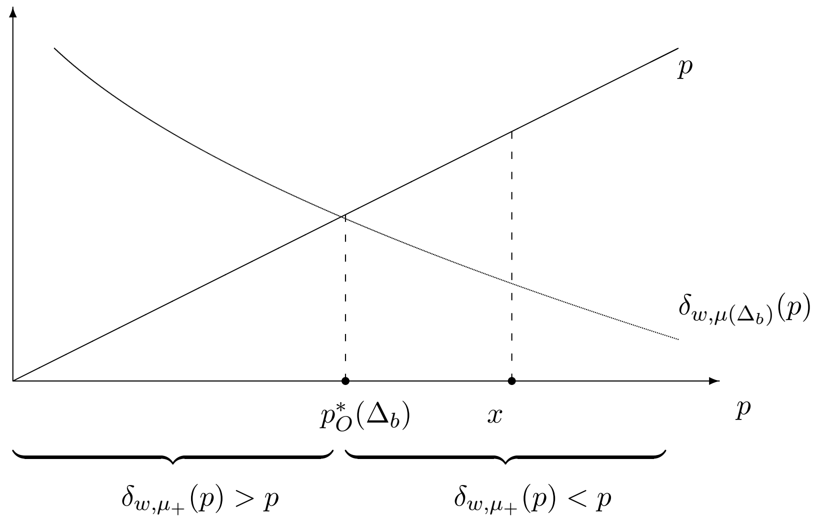

Figure A1.1: Illustration of the equivalence of $p_O^*(\Delta_b) \leq x$ and $\delta_{w,\mu(\Delta_b)}(x) \leq x$.

Case 1: $\Delta_b \geq w(1) + u_b^+ - u_b^0$

In this case, a separating equilibrium is played, both with public and private evaluation. Delegation is more likely when the evaluation is public if $p_O^*(\Delta_b)$, with $\mu(\Delta_b) = 1$, lies below $\frac{\eta}{2-\eta} = p_U^*(\Delta_b)$. Then $p_O^*(\Delta_b) \leq \frac{\eta}{2-\eta}$ is equivalent to $\delta_{w,1}\left(\frac{\eta}{2-\eta}\right) \leq \frac{\eta}{2-\eta}$, which yields

$$\frac{\eta}{2-\eta} w(1) - w\left(\frac{\eta}{2-\eta}\right) \leq 0.$$

As w is (weakly) convex, the above inequality does not hold and $p_O^*(\Delta_b)$ lies above $\frac{\eta}{2-\eta} = p_U^*(\Delta_b)$ for $\Delta_b \geq w(1) + u_b^+ - u_b^0$. Thus, delegation is weakly more likely with private evaluation. As $p_O^*(\Delta_b)$ is continuous and weakly increasing from zero, this result implies that $p_O^*(\Delta_b)$ cuts $p_U^*(\Delta_b)$ at some point $\rho \leq w(1) + u_b^+ - u_b^0$. Thus, the proposition follows immediately. In case 2, we describe the relation between $p_O^*(\Delta_b)$ and $p_U^*(\Delta_b)$ in further detail.

Case 2: $w(\eta) < \Delta_b < w(\eta) + u_b^+ - u_b^0$

In this case, $\mu(\Delta_b) = \eta$ if the evaluation is public, and $p_U^*(\Delta_b)$ is given by η^2 for $\Delta_b < u_b^+ - u_b^0$, and by $\frac{\eta}{2-\eta}$ if $\Delta_b > u_b^+ - u_b^0$. Now, three cases are possible: either $p_O^*(\Delta_b)$ lies below or above any of these thresholds, or in between. On the one hand, $p_O^*(\Delta_b) \leq \eta^2$ if and only if $\delta_{w,\eta}(\eta^2) \leq \eta^2$, or $\eta w(\eta) - w(\eta^2) \leq 0$, which is not the case if w is convex.

On the other hand, $p_O^*(\Delta_b) \leq \frac{\eta}{2-\eta}$ if and only if $\delta_{w,\eta}\left(\frac{\eta}{2-\eta}\right) \leq \frac{\eta}{2-\eta}$, which is equivalent to

$$\frac{\eta}{2-\eta}w(\eta) - \eta w\left(\frac{\eta}{2-\eta}\right) \leq (1-\eta)\frac{\eta}{2-\eta}(u_P^+ - u_P^0).$$

Hence, $p_O^*(\Delta_b)$ lies between η^2 and $\frac{\eta}{2-\eta}$ for $w(\eta) < \Delta_b < w(\eta) + u_b^+ - u_b^0$ if w is "not too convex", and above $\frac{\eta}{2-\eta}$ if w is "very convex". ■

Proof of Corollary 2. In order to prove Corollary 2, we use the proof of Proposition 3. A linear wage schedule implies that $p_O^*(\Delta_b) = \frac{\eta}{2-\eta} = p_U^*(\Delta_b)$ for $\Delta_b \geq w(1) + u_b^+ - u_b^0$. Furthermore, $p_O^*(\Delta_b) = \eta^2 \leq p_U^*(\Delta_b)$ if $w(\eta) < \Delta_b < w(\eta) + u_b^+ - u_b^0$. Taken together, this implies that $p_O^*(\Delta_b)$ lies weakly below the step function $p_U^*(\Delta_b)$, hence there is weakly more delegation when the evaluation is public. ■

Proof of Proposition 6. If $p > \eta$ and the principal privately evaluates the agent, he receives a payoff of

$$E[\Pi_U] = 2pu_P^+ + (1-p)\lambda(u_P^+ + u_P^-) + (1-p)(1-\lambda)(u_P^- + u_P^0) - w(p),$$

where $\lambda \in [0,1]$, covering both the pooling and the separating equilibrium. On the other hand, if $p > \eta$, the principal delegates in Period 1 if there is no evaluation. In this case, his expected payoff is given by

$$E[\Pi_N] = 2[pu_P^+ + (1-p)u_P^-] - w(p),$$

where we omit the first-period wage $w(p)$ in all cases. Thus, evaluating the agent is preferred whenever $E[\Pi_U] \geq E[\Pi_N]$, or

$$(1-p)\lambda(u_P^+ - u_P^0) + (1-p)(u_P^0 - u_P^-) \geq 0,$$

which is true for all $\lambda \in [0,1]$.

Alternatively, if $p \leq \eta$, the principal does not delegate if no evaluation takes place. Hence, comparing the principal's payoffs with and without evaluation yields the same analysis as comparing his payoffs after a delegation decision in Period 1 and after centralization, given that a private evaluation takes place. Hence, the principal prefers to evaluate the agent whenever p lies above the delegation threshold, and he is indifferent if there is no delegation in any case. ■

Proof of Proposition 7. As in the Proof of Proposition 6, if $p > \eta$, the principal's expected payoff without evaluation is given by $E[\Pi_N] = 2[pu_P^+ + (1-p)u_P^-] - w(p)$, omitting the first-period wage. With public evaluation, the principal's payoff is given by

$$E[\Pi_O] = p[2u_P^+ - w(\mu_+)] + (1-p)\lambda[u_P^+ + u_P^- - w(\mu_+)] + (1-p)(1-\lambda)(u_P^- + u_P^0),$$

where $\lambda \in [0,1]$ (including the pooling and the separating equilibrium) and again omitting first-period wages. Using $\mu_+ = \frac{p}{p+(1-p)\lambda}$, this yields for the net expected payoff

$$E[\Pi_O - \Pi_N] = p\frac{1-\mu_+}{\mu_+}(u_P^+ - u_P^0) + (1-p)(u_P^0 - u_P^-) + w(p) - \frac{p}{\mu_+}w(\mu_+),$$

which is non-negative if and only if

$$pw(\mu_+) - \mu_+ w(p) \leq p(1-\mu_+)(u_P^+ - u_P^0) + (1-p)\mu_+(u_P^0 - u_P^-).$$

The above inequality holds for $\mu_+ = p$. Furthermore, for all $\mu_+ \geq p$, the non-negative left-hand side, $pw(\mu_+) - \mu_+ w(p)$, increases in μ_+ because w is convex, while the right-hand side, $p(1-\mu_+)(u_P^+ - u_P^0) + (1-p)\mu_+(u_P^0 - u_P^-)$, is decreasing in μ_+ as $p > \eta$. Knowing that μ_+ weakly increases in Δ_b, we conclude that there is a threshold $\rho > 0$ (possibly infinity) such that a public evaluation is preferred to no evaluation for all $\Delta_b \leq \rho$.

If $p \leq \eta$, the principal centralizes without evaluation which yields an expected payoff of $E[\Pi_N] = 2u_P^0 - w(p)$ (omitting first-period wages). Comparing the principal's expected payoffs in case of evaluation vs. no evaluation yields the same analysis as comparing delegation in Period 1 to centralization in Period 1 in case of public evaluation. Hence, public evaluation is preferred by the principal whenever p lies above the delegation threshold, and he is indifferent if there is no delegation in any case. ■

Proof of Proposition 8. Let's first assume $p > \eta$. If $\Delta_b \leq w(p) + u_b^+ - u_b^0$, the biased agent pools in case the evaluation is public, whereas any mixed strategy by the biased agent is possible in the private case. Thus, the principle prefers to reveal the evaluation whenever

$$\begin{aligned} & 2pu_P^+ + (1-p)(u_P^+ + u_P^-) - w(p) \\ \geq \; & 2pu_P^+ + (1-p)\lambda(u_P^+ + u_P^-) + (1-p)(1-\lambda)(u_P^- + u_P^0) - w(p), \end{aligned}$$

which is true for all $\lambda \in [0,1]$. If $w(p) + u_b^+ - u_b^0 \leq \Delta_b$, the biased agent randomizes between strategies in case the evaluation is public, with $\lambda \in [0,1]$, covering the pooling and the separating case. On the other hand, a separating equilibrium is played in the private case. Thus, the principle prefers to reveal the evaluation whenever

$$\begin{aligned} & 2pu_P^+ + p\frac{1-\mu_+}{\mu_+}(u_P^+ + u_P^-) + \frac{\mu_+ - p}{\mu_+}(u_P^- + u_P^0) - \frac{p}{\mu_+}w(\mu_+) \\ \geq \; & 2pu_P^+ + (1-p)(u_P^- + u_P^0) - w(p), \end{aligned}$$

using $\lambda = \frac{p}{1-p}\frac{1-\mu_+}{\mu_+}$. This is equivalent to

$$pw(\mu_+) - \mu_+ w(p) \leq p(1-\mu_+)(u_P^+ - u_P^0).$$

Note that the above inequality is fulfilled for $\Delta_b = w(p) + u_b^+ - u_b^0$, where $\mu_+ = p$. On the other hand, it does not hold for $\mu_+ = 1$ because w is convex. As the left-hand side is increasing in $\mu_+ \geq p$, while the right-hand side decreases in μ_+, we conclude that for all $p > \eta$, there exists a threshold $\rho(p) \in [w(p) + u_b^+ - u_b^0, w(1) + u_b^+ - u_b^0]$ such that the principal prefers to reveal the evaluation for $\Delta_b < \rho(p)$ and to conceal it if $\Delta_b > \rho(p)$.

Now we assume $p \leq \eta$ and consider the case where the decision is delegated both in case of private and public evaluation. If $\Delta_b \leq w(\eta) \leq u_b^+ - u_b^0$, a mix2($\lambda_P$,C) equilibrium is played in the private case, whereas pool(C,C) or mix1(C,C) occurs with public evaluation. Hence, public evaluation is preferred if

$$E[\Pi_O] = \frac{p}{\mu_+}(u_P^+ - u_P^-) + u_P^0 + u_P^- - \frac{p}{\mu_+}w(\mu_+) \geq \frac{p}{\eta}(u_P^+ - u_P^-) + u_P^0 + u_P^- - w(p) = E[\Pi_U],$$

where $\mu_+ = p$ for $\Delta_b \leq w(p)$ and $\mu_+ = w^{-1}(\Delta_b)$ if $\Delta_b \in [w(p), w(\eta)]$. This is equivalent to

$$pw(\mu_+) - \mu_+ w(p) \leq p\left(1 - \frac{\mu_+}{\eta}\right)(u_P^+ - u_P^-).$$

The above inequality holds for $\mu_+ = p$, but not for $\mu_+ = \eta$. As the left-hand side increases in $\mu_+ \geq p$, while the right-hand side decreases in μ_+, there is a threshold $\rho_1(p) \in [w(p), w(\eta)]$ such that the principal reveals the evaluation for $\Delta_b < \rho_1(p)$.
On the other hand, if $\Delta_b \leq u_b^+ - u_b^0 \leq w(\eta)$, we face the same situation. Hence, for the case $u_b^+ - u_b^0 \leq w(\eta)$, we see that public evaluation is preferred at least up to some threshold $\rho_1(p) \in [w(p), w(\eta)]$. In order to determine $\rho_2(p)$, we distinguish two cases: w being "not too convex" and "very convex".

Case 1: $pw(\eta) - \eta w(p) \leq p(1-\eta)(u_P^+ - u_P^0)$
If $\Delta_b \geq w(\eta) + u_b^+ - u_b^0$, a mix1(D,C) or sep(D,C) equilibrium is played if the evaluation is public, while the biased agent always separates in the private case. Some algebra shows that the principal reveals the evaluation if

$$pw(\mu_+) - \mu_+ w(p) \leq p(1-\mu_+)(u_P^+ - u_P^0)$$

for $\mu_+ \in [\eta, 1]$. While the above inequality never holds if $\mu_+ = 1$, it holds for $\mu_+ = \eta$ given that w is "not too convex". Applying the same reasoning as above, we conclude that there is a threshold $\rho_2(p) \in [w(\eta) + u_b^+ - u_b^0, w(1) + u_b^+ - u_b^0]$ such that the principal reveals the evaluation for $\Delta_b \in [w(\eta) + u_b^+ - u_b^0, \rho_2(p))$ and conceals it for $\Delta_b > \rho_2(p)$.

Case 2: $pw(\eta) - \eta w(p) \geq p(1-\eta)(u_P^+ - u_P^0)$

From case 1 we conclude that the principal prefers to conceal the evaluation for all $\Delta_b \geq w(\eta) + u_b^+ - u_b^0$ if w is "very convex". Furthermore, if $\max\left\{w(\eta), u_b^+ - u_b^0\right\} \leq \Delta_b \leq w(\eta) + u_b^+ - u_b^0]$, both the principal and the biased agent randomize between strategies if the evaluation is public, while a separating equilibrium is played in the private case. Comparing the equilibria, it turns out that the principal prefers to conceal the evaluation if and only if w is "very convex", i.e. $pw(\eta) - \eta w(p) \geq p(1-\eta)(u_P^+ - u_P^0)$. If $u_b^+ - u_b^0 \leq \Delta_b \leq w(\eta)$, a pool(C,C) or mix1(C,C) equilibrium is played in the public case, while the biased agent separates if the evaluation is private. Here, the principal prefers to reveal the evaluation if

$$pw(\mu_+) - \mu_+ w(p) \leq p[1 - \mu_+(2-\eta)](u_P^+ - u_P^-)$$

with $\mu_+ \in [\max\left\{w^{-1}(u_b^+ - u_b^0), p\right\}, \eta]$. The above inequality does not hold for $\mu_+ = \eta$, as w is "very convex", and it holds for $\mu_+ = p$ because $p \leq \eta < \frac{1}{2-\eta}$. Applying the same arguments as before, there exists a threshold $\rho_2(p) \in [\rho_1(p), w(\eta)]$ such that the principal conceals the evaluation for $\Delta_b > \rho_2(p)$. On the other hand, if $w(\eta) \leq \Delta_b \leq u_b^+ - u_b^0$, a mix2($\lambda_P$,C) equilibrium is played in both cases and the principal prefers to conceal the evaluation as wages are convex. Together with the result on $\rho_1(p)$, we can conclude that $\rho_2(p) = \rho_1(p)$ in this case and the principal conceals the evaluation if and only if $\Delta_b > \rho_2(p) = \rho_1(p) \in [w(p), w(\eta)]$. ■

Proof of Corollary 4. If wages are linear, the expected wage if the evaluation is public equals the one in the private case. Hence, replacing $pw(\mu_+) - \mu_+ w(p) = 0$ in the proof of Proposition 8 yields that the principal prefers to reveal the evaluation to the market for any of the considered cases. ■

Proof of Proposition 9. If $p > \eta$, expected aggregate welfare without evaluation is given by

$$E[W_N] = 2p(u_P^+ + u_l^+) + 2(1-p)(u_b^+ + u_P^-).$$

When evaluation takes place and $p > \eta$ (regardless of whether it is public or private), the principal's strategy is $s_P = (D, C)$ in all equilibria and the loyal agent always chooses $s_l = x_l$. The biased agent's strategy is given by $(\lambda, 1-\lambda)$ with $\lambda \in [0,1]$. Thus, we can calculate the expected aggregate welfare with evaluation,

$$\begin{aligned} E[W_E] &= 2p(u_P^+ + u_l^+) + (1-p)\lambda[(u_P^+ + u_P^-) + (u_b^- + u_b^+)] \\ &+ (1-p)(1-\lambda)[(u_P^- + u_P^0) + (u_b^+ + u_b^0)]. \end{aligned}$$

Accordingly, expected net aggregate welfare is given by

$$E[W_E - W_N] = (1-p)\lambda[(u_P^+ - u_P^0) - (u_b^0 - u_b^-)] + (1-p)[(u_P^0 - u_P^-) - (u_b^+ - u_b^0)].$$

If $u_P^0 - u_P^- \geq u_b^+ - u_b^0$, the second summand is non-negative. Hence, if $\lambda = 0$ (separating equilibrium), we have $E[W_E - W_N] \geq 0$. Second, whenever $\lambda \in (0,1)$, it has to hold that $u_b^0 - u_b^- \leq w(1) \leq u_P^+ - u_P^0$ in the public case, or $u_b^0 - u_b^- = 0$ if the evaluation is private. Thus, in this case, the first summand is also non-negative which again yields $E[W_E - W_N] \geq 0$. Finally, if $\lambda = 1$ (pooling equilibrium), $E[W_E - W_N] \geq 0$ if and only if $u_P^+ - u_P^- \geq u_b^+ - u_b^-$, which holds by assumption. On the other hand, if $u_P^0 - u_P^- \leq u_b^+ - u_b^0$, aggregate welfare is weakly increased if and only if

$$\lambda[(u_P^+ - u_P^0) - (u_b^0 - u_b^-)] \geq (u_b^+ - u_b^0) - (u_P^0 - u_P^-) \geq 0.$$

Obviously, in a separating equilibrium ($\lambda = 0$), this inequality does not hold, while for $\lambda = 1$, it is fulfilled as $u_P^+ - u_P^- \geq u_b^+ - u_b^-$. In intermediate cases, we know that $u_P^+ - u_P^0 \geq u_b^0 - u_b^-$, thus welfare is weakly increased if and only if

$$\lambda \geq \frac{(u_b^+ - u_b^0) - (u_P^0 - u_P^-)}{(u_P^+ - u_P^0) - (u_b^0 - u_b^-)} \in [0,1].$$

Hence, there exists a threshold $\rho \in [w(\eta) + u_b^+ - u_b^0, w(1) + u_b^+ - u_b^0]$ if the evaluation is public, or $\rho = u_b^+ - u_b^0$ in the private case, such that welfare is increased for $\Delta_b < \rho$ and decreased for $\Delta_b > \rho$. ■

Proof of Lemma 4. If $p > \eta$, the derivative of the expected aggregate welfare in case of evaluation with regard to λ is given by

$$\frac{\partial E[W_E]}{\partial \lambda} = (1-p)[(u_P^+ - u_P^0) - (u_b^0 - u_b^-)].$$

If the biased agent does not separate in the public case, we have $u_b^0 - u_b^- \leq w(1) \leq u_P^+ - u_P^0$, thus $\frac{\partial E[W_E]}{\partial \lambda} \geq 0$ for all $\lambda > 0$. Furthermore, in this case we have that the randomization factor λ is always at least as large in the public case as in the private case, for a given (p, Δ_b). Thus, public evaluation yields a (weakly) higher aggregate welfare. If the biased agent separates in the public case, he also does so with private evaluation and aggregate welfare is the same in both cases. ■

Proof of Proposition 10.
As $E[W_N]$ is constant in λ, we conclude from the proof of Lemma 4 that $\frac{\partial E[W_E - W_N]}{\partial \lambda} = \frac{\partial E[W_E]}{\partial \lambda} \geq 0$. As the wage schedule influences the played equilibrium, increasing $w(p)$ increases the region for values of Δ_b where a pooling equilibrium is played if the evaluation is public. In turn, rising $w(1)$ reduces the region for values of Δ_b where a separating equilibrium is played. Furthermore, if $w(p) + u_b^+ - u_b^0 < \Delta_b < w(1) + u_b^+ - u_b^0$, λ decreases in μ_+, where $\mu_+ = w^{-1}(\Delta_b - (u_b^+ - u_b^0))$. Thus, for a given $\Delta_b \in (w(p) + u_b^+ - u_b^0, w(1) + u_b^+ - u_b^0)$, welfare is maximal if $w(\mu_+(\Delta_b))$ is maximal. Hence, high wages for $\mu \in [p, 1]$ (weakly) increase expected welfare in case of public evaluation. ■

Proof of Proposition 11. If $p \leq \eta$, expected aggregate welfare without evaluation is given by $E[W_N] = 2[u_P^0 + pu_l^0 + (1-p)u_b^0]$. If $\Delta_b \geq w(\eta) + u_b^+ - u_b^0$ if the evaluation is public, or $\Delta_b \geq u_b^+ - u_b^0$ in the private case, either a mix1(D,C) or a sep(D,C) equilibrium is played. Using the result for $E[W_E]$ from Proposition 9 and $p + (1-p)\lambda = \frac{p}{\mu_+}$, we conclude that the net expected welfare is given by

$$\begin{aligned} E[W_E - W_N] &= [(u_b^+ - u_b^0) - (u_P^0 - u_P^-)] + \frac{p}{\mu_+}[(u_P^+ - u_P^0) - (u_b^0 - u_b^-)] \\ &+ p[2(u_l^+ - u_l^0) + (u_b^0 - u_b^-) + (u_P^+ - u_P^-) - (u_b^+ - u_b^0)], \end{aligned}$$

where $\mu_+ \in (\eta, 1]$. We assume throughout the proof that

$$u_b^+ - u_b^0 \geq u_P^0 - u_P^-. \tag{A1.1}$$

Condition A1.1 ensures that the first of the three summands in $E[W_E - W_N]$ is non-negative. If $\mu_+ < 1$, we have $u_b^0 - u_b^- = 0$ in the private case and $u_b^0 - u_b^- \leq w(1) \leq u_P^+ - u_P^0$ if the evaluation is public. Hence, also the second summand is non-negative in this case. Finally, we have $u_b^+ - u_b^0 \leq \Delta_b < u_b^+ - u_b^- \leq u_P^+ - u_P^-$, which yields non-negativity for the third summand. If $\mu_+ = 1$, net expected welfare is given by

$$E[W_E - W_N] = [(u_b^+ - u_b^0) - (u_P^0 - u_P^-)] + p[2(u_l^+ - u_l^0) + (u_P^+ - u_P^-) + (u_P^+ - u_P^0) - (u_b^+ - u_b^0)].$$

The second summand is non-negative because $u_b^+ - u_b^0 \leq u_P^+ - u_P^-$ as above. ■

Proof of Lemma 5. If $p \leq \eta$ and $\Delta_b \geq w(\eta) + u_b^+ - u_b^0$, the mix1(D,C) or the sep(D,C) equilibrium is played in case of public evaluation, whereas the separating equilibrium is played if the evaluation is private. Following the same argument as in the proof of Lemma 4, it is immediate that the aggregate welfare is weakly increased by revealing the evaluation. ■

Proofs of Section 1.2

Proof of Lemma 6. If the evaluation is private and $p > \eta$, the equilibrium analysis is analogous to Lemma 1. Here, we only briefly discuss the resulting equilibria.

In all equilibria, the principal delegates after observing x_l and centralizes otherwise. The biased agent then strictly prefers x_l if $\alpha u_b^- + (1-\alpha)u_b^+ > \alpha u_b^+ + (1-\alpha)u_b^0$, which is equivalent to

$$\alpha < \frac{u_b^+ - u_b^0}{(u_b^+ - u_b^0) + (u_b^+ - u_b^-)} = \alpha_b(0) =: \alpha_b.$$

Otherwise, he plays a separating equilibrium, and he mixes between strategies in case of equality. The principal's payoff in the pooling equilibrium is given by

$$E[\Pi_U|\alpha < \alpha_b] = pu_P^+ + (1-p)[\alpha u_P^+ + (1-\alpha)u_P^-] - w(p),$$

which increases in α. On the other hand, the principal's payoff in the separating equilibrium is given by

$$E[\Pi_U|\alpha > \alpha_b] = pu_P^+ + (1-p)[\alpha u_P^- + (1-\alpha)u_P^0] - w(p),$$

which decreases in α. Thus, it is optimal for the principal to choose $\alpha = \alpha_b$, in which case his payoff is given by

$$\begin{aligned} E[\Pi_U|\alpha = \alpha_b] &= pu_P^+ + (1-p)\lambda[\alpha_b u_P^+ + (1-\alpha_b)u_P^-] \\ &+ (1-p)(1-\lambda)[\alpha_b u_P^- + (1-\alpha_b)u_P^0] - w(p). \end{aligned}$$

■

Proof of Lemma 7. If the evaluation is public and $p > \eta$, the equilibrium analysis is analogous to Lemma 2. Here, we only briefly discuss the resulting equilibria.
In all equilibria, the principal delegates after observing x_l and centralizes otherwise. In a pooling equilibrium, for the biased agent to strictly prefer x_l, it has to hold that

$$\begin{aligned} & \alpha u_b^- + (1-\alpha)u_b^+ + (1-\alpha)w(p) > \alpha u_b^+ + (1-\alpha)u_b^0 \\ \Leftrightarrow \quad & \alpha < \frac{w(p) + (u_b^+ - u_b^0)}{w(p) + (u_b^+ - u_b^0) + (u_b^+ - u_b^-)} = \alpha_b(p). \end{aligned}$$

On the other hand, a separating equilibrium is played if

$$\begin{aligned} & \alpha u_b^- + (1-\alpha)u_b^+ + (1-\alpha)w(1) < \alpha u_b^+ + (1-\alpha)u_b^0 \\ \Leftrightarrow \quad & \alpha > \frac{w(1) + (u_b^+ - u_b^0)}{w(1) + (u_b^+ - u_b^0) + (u_b^+ - u_b^-)} = \alpha_b(1). \end{aligned}$$

Finally, the biased agent randomizes between strategies if

$$\begin{aligned} & \alpha u_b^- + (1-\alpha)u_b^+ + (1-\alpha)w(\mu_+) = \alpha u_b^+ + (1-\alpha)u_b^0 \\ \Leftrightarrow \quad & \alpha = \frac{w(\mu_+) + (u_b^+ - u_b^0)}{w(\mu_+) + (u_b^+ - u_b^0) + (u_b^+ - u_b^-)} = \alpha_b(\mu_+) \end{aligned} \tag{A1.2}$$

and $\mu_+ \in [p, 1]$ if and only if $\alpha \in [\alpha_b(p), \alpha_b(1)]$. The principal's payoff in the pooling equilibrium is given by

$$E[\Pi_O|\alpha < \alpha_b(p)] = pu_P^+ + (1-p)[\alpha u_P^+ + (1-\alpha)u_P^-] - w(p),$$

which increases in α. On the other hand, the principal's payoff in the separating equilibrium is given by

$$E[\Pi_O|\alpha > \alpha_b(1)] = pu_P^+ + (1-p)[\alpha u_P^- + (1-\alpha)u_P^0] - \alpha w(p) - (1-\alpha)pw(1),$$

which decreases in α if $pw(1) - w(p) \leq (1-p)(u_P^0 - u_P^-)$. Thus, in this case it is optimal for the principal to choose $\alpha \in [\alpha_b(p), \alpha_b(1)]$, in which case his payoff is given by

$$\begin{aligned} E[\Pi_O|\alpha] &= pu_P^+ + (1-p)\lambda[\alpha u_P^+ + (1-\alpha)u_P^-] + (1-p)(1-\lambda)[\alpha u_P^- + (1-\alpha)u_P^0] \\ &- \alpha w(p) - (1-\alpha)[p + (1-p)\lambda]w(\mu_+). \end{aligned}$$

Furthermore, from (A1.2) we conclude that $w(\mu_+) = \frac{\alpha}{1-\alpha}(u_b^+ - u_b^-) - (u_b^+ - u_b^0)$. ■

Proof of Lemma 8.

If $u_b^0 > u_b^-$, the biased agent chooses x_b in the penultimate period. He chooses x_b again in the period before if $u_b^+ + \delta u_b^0 + \delta^2 u_b^0 \geq u_b^- + \delta u_b^+ + \delta^2 u_b^0$, or $\delta \leq \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0}$, which is always the case as $u_b^0 > u_b^-$ and hence $\frac{u_b^+ - u_b^-}{u_b^+ - u_b^0} > 1$. So the biased agent chooses x_b again in period $N-2$. By induction, we see that the biased agent chooses x_b in period $N-n$, given that he has chosen x_b in period $N-(n-1)$ if

$$u_b^+ + \underbrace{\sum_{k=1}^{n} \delta^k u_b^0}_{\text{C after } x_b} \geq u_b^- + \underbrace{\delta u_b^+ + \sum_{k=2}^{n} \delta^k u_b^0}_{x_b \text{ in } N-(n-1), \text{ then C}},$$

or $\delta \leq \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0}$. If $u_b^0 = u_b^-$, the biased agent plays a mixed strategy in the penultimate period. He chooses x_b in the period before if

$$u_b^+ + \delta u_b^0 + \delta^2 u_b^0 \geq u_b^- + \delta[\lambda(u_b^- + \delta u_b^+) + (1-\lambda)(u_b^+ + \delta u_b^0)],$$

which is equivalent to $\delta \leq 1$, where we used $u_b^- = u_b^0$. For all prior periods, the same argument applies as in case $u_b^0 > u_b^-$. Furthermore, if the principal observes x_l, he knows that the agent is loyal, thus his best reply is to delegate the decision again. However, if he observes x_b, the agent is biased with certainty and the principal maximizes his future payoff by centralizing until the last period. ■

Proof of Lemma 9.

If $u_b^0 < u_b^-$, the biased agent chooses x_l in the penultimate period. If we add a period, the biased agent again chooses x_l if and only if $u_b^- + \delta u_b^- + \delta^2 u_b^+ \geq u_b^+ + \delta u_b^0 + \delta^2 u_b^0$, or, equivalently, $\delta \geq \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0}$. Now $\frac{u_b^+ - u_b^-}{u_b^+ - u_b^0} < 1$ as $u_b^0 < u_b^-$ and the biased agent pools if δ is large enough. By induction, the biased agent pools in period $N-n$ given that he chooses x_l in period $N-(n-1)$ if and only if

$$u_b^- + \underbrace{\sum_{k=1}^{n-1} \delta^k u_b^- + \delta^n u_b^+}_{x_l \text{ until last period}} \geq u_b^+ + \underbrace{\sum_{k=1}^{n} \delta^k u_b^0}_{\text{C after } x_b},$$

or $\delta \geq \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0}$, which gives us the same constraint as before. Moreover, if the principal observes x_l, he does not update his prior belief. But as we assume $p \geq \eta$, his best reply is to delegate the decision. If he observed the off-equilibrium outcome x_b, the rational beliefs assumption requests that the principal infers that the agent is biased. Thus, his best reply is to centralize in all future periods if he observes x_b. ■

Proof of Lemma 10.

If $\Delta_b \geq w(1) + u_b^+ - u_b^0$, that is, $u_b^0 - u_b^- \geq w(1)$, the biased agent chooses x_b in the penultimate period. He chooses x_b again in the period before if

$$u_b^+ + \delta u_b^0 + \delta^2 u_b^0 \geq u_b^- + \delta(u_b^+ + w(p)) + \delta^2 u_b^0,$$

which is equivalent to $\delta \leq \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0 + w(p)}$, which is always the case as $w(p) < w(1) \leq u_b^0 - u_b^-$ in the separating equilibrium and hence $\frac{u_b^+ - u_b^-}{u_b^+ - u_b^0 + w(p)} > 1$. So the biased agent chooses x_b again in period $t = N - 2$ – recall that Period N is the last one. By induction, we see that the biased agent chooses x_b in period $N - n$, given that he has chosen x_b in period $N - (n-1)$ if

$$u_b^+ + \underbrace{\sum_{k=1}^{n} \delta^k u_b^0}_{\text{C after } x_b} \geq u_b^- + \underbrace{\delta(u_b^+ + w(p)) + \sum_{k=2}^{n} \delta^k u_b^0}_{x_b \text{ in } T-(n-1)\text{, then C}},$$

which again yields $\delta \leq \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0 + w(p)}$. If $\Delta_b \in (w(p) + u_b^+ - u_b^-, w(1) + u_b^+ - u_b^-)$, the biased agent plays a mixed strategy in the penultimate period. He chooses x_b before if

$$\begin{aligned} & u_b^+ + \delta u_b^0 + \delta^2 u_b^0 > u_b^- + \delta[w(p) + \lambda(u_b^- + \delta(u_b^+ + w(\mu_+))) + (1-\lambda)(u_b^+ + \delta u_b^0)] \\ \Leftrightarrow \quad & \delta[w(p) + \delta\lambda w(\mu_+)] > (1 + \delta\lambda)[(u_b^+ - u_b^-) - \delta(u_b^+ - u_b^0)]. \end{aligned}$$

The latter inequality holds because $w(p) < w(\mu_+) = u_b^0 - u_b^-$, thus

$$\begin{aligned} \delta[w(p) + \delta\lambda w(\mu_+)] &< \delta(1 + \delta\lambda)w(\mu_+) = \delta(1 + \delta\lambda)(u_b^0 - u_b^-) \\ &\leq (1 + \delta\lambda)[(u_b^+ - u_b^-) - \delta(u_b^+ - u_b^0)]. \end{aligned}$$

With regard to the principal's equilibrium strategy, the same argument as in the proof of Lemma 8 applies and the grim-trigger strategy is the principal's unique subgame perfect best reply. ■

Proof of Lemma 11.

If $\Delta_b \leq w(p) + u_b^+ - u_b^0$, the biased agent chooses x_l in the penultimate period. If we add a period, the biased agent again chooses x_l if and only if

$$u_b^- + \delta(u_b^- + w(p)) + \delta^2(u_b^+ + w(p)) \geq u_b^+ + \delta u_b^0 + \delta^2 u_b^0,$$

or $\delta \geq \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0 + w(p)}$. Now $\frac{u_b^+ - u_b^-}{u_b^+ - u_b^0 + w(p)} \leq 1$ as $w(p) \geq u_b^0 - u_b^-$ and the biased agent pools if δ is large enough. By induction, the biased agent pools in Period $N - n$ given that he chooses x_l in Period $N - (n-1)$ if and only if

$$u_b^- + \underbrace{\sum_{k=1}^{n-1} \delta^k (u_b^- + w(p)) + \delta^n (u_b^+ + w(p))}_{x_l \text{ until last period}} \geq u_b^+ + \underbrace{\sum_{k=1}^{n} \delta^k u_b^0}_{\text{C after } x_b},$$

which yields again $\delta \geq \frac{u_b^+ - u_b^-}{u_b^+ - u_b^0 + w(p)}$. With the same argument as in Lemma 9, the grim-trigger strategy is the principal's unique subgame perfect best reply consistent with the Intuitive Criterion. ∎

Chapter 2

Communicating with Extreme or Conservative Agents

A central challenge organizations face is that information is often dispersed across hierarchies, yielding a lack of relevant factual knowledge among decision makers. Subordinate managers, for example, are often much better informed about consumer needs, competitive pressure, or market opportunities than their superiors. However, sharing this information, even if communication is costless, turns out to be difficult if the informed agent follows a private agenda. While the agent has some incentive to steer the uninformed principal into the right direction by disclosing relevant information, she conveys her knowledge to a certain extent only, in order to prevent the principal from overruling her preferences. This imperfect transmission of information due to conflicts of interest might cause the organization to take erroneous project decisions or allocate resources inefficiently, thereby reducing firm value. In this chapter, we first analyze how effective strategic communication can be in a general setting, where the preference divergence between principal and agent might depend on the environment the organization faces. Second, we study delegation as an alternative to strategic communication and analyze the optimal allocation of authority.

Starting from the seminal paper by Crawford and Sobel (1982; henceforth: [CS]), a vast majority of the literature on cheap talk has so far focused on situations where the principal and the agent disagree about the right course of action in all states of the world. In other words, the agent exhibits a systematic bias, independent of the actual state of the world, implying that his preference constantly differs from the principal's. Examples for such systematic biases include status-quo biases or empire building, that is, managers are inclined to cause their division or department to grow beyond the optimal size (see Jensen, 1986). In fact, there exists ample empirical evidence for empire building in companies (e.g. Shin and Kim, 2002; Chen at al., 2008; Hope and Thomas, 2008; Mortal and Reisel, 2012; Du et al., 2013) and governments (Levinson, 2005). Similarly, Dessein (2002; henceforth: [D]) assumes the agent to exhibit a purely systematic bias when comparing communication to delegation in order to characterize the optimal allocation of decision rights by the principal.

In contrast, Alonso, Dessein and Matouschek (2008; henceforth: [ADM]) analyze a situation where the agent has no systematic bias, but the preference divergence between principal and agent depends on some state of the world. In their model, two division managers need to coordinate and adapt to specific local circumstances. Hence, if both divisions operate in some status quo, adaptation is not an issue and both division managers coordinate on the same project. However, if local conditions change, their preferences increasingly differ from each other. Generally speaking, the principal and the agent agree on the preferred project choice as long as the state of the world represents some status quo, and they increasingly disagree in more extreme situations. However, [ADM] assume that the agent prefers more extreme projects than the principal, in the sense that he exaggerates extreme states relative to the principal.

In reality, it is likely that a manager exhibits both a state-dependent and a systematic bias. Imagine for example a production manager who on the one hand exaggerates the need for investment into a new plant in times of increasing demand if his preferences are extreme, or understates deconstruction exigences in a shrinking market if he is rather conservative. On the other hand, he requests a minimum investment level independent of the demand situation in order to maintain or strengthen his position in the organization. In our model, we thus combine and generalize the above approaches by allowing the agent to exhibit both a systematic *and* a state-dependent bias, such that the preference divergence is an arbitrary linear function of the state of the world.[1] In particular, we allow for the possibility that the agent prefers more conservative choices than the principal. Furthermore, the agent might exhibit reversed preferences with regard to the principal's, that is, his preferred project decreases in the state of the world while the principal's preference increases.

In line with [CS], [D] and [ADM], we assume that information is soft in the sense that it cannot be verified by the principal. Furthermore, the principal is not able to credibly condition his decision on the received information, thus communication takes the form of cheap talk (see [CS]). The only formal mechanism the principal is able to commit to is the ex ante allocation of decision rights. In this setting, assuming that the state of the world is uniformly distributed, we first characterize the equilibria of the communication game for an arbitrary linear preference divergence between principal and agent. Second, we compare the principal's payoff from retaining authority and communicating with the agent to his utility from delegating the agent decision rights.

Our first major finding is that communication can be infinitely informative in equilibrium, in the sense that the agent can send one out of countably infinitely many messages, if and only if the state of agreement between the principal and the agent is realized with positive probability – provided that the agent prefers more extreme projects than the principal. This finding implies the respective results by both [CS] and [ADM] as limiting cases. While [CS] shows that the agent can only choose between finitely many messages in equilibrium if the principal and the agent never agree about the right course of action, [ADM] assume that the agreement state coincides with the expected state of the world and find that communication can be infinitely informative. However, if the agent is conservative in the sense that he understates the impact of extreme states, we show that communication can never be infinitely informative. Finally, if the agent's preference is even reversed, communication does not reveal any information.

[1] To the best of our knowledge, the only paper considering general linear biases, like the present analysis, is Melumad and Shibano (1991). However, they focus on finite equilibrium partitions, while we characterize when infinitely informative communication is feasible.

Essential for these findings is that the agent's incentive to misrepresent his private information is qualitatively different depending on whether he is more extreme or more conservative than the principal. If the agent is extreme, messages far away from the agreement state have to be very noisy in equilibrium in order to make the exaggeration of private information costly for the agent. In contrast, if the state of the world is close to the agreement state, the agent has only slight incentives to distend it and the noise in communication becomes infinitesimally small, yielding an infinitely fine choice set of messages in equilibrium. On the other hand, a conservative agent tends to distort the report about the actual situation towards the agreement state, hence messages around the agreement state are very noisy in equilibrium. While this noisiness decreases if the realized state becomes more extreme, it never reduces to zero as the preference divergence increases in the extremeness of the state. Hence, a conservative agent chooses between finitely many different messages in equilibrium.

Our second major result is that communication may be informative *and* dominate delegation from the principal's point of view if the agent's bias is sufficiently large. This stands in contrast to [D], who shows that in case the agent exhibits a purely systematic bias, delegation is the principal's preferred policy as soon as communication is informative, rendering informative communication impossible from an ex ante perspective.

Finally, we test the robustness of these two major findings with regard to general distributions of the state of the world. In this context, we provide necessary and sufficient conditions for communication to be informative and for communication to be infinitely informative. Comparing the performance of delegation and communication for symmetric distributions with compact support, we find that the principal's expected utility loss incurred by communication is an infinite times larger than his loss from delegation if the preference divergence vanishes – which is in line with [D]. If the agent's bias is large, however, we show that the principal prefers communication to delegation as soon as the agent's preference is extreme enough with regard to the principal's, independent of the distribution of the state of the world. Hence, it is always possible to find a bias such that communication is informative *and* dominates delegation. This is in contrast to [D], as his result implies that communication is uninformative whenever it is the preferred policy, given that the variance of the state of the world is sufficiently large.

Further Related Literature.

Allowing for the preference divergence to depend on the state of the world, we first see that communication can become more informative than assuming a purely systematic bias. Building on the seminal paper by [CS], several articles propose other remedies to improve the precision of information transmission, such as including noise into the agent's signal (Blume et al., 2007), repeated communication (Farrell and Gibbons, 1989; Krishna and Morgan, 2004), multiple receivers (Battaglini, 2002), or multi-dimensional action spaces (Chakraborty and Harbaugh, 2007, 2010). Goltsman et al. (2009) analyze the efficiency bounds of communication by comparing different institutional arrangements. In fact, a fully revealing communication equilibrium can be achieved if the sender incurs lying cost or the receiver is credulous (Kartik et al., 2007), or if the receiver has private information while the sender suffers honesty concerns (Olszewski, 2004). However, a vast majority of the literature following [CS] assumes the agent's bias to be purely systematic. We are aware of two exceptions, Alonso et al. (2008, 2010), who explicitly consider strategic communication with a state-dependent bias.

Second, our model is based on the assumption that the principal cannot commit to message-contingent project choices, following [D]. In contrast, a recent strand of the delegation literature assumes that the principal can credibly commit to a mechanism selecting a predefined project for any given message by the agent (e.g. Alonso and Matouschek, 2008; Kováč and Mylovanov, 2009; Koessler and Martimort, 2012). The optimal mechanism from the principal's point of view then mostly takes the form of a partly separating equilibrium, that is, the principal rubber-stamps the agent's recommendation in a predefined range of the state of the world. It is worth noting that the principal chooses menu delegation in this framework if his commitment power is reduced, which is equivalent to a partition equilibrium in the analysis of strategic communication (see Melumad and Shibano, 1991, or Alonso and Matouschek, 2007). An alternative perspective on delegation is that it might provide incentives for the agent to acquire relevant information (Aghion and Tirole, 1997) or to exert project implementation effort (Bester and Krähmer, 2008). However, we assume the information structure to be exogenous and refrain from modeling the project implementation stage.

The remainder of the chapter is structured as follows. Section 2.1 introduces the model for uniformly distributed states of the world. Communication equilibria are characterized in Section 2.1.1, while Section 2.1.2 analyzes the optimal allocation of authority. Section 2.2 generalizes the analysis to arbitrary distributions and Section 2.3 concludes. All technical proofs are gathered in the appendix.

2.1 The Model

We consider an organization which has to decide about the implementation of a project $x \in X \subset \mathbb{R}$. A principal, such as the owner of the firm or the group of shareholders, hires an agent (e.g. a manager) to implement the project, however the principal has the formal authority to decide which project shall be implemented – for instance because he controls firm assets or other resources needed for the implementation. Projects are assumed to differ only in one dimension.[2] A common way to interpret this dimension is to think of investment sizes or the allocation of (human) resources to a specific project. While there is an infinity of potential project choices, only one of them can be implemented.

Preferences. Each project choice is associated with a monetary gain or a personal benefit $U_P(x, \theta)$ for the principal and $U_A(x, \theta, b)$ for the agent, where $\theta \in \Theta := [-L, L]$ denotes a randomly drawn state of the world and $b \in \mathbb{R}^2$ parametrizes the dissonance between principal and agent. Without loss of generality, we assume that the state of the world reflects the principal's optimal project choice, for example the optimal investment level

[2] Alternatively, projects differ along several qualitative dimensions, but only one dimension causes disagreement between the principal and the agent.

for a new plant from the perspective of the firm. Hence, the principal wants to match the state of the world and his preferred project, conditional on the realization of θ, is given by $x_P(\theta) = \theta$.

The agent, in turn, exhibits deviating preferences. We assume that the agent might exhibit both a systematic *and* a state-dependent bias. In other words, we allow for a general linear bias $b(\theta) := b_a + (b_m + 1)\theta$, such that the agent's preferred project is given by $x_A(\theta) = b(\theta)$, with $b_m, b_a \in \mathbb{R}$. Thus, preferences are everywhere aligned if and only if $b_a = b_m = 0$. Slightly abusing notation, we often refer to the bias as $b = (b_m, b_a) \in \mathbb{R}^2$, which parametrizes the preference divergence by its multiplicative part (the agent's relative sensitivity to changes in θ), b_m, and its additive part (the agent's systematic bias), b_a. Note that the agent either reacts stronger ($b_m > 0$) or weaker ($b_m < 0$) to changes in θ than the principal. In particular, his preference might even be reversed with regard to the principal's ($b_m < -1$).

By allowing for a general linear preference divergence between the principal and the agent, our setup embeds the models of [CS], [D] and [ADM]. Indeed, if $b_m = 0$, the agent exhibits a purely systematic bias towards larger (smaller) projects, such that his optimal project choice differs from the principal's preference by some constant $b_a \in \mathbb{R}$ and is given by $\theta + b_a$. In our investment example, this is reminiscent of a production manager willing to undertake inefficiently high investments into a new plant, independent of the market situation, thereby growing his division beyond the optimal size in order to strengthen his position in the organization (see, for instance, Jensen, 1986). This kind of systematic bias is at the heart of both [CS] and [D]. On the other hand, if we fix $b_a = 0$ and assume $b_m > 0$, the model setup is applicable to the situation in [ADM], where two division managers in an organization need to coordinate and adapt their strategy to local conditions, yielding a natural multiplicative structure. In their model, the agent prefers project $(b_m + 1)\theta$ with $b_m > 0$ in any given state of the world. Hence, both principal and agent agree on the optimal project choice in state $\theta_0 = 0$, but increasingly disagree the more θ deviates from there. Table 2.1 illustrates the extended scope our model covers.

	$b_m < 0$	$b_m = 0$	$b_m > 0$
$b_a < 0$		[CS]* [D]**	
$b_a = 0$			[ADM]*
$b_a > 0$			

*Characterization of communication equilibria

**Optimal allocation of authority

Table 2.1: Illustration how the special cases of [CS], [D] and [ADM] are embedded in our model. Communication equilibria have been analyzed for $b_m = 0$ ([CS]) and $b_a = 0$, $b_m > 0$ ([ADM]). [D] investigates the optimal allocation of authority in case $b_m = 0$. We analyze both aspects for all $b = (b_m, b_a) \in \mathbb{R}^2$.

We assume that utility functions are strictly concave in x for a given θ, such that the optimal project choice for any given state of the world is unique. For simplicity, we assume a quadratic functional form, that is

$$\begin{aligned} U_P(x, \theta) &= -(x - \theta)^2 \\ U_A(x, \theta, b) &= -(x - b(\theta))^2. \end{aligned}$$

Information Structure. The state of the world, θ, is assumed to be uniformly distributed on $\Theta = [-L, L]$, where $L > 0$. The variance of the distribution is denoted by σ^2 with $\sigma^2 = \frac{L^2}{3}$ for the uniform distribution.[3] While the agent fully learns the state of the world, the principal stays uninformed. The superior information of the agent can be interpreted as a manager having deeper knowledge of the market environment, while shareholders do not have business insights. However, we assume that the agent's information is soft in the sense that it cannot be verified by the principal. All other aspects of the relationship, such as the agent's bias $b(\theta)$ and the maximal size of his informational advantage due to the uncertainty of the environment, L, are common knowledge.

[3] Though the uniform distribution exhibits very special properties, it is conventional to use it in this kind of problems. We relax this assumption and consider general distributions in Section 2.2.

Contracts. We assume that the principal has the formal right to decide about the project, for instance because he controls the assets or human resources of the organization, or, referring to our investment example, he is in charge of its budget. We follow the literature on incomplete contracts (e.g. Grossman and Hart, 1986; Hart and Moore, 1990) and assume that the project choice itself is not contractible. The principal can only commit to the ex ante allocation of decision rights, for example by granting the agent the formal right to underwrite investment contracts.[4] Once this decision right is allocated, it cannot be transferred before the project decision is made. If the principal delegates the formal authority to the agent, the agent can initiate a project without the principal's support. On the other hand, if the principal retains decision authority, he initiates a project which is then implemented by the agent and cannot be changed. In the latter case, the principal can communicate with the better informed agent. However, the principal can neither commit to message-contingent monetary transfers, nor to a mechanism conditioning the project choice on the received information. Thus, communication takes the form of cheap talk.

The timing of the model is as follows. (i) The principal decides whether or not to delegate the formal decision right about the project to the agent. (ii) The agent learns the state of the world θ and initiates her preferred project if he has decision authority. If the principal has retained the decision rights, he may ask the agent for a recommendation and then initiates a project. (iii) The agent implements the project.

The following section consists of two parts. In Section 2.1.1, we consider the case where the principal keeps the decision right and characterize the communication equilibria for all $b := (b_m, b_a) \in \mathbb{R}^2$, thereby generalizing the leading example in [CS] and the model by [ADM]. Section 2.1.2 considers delegation as an alternative to communication and analyzes the optimal allocation of decision rights.

[4]Similar to [D], we consider delegation a binary decision and abstain from analyzing optimal (partial) delegation mechanisms.

2.1.1 Communication Equilibria

The following characterization of communication equilibria for a general linear bias derives from [CS], who analyzed the case where the principal and the agent disagree about the preferred project in all states of the world.

We assume that the principal retains authority, but communication of soft information between principal and agent is feasible. Since the project choice is not contractible, the principal chooses his preferred project given his belief about the state of the world, μ, where μ is a distribution on Θ. Communication is cheap in the sense that it can only influence the principal's belief, thus μ may depend on the message $m \in \Theta$ sent by the agent. The agent, in turn, strategically transmits a possibly noisy message in order to manipulate the principal's belief.

Formally, a Perfect Bayesian Equilibrium of the communication game is given by a mixed-strategy communication rule $p(m|\theta)$ of the agent, a decision rule $x(m)$ for the principal,[5] and belief functions $\mu(\theta|m)$ for the principal such that

- The agent's communication rule is optimal given the principal's decision rule.
  For all $\theta \in \Theta$, if m^* is in the support of $p(\cdot|\theta)$, then m^* solves

$$\max_{m \in \Theta} U_A(x(m), \theta, b).$$

- The principal's decision rule is optimal given his belief function.
  For each message m, $x(m)$ solves

$$\max_x \int_\Theta U_P(x, \theta)\mu(\theta|m)d\theta.$$

- The belief functions are derived by Bayes' rule whenever possible.

$$\mu(\theta|m) = \frac{p(m|\theta)}{\int_\Theta p(m|t)dt}.$$

[5] Since $U_P(x, \theta)$ is strictly concave in x, the principal never uses mixed strategies in equilibrium; see [CS].

Following the argument by [CS], all equilibria in this communication game are given by a partition of Θ, where the agent introduces noise to his signal by only specifying the partition interval containing θ. Correspondingly, the principal implements his preferred project given that θ lies in the specified partition interval.

In their leading example, [CS] consider the case $b_m = 0$ and $b_a \neq 0$, that is, $x_A(\theta) \neq x_P(\theta)$ for all $\theta \in \Theta$. In this case, any equilibrium partition consists of at most finitely many intervals. Considering a general linear bias, the principal's and the agent's preferences are aligned in $\theta_0 := -\frac{b_a}{b_m}$, which realizes with positive probability if $\theta_0 \in \Theta = [-L, L]$. In the following, we show that Θ can be partitioned into countably infinitely many intervals in the most informative equilibrium of the communication game if and only if this agreement state θ_0 lies in Θ and $b_m > 0$. This is in line with an analysis by [ADM], who consider a purely multiplicative positive bias, i.e. $b_a = 0$ and $b_m > 0$. Here, the agreement state is $\theta_0 = 0$, which is contained in $\Theta = [-L, L]$, thus the most informative equilibrium partition consists of infinitely many intervals.

The Case $b_m > 0$

We first analyze communication equilibria for the case $b_m > 0$, indicating that the agent is extreme in the sense that $x_A(\theta)$ reacts stronger to a change in θ than $x_P(\theta)$.

Let $a^N := (a_0, a_1, \ldots, a_{N-1}, a_N)$ denote a partition of $[-L, L]$ in N intervals, where $-L = a_0 < a_1 < \ldots < a_{N-1} < a_N = L$. If the principal believes that θ is uniformly distributed on $[a_{i-1}, a_i]$ for some $i \in \{1, \ldots, N\}$, he implements

$$x_i = \arg\max_x \int_{a_{i-1}}^{a_i} U_P(x, \theta) d\theta = E[\theta | \theta \in [a_{i-1}, a_i]] = \frac{a_{i-1} + a_i}{2}.$$

In equilibrium, given that the agent truthfully reports the interval containing θ by sending a message $m \in [a_{i-1}, a_i]$, the principal's best response is to implement $x(m) = x_i$.

On the other hand, if $\theta \in [a_{i-1}, a_i]$, the agent sends a truthful message $m \in [a_{i-1}, a_i]$ if $U_A(x_i, \theta, b) \geq U_A(x_j, \theta, b)$ for all $j \in \{1, \ldots, N\}$. As $x_A(\theta)$ is increasing in θ for $b_m > 0$, this requirement is equivalent to the agent being indifferent between implementing x_i and x_{i+1} if $\theta = a_i$, which yields the *arbitrage condition*

$$U_A(x_{i+1}, a_i) = U_A(x_i, a_i) \quad \forall i \in \{1 \dots N-1\}\,. \tag{A}$$

As $a_i < a_{i+1}$, (A) is equivalent to

$$a_{i+1} - a_i = a_i - a_{i-1} + 4(b_a + b_m a_i).$$

Condition (A) illustrates the "screening" resulting in equilibrium. The agent is induced to tell the truth by the fact that messages are noisy if they are "large" ("small") whenever the agent exhibits a positive (negative) systematic bias, or if the realized state of the world is far away from θ_0. Intuitively, the agent's message contains more noise, and is therefore less credible, if the agent recommends a project which lies in the direction of his bias.

More precisely, the size of the partition intervals increase by the difference between the agent's and the principal's preferred project in the cutoff point a_i between two partition intervals. It turns out that the size of the partition intervals decreases up to the interval $I_0 \ni \theta_0$, which includes the intersection of the principal's and the agent's preferences, and increases thereafter. The smallest interval I_0 lies further to the left (right) the larger (smaller) b_a is, and if $|b_a| > b_m L$, I_0 is the first or last interval, thus interval lengths increase (decrease) from left to right, resulting in a finite partition. The partition is symmetric in the sense that the ith interval is of equal size as the $(N-i+1)$th one if and only if $b_a = 0$.

Note that, if $b_m = 0$, we obtain the standard result by [CS]. In this case, interval lengths in an equilibrium partition constantly increase from left to right (right to left) by $4|b_a|$ if $b_a > 0$ ($b_a < 0$), thus inducing a finite partition of the compact set $[-L, L]$. If $b_a = 0$, the relation between the sizes of partition intervals in equilibrium can be found in [ADM].

This yields Proposition 1, where we characterize the maximum number of partition intervals in equilibrium, $N(b)$, for any bias $b = (b_m, b_a)$ with $b_m > 0$.

Proposition 1 (Communication Equilibria)**.** *If* $b_m > 0$*, there exists a positive integer* $N(b) \leq \infty$ *such that, for every integer* N *with* $1 \leq N \leq N(b)$*, there exists at least one Perfect Bayesian Equilibrium* $(p(m|\theta), x(m), \mu(\theta|m))$ *of the communication game, where*

1. $x(m) = \frac{a_{i-1}+a_i}{2}$ *if* $m \in (a_{i-1}, a_i)$
2. $p(m|\theta)$ *is uniform on* $[a_{i-1}, a_i]$ *if* $\theta \in (a_{i-1}, a_i)$
3. $\mu(\theta|m))$ *is uniform on* $[a_{i-1}, a_i]$ *if* $m \in (a_{i-1}, a_i)$
4. $a_{i+1} - a_i = a_i - a_{i-1} + 4(b_a + b_m a_i)$ *for all* $i \in \{1 \dots N-1\}$
5. $-L = a_0 < a_1 < \dots < a_{N-1} < a_N = L$*.*

Moreover, $N(b)$ *is implicitly given by the largest positive integer* N *such that*

$$|b_a| < \frac{(c^N+1)(c^{N-1}+1)}{(c^N-1)(c^{N-1}-1)} b_m L, \tag{M}$$

where $c := 1 + 2b_m + \sqrt{(1+2b_m)^2 - 1}$*. In particular,* $N(b) = \infty$ *if and only if*

$$|b_a| \leq b_m L.$$

In the following, we say that *communication is informative* if there exists a Perfect Bayesian Equilibrium of the communication game with $N \geq 2$. Moreover, we define communication in b^1 to be *more (less) informative* than communication in b^2 if and only if $N(b^1) > (<) N(b^2)$. If $N(b) = \infty$, we say that communication is *infinitely informative.* Note that in this case, communication is still noisy because $[-L, L]$ is partitioned into countably infinitely many intervals, where the agent truthfully reveals the interval which contains θ. In contrast, if the agent directly communicates θ, which is the case if and only if $b_m = b_a = 0$, we refer to communication as being *perfect.*

Proposition 1 shows in particular that the number of partition intervals in equilibrium is not bounded above if and only if the state where the principal and the agent agree about the project choice, $\theta_0 = -\frac{b_a}{b_m}$, realizes with positive probability. In that case, the agent's incentive to add noise to his message is rather weak as θ approaches θ_0 because the principal's preferred project does not deviate too much from the agent's preference. This result generalizes both the analysis by [CS] and by [ADM].

First, if $b_m = 0$ as in [CS], the preferences of principal and agent do not intersect, and the number of partition intervals is finite in equilibrium since interval lengths are constantly increasing. Indeed, taking the limit of the right-hand side in (M) as b_m vanishes, or equivalently, as c approaches 1, yields the corresponding result in [CS], as summarized in Corollary 1.

Corollary 1. *If $b_m = 0$, the number of partition intervals in equilibrium is finite, i.e. $N(b) < \infty$, and $N(b)$ is given by the maximum N such that*

$$|b_a| < \frac{L}{N(N-1)}.$$

Second, [ADM] consider the case $b_a = 0$ and $b_m > 0$, thus the preferences of the principal and the agent intersect in $\theta_0 = 0$, which realizes with positive probability for any $b_m > 0$. In this case, we can find a partition equilibrium for any $N \in \mathbb{N}$, such that interval lengths decrease up to $\theta_0 = 0$, and symmetrically increase thereafter. The status quo, θ_0, either lies in the middle of the centered partition interval (if N odd) or constitutes the interval border between the two middle intervals (if N even). Their result is obtained by assuming $b_a = 0$ in Proposition 1, which yields $N(b) = \infty$ for all $b = (0, b_m)$ with $b_m > 0$.

Corollary 2. *If $b_a = 0$, for all $N \in \mathbb{N}$ there exists an equilibrium of the communication game with N partition intervals, that is, $N(b) = \infty$ for all $b_m > 0$.*

Comparative Statics. Qualitatively analogous to [CS], if we fix the state-dependent part $b_m > 0$ of the bias, a reduction in the systematic preference divergence, $|b_a|$, results in a more informative communication equilibrium. In fact, $N(b)$ increases up to infinity for small values of $|b_a|$. In other words, reducing the systematic bias of the agent eventually yields a communication equilibrium with infinitely many partition intervals. On the other hand, if we set the systematic bias b_a to some non-zero value, reducing the slope difference between the preferences, b_m, eventually renders communication *less* informative, where the number of partition intervals in the most informative equilibrium depends on the size of $|b_a|$ if b_m vanishes. Intuitively, if the slopes of both preferences become more and more aligned, the potential state of agreement moves further to the border of the support of θ until it is not realized any more with positive probability and $N(b) < \infty$. The further this agreement state moves away from the possibility of being realized, the larger is the minimal preference divergence between principal and agent in the support of θ and the less informative is the communication equilibrium. Finally, considering a fixed bias $b = (b_m, b_a)$ with $b_m > 0$, increasing the agent's informational advantage, L, results in more informative communication because the minimal preference divergence within the relevant range of θ decreases.

Equilibrium Selection. It remains to argue which of the equilibria will be played by the agent and the principal. In an equilibrium with N partition intervals, the principal's ex ante expected utility is given by

$$E[U_P(x, \theta)] = -\sigma_N^2,$$

where $\sigma_N^2 := E[(\theta - E[\theta|m])^2]$ denotes the residual variance of θ in an equilibrium with N partition intervals. Since σ_N^2 decreases in N if $b_m > 0$, as we will show in Lemma 2, the principal prefers the equilibrium with the maximum number of partition intervals. Following the argument in [ADM], we assume that the organization is able to coordinate on the equilibrium that maximizes corporate profits, represented by the principal's utility. Thus, we focus on the equilibrium with the largest number of partition intervals if $b_m > 0$.

It is worth noting that in case the preference divergence depends on the state of the world, the agent's ex ante expected utility is not necessarily increasing in the number of intervals in the equilibrium partition. In fact, the interaction of the project implemented by the principal after communication and the state of the world enters the agent's expected utility, thus it depends on the particular bias $b = (b_m, b_a)$ whether the agent is better off with a finer equilibrium partition. In contrast, if the dissonance is independent of the state of the world, as in [CS], the agent's expected loss equals the residual variance plus the square of the systematic bias, thus the more informative equilibrium Pareto-dominates the less informative one.

It remains to confirm that the construction in Proposition 1 continues to constitute a Perfect Bayesian Equilibrium if the number of partition intervals is countably infinite. The according result in [ADM] turns out to hold for the general context considered here.

Proposition 2. *Let $|b_a| \leq b_m L$. Then the limit of strategy profiles $(p(m|\theta), x(m))$ and beliefs $\mu(\theta|m)$ for $N \to \infty$ is a Perfect Bayesian Equilibrium.*

The proof of Proposition 2 follows closely the proof of Proposition 2 in [ADM]. Figure 2.1 illustrates the number of partition intervals in the most informative equilibrium of the communication game, for any given $b = (b_m, b_a)$ with $b_m > 0$.

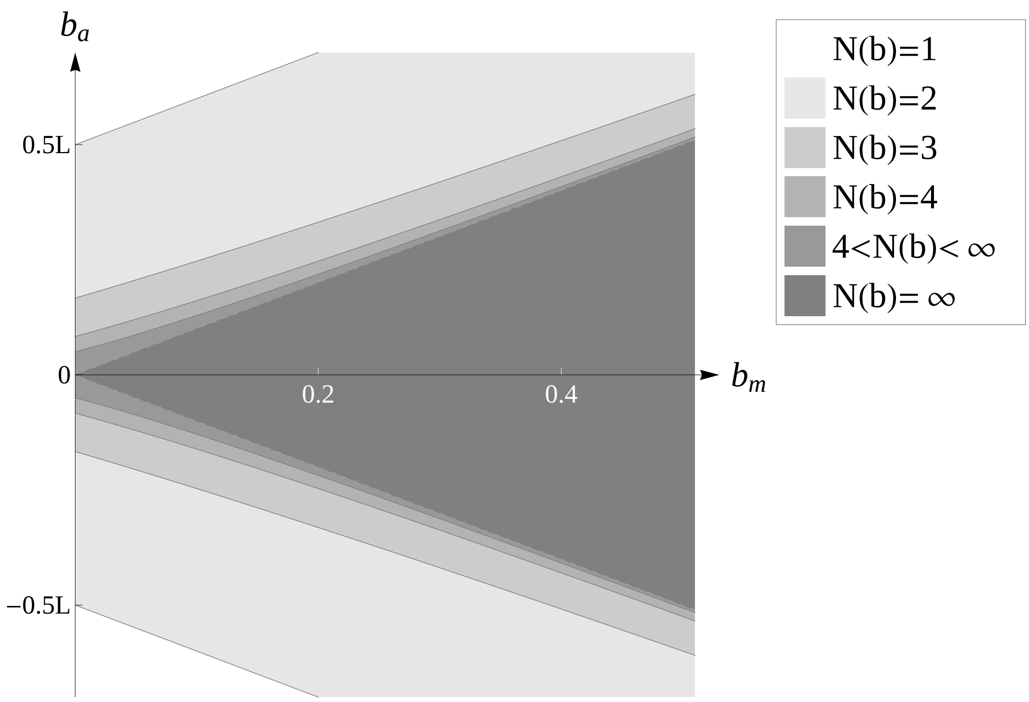

Figure 2.1: Maximum number of partition intervals in an equilibrium of the communication game for a given $b = (b_m, b_a)$ with $b_m > 0$.

The Case $b_m < 0$

If the multiplicative part of the agent's bias, b_m, is negative, we distinguish two cases. First, if $b_m < -1$, the agent's preference is reversed with regard to the principal's preferred project. Hence, in order to induce the agent to reveal the interval containing θ, the project implemented by the principal has to decrease in θ. But such a decreasing pattern can never be a best response of the principal to a truthful message by the agent, thus communication is not informative in equilibrium.

Second, if $b_m \in (-1, 0)$, the agent is conservative in the sense that his preference reacts weaker to a deviation of θ from the agreement state than the principal's. Though this is not obvious at first glance, it turns out that the direction of the preference divergence has a fundamental impact on the informativeness of communication. In fact, in Section 2.1.1, we considered an extreme agent who has an incentive to exaggerate the deviation of θ from the agreement state. In order to induce truth telling, the agent's message thus has to be rather noisy in equilibrium for states far away from the agreement state θ_0. As θ approaches θ_0, however, truth telling is a negligible issue and communication can be infinitely informative.

In contrast, if $b_m \in (-1, 0)$, the agent is conservative, hence he underestimates deviations of the state of the world from the agreement state. In this case, the agent has an incentive to bias his report towards θ_0, thus an equilibrium has to induce more noise on messages around the agreement state. However, for states near the borders of $[-L, L]$, the preference divergence between principal and agent is bounded away from zero, thus communication has to include some minimum noise in order to induce truth telling by the agent. Hence, communication cannot be infinitely informative in this case, as opposed to the case of an extreme agent.

From a technical point of view, while the expression for a_i in terms of c in the proof of Proposition 1 continues to hold, the main monotonicity results are no longer true. Indeed, if $b_m < -1$, it is not clear whether c is positive or negative, and for $b_m \in (-1, 0)$,

c is no longer a real number.[6] Hence, communication equilibria have to be calculated recursively on a case-by-case basis. Still, we obtain a clear result for when communication is informative.

Proposition 3 (Communication Equilibria)**.** *If $b_m < 0$, there exists a positive integer $N(b) < \infty$ such that, for every integer N with $1 \leq N \leq N(b)$, there exists at least one Perfect Bayesian Equilibrium $(p(m|\theta), x(m), \mu(\theta|m))$ of the communication game satisfying conditions 1 to 5 in Proposition 1.*

In particular, communication is informative ($N(b) \geq 2$) if and only if $b_m \geq -1$ and

$$|b_a| < \frac{|2b_m + 1|}{2} L.$$

Furthermore, $N(b) > 2$ is only possible if $b_m > -\frac{1}{2}$.

Note that, if the agent and the principal never agree about the project choice, the agent's systematic bias has to be small enough relative to his informational advantage in order to render communication informative. Corollary 3 derives the respective result by [CS].

Corollary 3. *If $b_m = 0$, communication is informative ($N(b) \geq 2$) if and only if*

$$|b_a| < \frac{L}{2}.$$

In contrast to the case $b_m > 0$, the length of partition intervals increases up to the interval $I_0 \ni \theta_0$, and decreases thereafter. Due to its compactness, the support $[-L, L]$ thus can not contain an accumulation point of $\{a_i\}$, hence $N(b) < \infty$. Figure 2.2 provides an intuition for the informativeness of communication for any bias $b = (b_m, b_a)$ with $b_m < 0$.

Comparative Statics. Analogous to [CS], we find that reducing the systematic bias $|b_a|$ for a given $b_m < 0$ increases the informativeness of communication up to a finite number, which depends on b_m. However, if we fix the systematic bias b_a, the informativeness of

[6] Of course, we still have $a_i \in \mathbb{R}$.

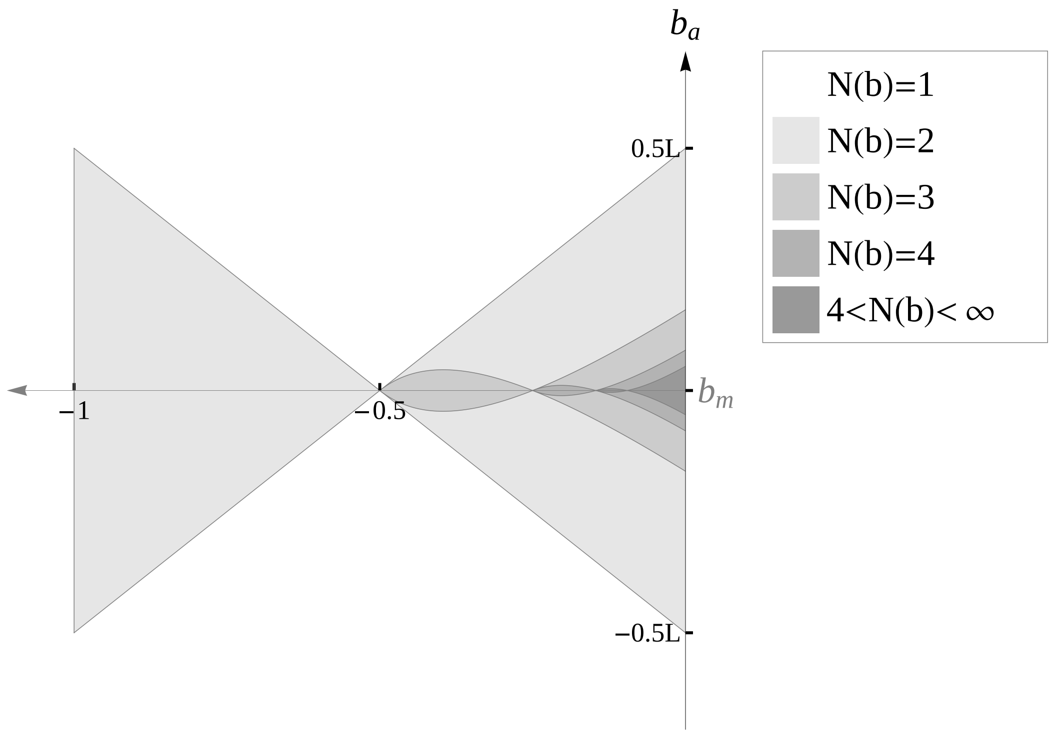

Figure 2.2: Maximum number of partition intervals in an equilibrium of the communication game for a given $b = (b_m, b_a)$ with $b_m < 0$.

communication is no longer monotone in $|b_m|$. Still, the most informative equilibrium is reached if $|b_m|$ vanishes. In contrast to the case $b_m > 0$, aligning the sensitivity of the agent's and the principal's preference with regard to deviations in θ eventually results in more informative communication. The reason here is that the agent's incentive to bias his report towards the agreement state decreases in $|b_m|$, hence the noisiness of messages in equilibrium is reduced for states near θ_0 and the equilibrium partition can be finer. Finally, if the agent's informational advantage, L, increases, communication is more likely to be informative ($N(b) \geq 2$).

Equilibrium Selection. In addition to the non-monotonicity of the informativeness of communication, we find that, if the agent reacts weaker to changes in θ than the principal ($b_m \in (-1, 0)$), the residual variance in an equilibrium with N partition intervals, σ_N^2, is no longer monotone in N. Hence, the most informative equilibrium is not necessarily preferred to the remaining ones from the principal's point of view, and we cannot apply the equilibrium selection criterion as in the case $b_m > 0$. Still, we are able to state that whenever communication is informative, the principal prefers the coarsest communication equilibrium with $N = 2$ to taking an uninformed decision.

Remark 1. *The principal prefers communication with two intervals in the equilibrium partition to uninformative communication, i.e.* $\sigma_2^2 < \sigma_1^2 = \sigma^2 = \frac{L^2}{3}$.

Yet, we will see in Section 2.1.2 that the analysis of the optimal allocation of authority does not rely on equilibrium selection if $b_m \in (-1, 0)$.

2.1.2 Communication Versus Delegation

After having characterized communication equilibria for all $b \in \mathbb{R}^2$, we now consider delegation as an alternative to communication. By comparing the principal's expected utilities of both alternatives, we analyze the optimal allocation of authority from the principal's perspective.

If the principal formally delegates decision authority to the agent, he expects a disutility, or loss, of

$$\delta^2 := E\left[(b_a + (b_m + 1)\theta - \theta)^2\right] = b_a^2 + b_m^2\sigma^2 = b_a^2 + b_m^2\frac{L^2}{3}.$$

In contrast, his respective expected loss if he retains authority and communicates with the agent, is given by the residual variance of θ, conditional on receiving a message $m \in \Theta$ in an equilibrium with N partition intervals,

$$\sigma_N^2 = E[(\theta - E[\theta|m])^2] = \frac{1}{2L}\sum_{i=1}^{N}\int_{a_{i-1}}^{a_i}(\theta - E[\theta|m \in [a_{i-1}, a_i]])^2 d\theta.$$

Lemma 1 determines the residual variance depending on the number of partition intervals for any $b \in \mathbb{R}^2$.

Lemma 1. *The residual variance in an equilibrium with N partition intervals is given by*

$$\begin{aligned}\sigma_N^2 &= \frac{L^2}{12}\frac{(c-1)^2}{(c^N-1)^2}\left[\frac{c^{2N}+c^N+1}{c^2+c+1}+3c^{N-1}\right] \\ &+ \frac{4c^2b_a^2}{(c-1)^2(c^N+1)^2}\left[\frac{c^{2N}+c^N+1}{c^2+c+1}-c^{N-1}\right],\end{aligned}$$

with $c = 1+2b_m+\sqrt{(1+2b_m)^2-1}$. Furthermore, $\sigma_N^2 \in \mathbb{R}$ for all $b \in \mathbb{R}^2$.

The latter remark ensures that, even if $c \in \mathbb{C}$ for $b_m \in (-1,0)$, the residual variance is always a real number. In particular, if the agent's and the principal's preferences have the same slope, i.e. $b_m = 0$, the residual variance coincides with the result in [CS]. Furthermore, the residual variance in case $b_a = 0$ can be found in [ADM].

Corollary 4. *If b_m vanishes, the residual variance in an equilibrium with N partition intervals is given by*

$$\lim_{b_m\to 0}\sigma_N^2 = \frac{L^2}{3N^2}+\frac{N^2-1}{3}b_a^2.$$

In the following, we will say that communication *dominates* delegation if the principal's expected disutility from communication is smaller than his expected disutility from delegation, and vice versa.

The Case $b_m > 0$

In order to compare the expected loss incurred by delegation to the expected loss from communication, we first show that the quality of communication increases in the number of partition intervals up to the maximally possible number, $N(b)$, if $b_m > 0$.

Lemma 2. *For any $b \in \mathbb{R}^2$ with $b_m > 0$, the residual variance σ_N^2 decreases in N up to $N(b)$. Furthermore, if $|b_a| \leq b_m L$, σ_N^2 decreases for all $N \in \mathbb{N}$, with*

$$\lim_{N\to\infty}\sigma_N^2 = \frac{b_m}{4b_m+3}\frac{L^2}{3}+\frac{b_a^2}{b_m(4b_m+3)}.$$

It is worth noting that the residual variance increases in N for $N > N(b)$. Hence, the maximum number of partition intervals can also be derived from the maximum N such that σ_N^2 decreases. Proposition 4 implicitly defines curves in the (b_m, b_a)-plane for any $N(b) \in \mathbb{N}$, such that the principal prefers to communicate for biases to the right of the respective curve, and delegates otherwise. It turns out that these curves converge to the vertical line $b_m = \frac{1}{4}$ for $N \to \infty$, which delimits the area where communication performs better than delegation if an equilibrium partition with infinitely many intervals is possible.

Proposition 4. *Let $b_m > 0$. Then communication dominates delegation in $b \in \mathbb{R}^2$ if and only if*

$$\delta^2 > \sigma_{N(b)}^2.$$

If $|b_a| \leq b_m L$, communication dominates delegation if and only if $b_m > \frac{1}{4}$.

Figure 2.3 illustrates the curves limiting the area where communication outperforms delegation for some $N(b)$. We see that communication yields the highest payoffs, as indicated by the most inner curve, exactly in an equilibrium with $N(b)$ partition intervals for the respective $b \in \mathbb{R}^2$. Connecting the segments of these inner curves for the respective $N(b)$ yields a separation between the inner area, where the principal delegates, and the area outside the thick line, where communication is preferred.

In particular, we show that the case $b_m = 0$, as considered in [D], is a limiting case. [D] finds that communication is uninformative if it yields higher payoffs to the principal than delegation, which is the case for large $|b_a|$. Indeed, considering only the vertical axis, we see that $N(b) = 1$ if $b_m = 0$ and $|b_a|$ is larger than some value indicated by the thick line. However, if we allow for a general linear bias, we find that communication can outperform delegation *and* be informative if $b = (b_m, b_a)$ is outside the thick-bordered area.

Comparative Statics. If $b_m = 0$, we find in line with [D], that increasing the agent's informational advantage, L, renders delegation more likely. This is due to the fact that the principal incurs a constant utility loss of b_a^2 by delegating, while his minimal loss by communicating is given by $\frac{L^2}{3N(b)^2} + \frac{N(b)^2-1}{3} b_a^2$. Since $N(b)$ increases weakly (step-wise) in

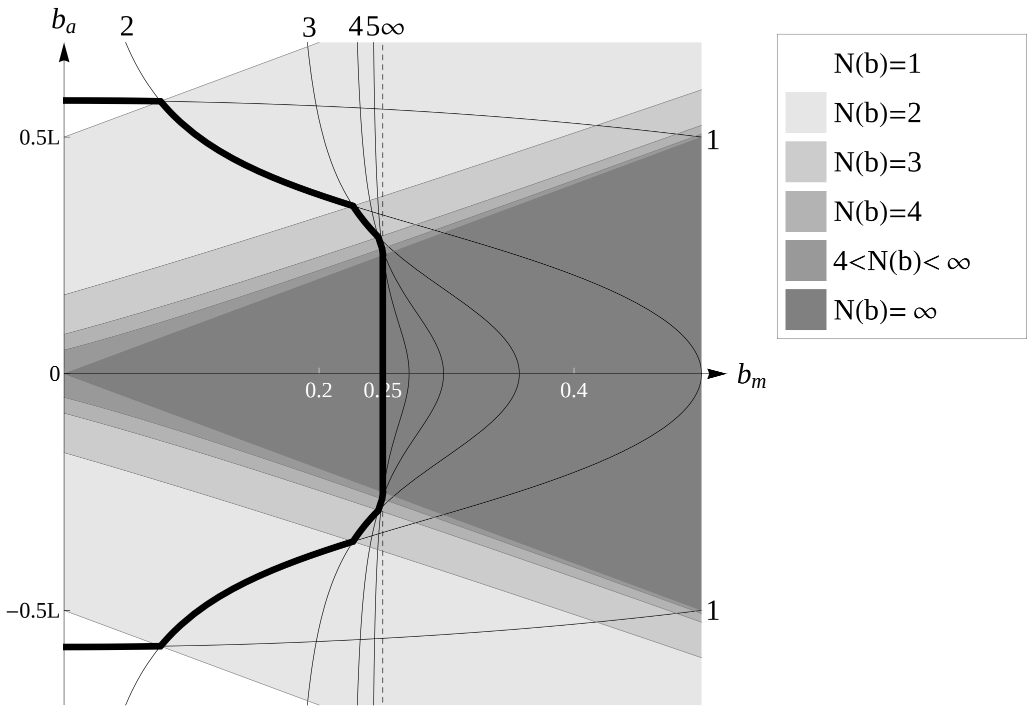

Figure 2.3: Comparison of the principal's expected payoffs from delegation and communication in case $b_m > 0$. Black curves indicate that for $|b_a|$ larger than the respective curve, communication performs better than delegation in an equilibrium with $N(b)$ partition intervals, where $N(b)$ is given by the integer attached to the respective curve. Assuming that the most informative equilibrium is played, communication performs better than delegation outside of the area limited by the thick line.

L, the principal's minimal loss of communication increases, thus delegation becomes more likely for large L. In contrast, if we consider $b_m > 0$, we observe that an increase in L can promote the likeliness of either communication or delegation, depending on b_m. The reason for this qualitative change is that the principal risks to incur a relatively huge utility loss from delegation if the realized state is far away from the agreement state and $b_m > 0$. Hence, increasing the support $[-L, L]$ of θ increases the principal's expected utility loss from delegation. In turn, the informativeness of communication eventually reaches infinity as L increases. Furthermore, the expected utility loss from both communication and delegation are of second order in L, thus the size of b_m determines whether the loss of communication or delegation increases faster with the agent's informational advantage.

The Case $b_m < 0$

Considering negative biases, delegation is detrimental to the principal if preferences are reversed, i.e. $b_m < -1$. In this case, he retains authority and takes an uninformed decision as communicating with the agent does not convey any information. On the other hand, if $b_m \in (-1, 0)$ and communication is informative, it is preferred to delegation if $\|b\|$ is large enough. In particular, it turns out that whenever an equilibrium with $N > 2$ partition intervals is possible, $\|b\|$ is so small that delegation performs better than communication. Hence, communication can only outperform delegation if it is not too informative.

Proposition 5. *If $b_m < 0$, communication dominates delegation if and only if*

$$|b_a| > \frac{L}{\sqrt{3}}\sqrt{1 - b_m^2}$$

or

$$|b_a| < \frac{L}{4\sqrt{3}}|2b_m + 1|\sqrt{\frac{1 - 4b_m^2}{b_m(b_m + 1)}}.$$

Furthermore, if communication dominates delegation, then $N(b) \leq 2$.

It is worth noting that we do not need any monotonicity result nor an equilibrium selection device in order to characterize the area where communication dominates delegation. Figure 2.4 summarizes the analysis of Section 2.1 for positive and negative b_m. For biases outside the thick-bordered area, communication performs better than delegation from the principal's point of view and colors indicate how informative a communication equilibrium can maximally be.

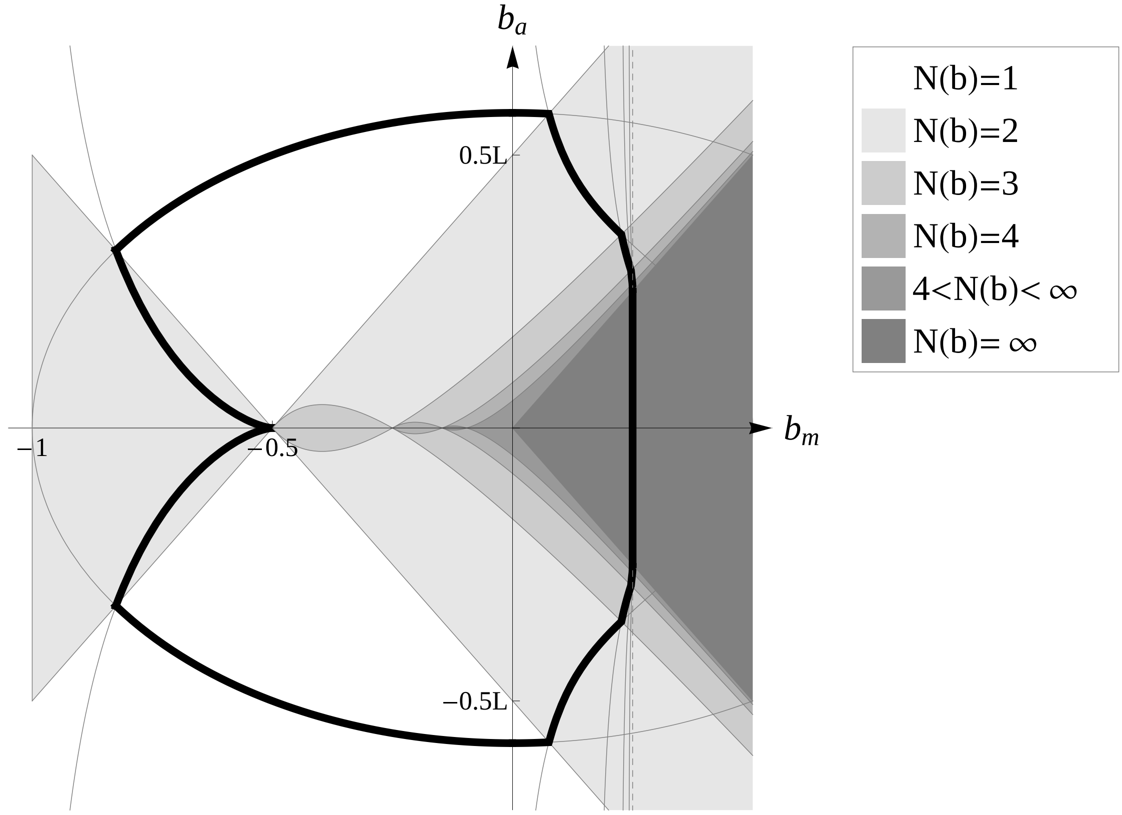

Figure 2.4: Comparison of delegation and communication and the maximum number of partition intervals in equilibrium for any given $b \in \mathbb{R}^2$. Outside the area limited by the thick curve, communication dominates delegation.

2.2 General Distributions

In this section, we test the robustness of our findings if we consider general distributions. Let thus θ be distributed according to $F(\theta)$, with $f(\theta)$ denoting its density. We assume that f is continuous and twice differentiable on its support $supp(f) = \overline{(\theta_L, \theta_H)}$, with $-\infty \leq \theta_L < \theta_H \leq +\infty$, thereby allowing for unbounded supports. Throughout this section, we assume that $b_m > -1$, such that $b(\theta) = b_a + (b_m + 1)\theta$ is strictly increasing in θ.

In Section 2.2.1, we provide a necessary and sufficient condition for communication to be informative for arbitrary distributions of θ. Moreover, we find a necessary condition for the existence of communication equilibria with more than two partition intervals and derive a condition for communication to be infinitely informative. Hence, our analysis is complementary to [CS], who show that an equilibrium partition cannot contain an accumulation point for any distribution of θ if $b_m = 0$.

Considering symmetric distributions with compact support, Section 2.2.2 compares communication to delegation and analyzes the optimal allocation of authority from the principal's point of view. In line with [D], we observe that communication performs badly compared to delegation if $\|b\|$ is small. If the preference divergence is large enough, however, we show that the principal prefers communication to delegation, independent of the distribution of θ. This differs from the result in [D], which implies that for distributions with large variance, communication is uninformative if it dominates delegation from the principal's point of view.

2.2.1 General Properties of Communication Equilibria

Let $\theta_L = a_0 < a_1 < \ldots < a_{N-1} < a_N = \theta_H$ denote a partition of $[\theta_L, \theta_H]$.[7] Then, given that the agent truthfully signals the partition interval containing θ, the principal implements $x_i = E[\theta|\theta \in (a_{i-1}, a_i)]$ if $m \in (a_{i-1}, a_i)$. As $x_A(\theta) = b(\theta)$ is strictly increasing in θ, the agent has an incentive to truthfully report $[a_{i-1}, a_i] \ni \theta$ if and only if the partition satisfies the arbitrage condition

$$x_i + x_{i+1} = 2b(a_i). \tag{A}$$

for all $i \in \{1, \ldots, N-1\}$.[8] If we assume $N = 2$, and define $h(a) := E[\theta|\theta < a] + E[\theta|\theta > a]$ for $a \in supp(f)$, the above condition is equivalent to

$$h(a_1) = 2b(a_1).$$

[7]If $\theta_L = -\infty$ and $\theta_H = +\infty$, we have a partition of the set of real numbers, where a_0 and a_N are usually omitted.

[8]This holds independent of the distribution of θ.

This yields the following lemma, providing a necessary and sufficient condition for communication to be informative for arbitrary distributions of θ.

Lemma 3. *Let supp(f)* $= \overline{(\theta_L, \theta_H)}$ *with* $-\infty \leq \theta_L < \theta_H \leq +\infty$ *and* $b_m > -1$. *For* $a \in$ *supp*(f), *we define*

$$h(a) := E[\theta|\theta < a] + E[\theta|\theta > a].$$

Then communication is informative ($N(b) \geq 2$) *if and only if one of the following two cases holds:*

1. $\lim_{a\to\theta_L} h(a) > \lim_{a\to\theta_L} 2b(a)$ *and* $\lim_{a\to\theta_H} h(a) < \lim_{a\to\theta_H} 2b(a)$, *or*
2. $\lim_{a\to\theta_L} h(a) < \lim_{a\to\theta_L} 2b(a)$ *and* $\lim_{a\to\theta_H} h(a) > \lim_{a\to\theta_H} 2b(a)$.

Indeed, if we take a step back to the example where f is the uniform distribution on $[-L, L]$, the above two cases are equivalent to

1. $b_m > -\frac{1}{2}$ and $|b_a| < \frac{2b_m+1}{2}L$, or
2. $b_m < -\frac{1}{2}$ and $|b_a| < -\frac{2b_m+1}{2}L$, respectively,

yielding exactly the condition for communication to be informative if θ is uniformly distributed, as shown in Section 2.1.1. Furthermore, this same condition continues to characterize informative communication if we generalize f to a symmetric and compactly supported distribution. Lemma 3 illustrates this finding, which is akin to a result by [D].

Corollary 5. *If* $f(\theta)$ *is symmetric and supported on* $[-L, L]$, *then communication is informative* ($N(b) \geq 2$) *if and only if*

$$|b_a| < \frac{|2b_m + 1|}{2}L.$$

Finally, for the uniform distribution we have seen that $b_m > -\frac{1}{2}$ is a necessary condition for $N(b) > 2$. In other words, a communication equilibrium with more than two partition intervals can only exist in case 1 of Proposition 3 if θ is uniformly distributed. An analogous condition for general distributions is derived in Lemma 4.

Lemma 4. *In an equilibrium with $N > 2$ partition intervals, it has to hold that*

1. $\lim_{a\to\theta_L} h(a) > \lim_{a\to\theta_L} 2b(a)$ *and* $\lim_{a\to\theta_H} h(a) < \lim_{a\to\theta_H} 2b(a)$.

According to [CS], if $x_P(\theta) \neq x_A(\theta)$ for all $\theta \in supp(f)$, the sequence $\{a_i\}_{i=0}^{\infty}$ representing an equilibrium partition cannot have an accumulation point (though it is infinite if $supp(f) = \mathbb{R}$). However, if $x_P(\theta_0) = x_A(\theta_0)$ for some $\theta_0 \in supp(f)$, it is possible that θ_0 is an accumulation point for $\{a_i\}_{i=0}^{\infty}$, in which case communication becomes infinitely informative near θ_0. Proposition 6 provides a necessary and sufficient condition for the existence of a communication equilibrium with infinitely many partition intervals.

Proposition 6. *Let $a_1 < \ldots < a_i < \theta_0$ be a sequence in supp(f) satisfying the arbitrage condition* (A). *Then there exists exactly one a_{i+1} satisfying* (A) *with $a_i < a_{i+1} < \theta_0$ if and only if*

$$2b(a_i) - E[\theta|\theta \in (a_i, \theta_0)] < x_i^a < 2b(a_i) - a_i,$$

where $x_i^a := E[\theta|\theta \in (a_{i-1}, a_i)]$.
Analogously, let $A_1 > \ldots > A_i > \theta_0$ be a sequence in supp(f) satisfying the arbitrage condition (A). *Then there exists exactly one A_{i+1} satisfying* (A) *with $A_i > A_{i+1} > \theta_0$ if and only if*

$$2b(A_i) - A_i < x_i^A < 2b(A_i) - E[\theta|\theta \in (\theta_0, A_i)],$$

where $x_i^A := E[\theta|\theta \in (A_i, A_{i-1})]$.

Proposition 6 imposes a geometrical constraint on the choice of the sequence $\{a_i\}_{i=1}^{\infty}$, depending on the distribution of θ. Figure 2.5 illustrates this constraint.

Note that, in the uniform case, the conditions for an ascending sequence $\{a_i\}_{i=0}^{\infty}$ in Proposition 6 are equivalent to

$$\begin{aligned} 4b_a &< a_{i-1} - (4b_m + 2)a_i + \theta_0 \quad \text{and} \\ 4b_a &> a_{i-1} - (4b_m + 1)a_i. \end{aligned}$$

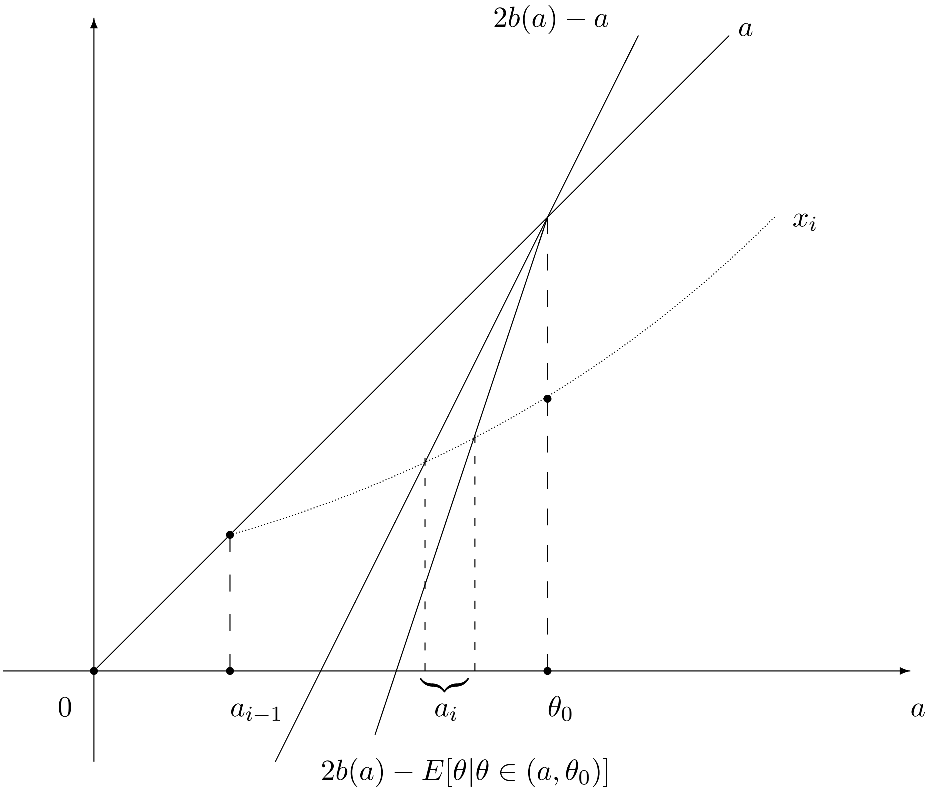

Figure 2.5: Illustration of Proposition 6. Depending on $a_{i-1} < \theta_0$, there exists a $a_{i+1} < \theta_0$ with $a_{i+1} > a_i > a_{i-1}$ if and only if a_i lies in the bracketed line segment.

Applying the recursive formula for a_i based on a_1, and assuming $b_m > 0$, we retrieve the known condition for communication to be infinitely informative if θ is uniformly distributed.

Corollary 6. *If θ is uniformly distributed on $[-L, L]$ and $b_m > 0$, applying Proposition 6 yields that $\{a_i\}_{i=0}^{\infty}$ has an accumulation point if and only if*

$$|b_a| < b_m L.$$

2.2.2 Communication Versus Delegation

In this section, we compare the principal's expected utility from delegation and communication for an arbitrary symmetric and compactly supported distribution of θ. Though we have seen that the nature of communication equilibria changes substantially by allowing for a state of agreement between the principal and the agent, this is not necessarily the case if we consider its performance compared to delegation. Indeed, for "small" biases, we obtain the same result as in [D], namely that the principal prefers delegation to communication. For "large" biases, however, we find that the threshold above which communication dominates delegation is independent of the distribution of θ. This indicates that, in contrast to [D], it is possible to find a bias for any distribution of θ, such that communication is informative *and* dominates delegation from the principal's point of view.

Small Biases. Before analyzing general distributions, we first have a look at the uniform one. Lemma 5 shows that, in line with the result in [D], the expected noisiness of the project choice is an infinite times larger with communication than with delegation if the bias is infinitesimally small and the state of the world is uniformly distributed – even in the most informative equilibrium.

Lemma 5. *If θ is uniformly distributed and $N(b) = \infty$, the principal's expected loss from retaining authority and communicating, as given by the residual variance, is an infinite times larger than his expected loss from delegating the project choice:*

$$\lim_{(b_m,b_a)\to(0,0)} \frac{\sigma_\infty^2}{b_a^2 + b_m^2\sigma^2} = \lim_{(b_m,b_a)\to(0,0)} \frac{L^2b_m^2 + 3b_a^2}{(4b_m+3)b_m(b_a^2 + L^2b_m^2)} = +\infty.$$

We will show in Proposition 7 that the above result continues to hold for general symmetric distributions with compact support.[9] In other words, we find that the analogous result for $b_m = 0$, as proven by [D], can be generalized to linear biases $b = (b_m, b_a)$. In the following, we denote by $x^*(\theta)$ the project implemented by the principal in the most informative

[9] In fact, we do not require symmetry, but just $E[\theta] = 0$.

equilibrium of the communication game, given that the state θ is realized and the agent plays the equilibrium communication strategy.

Proposition 7. *Consider the most informative equilibrium given $b = (b_m, b_a)$. For any symmetric distribution $f(\theta)$ with compact support, in the limit as $b = (b_m, b_a)$ tends to zero, the principal's expected loss from retaining authority and communicating is an infinite times larger than his expected loss from delegating the project choice:*

$$\lim_{(b_m,b_a)\to(0,0)} \frac{E[(x^*(\theta)-\theta)^2]}{b_a^2 + b_m^2\sigma^2} = +\infty.$$

Large Biases. On the other hand, if we consider large biases, communication is at least as profitable for the principal as taking an uninformed decision. In the latter case, he chooses $x = E[\theta]$ and his expected loss is given by $\sigma^2 = \sigma_1^2$. Comparing this to the expected loss incurred by delegation, $b_a^2 + b_m^2\sigma^2$, we can provide a sufficient condition such that communication dominates delegation.

Lemma 6. *For any distribution $F(\theta)$, communication dominates delegation at least if $|b_m| > 1$, independent of σ.*

In the limiting case $b_m = 0$, that is, the principal and the agent disagree about the project choice in all states of the world, [D] shows that the principal prefers communication to delegation at least if $b_a > \sigma$. At the same time, communication is informative only if $b_a < \frac{L}{2}$. Consequently, if the distribution of θ exhibits a large variance, it might be the case that for all biases allowing for informative communication, the principal prefers to delegate the project choice.

In contrast, we show that the threshold above which communication dominates delegation does not depend on the distribution of θ. Hence, for any distribution $F(\theta)$, irrespective of its variance, there is a bias $b = (b_m, b_a)$ such that communication is informative *and* dominates delegation.

2.3 Conclusion

In this paper, we have generalized the models of [CS], [D], and [ADM] by considering a general linear preference divergence between principal and agent. If the state of the world is uniformly distributed, we have shown that communication can be infinitely informative in equilibrium if and only if the state of agreement between the principal and the agent realizes with positive probability, which includes the results by [CS] and [ADM] as limiting cases. Based on this analysis, in line with [D], we have found that the principal prefers to delegate formal decision rights to the agent instead of retaining authority and communicating if the preference divergence is “small”. However, in our model with uniformly distributed states, communication can be informative *and* dominate delegation from the principal’s point of view, in contrast to [D]. To the best of our knowledge, this is the first paper characterizing communication equilibria for a general linear preference divergence between the principal and the agent, and analyzing the optimal allocation of authority in this generality.

The present analysis sheds light on the optimal structure of organizations, indicating whether decision making should be centralized or delegated, depending on the preference divergence between a principal and a subordinate agent. If the agent exhibits a purely systematic bias, such as empire building, the principal delegates decision making if the agent’s informational advantage is large relative to his bias. Following [D], we conclude that for a given bias centralization is only optimal if the corporate management has some important information, as reflected by firms focusing on their core activities while outsourcing others. In the same vein, operating in a rapidly changing environment favors decentralization, since top management information rapidly becomes obsolete.

Our results put this interpretation into a broader perspective by allowing the agent to react stronger or weaker to changes in the environment than the principal. In fact, as long as the agent’s preference is not too extreme nor too conservative with regard to the principal’s, it is true that decentralization is optimal in rather uncertain environments, as indicated by a large L. However, if the agent excessively under- or overreacts to changes in the state of the world, our results suggest that it is optimal to centralize decision making, no matter how unstable the market situation is.

Consider for example an organization entering an emerging market, where large external shocks influencing the local economy are relatively likely. On the one hand, if experience in the local market is the key for success, the firm might want to deploy local people with a profound market knowledge, who react adequately to changes in the environment. However, these agents are likely to exhibit a systematic bias, since they have no experience with the foreign organization, thus might prefer to build up their own empire. In case of centralization, communication is therefore rather imprecise if the environmental risks are large. Hence, in line with [D], headquarters is likely to delegate decision rights to the local managers in order to maximally profit from their local knowledge (while incurring a constant and predictable loss due to the systematic bias). This situation is reminiscent of organizations entering a new market via contracting with local firms, for example in procurement or sales.

On the other hand, imagine a business where the application of core competences is crucial for success in a new market, as for example for a consultancy with specific experience in supply chain management across several industries. In this case, the organization might prefer to install own employees in the new market who are familiar with the firm and its specific know-how. While these agents are unlikely to exhibit a large systematic bias, they might however overstate market distortions since they lack experience in the new environment. In that case, delegating project decisions is rather risky as overreactions to external shocks might cause huge organizational losses, which favors centralization. In line with this reasoning, consultancies tend to enter a new market by installing a subsidiary which is hierarchically subordinated to an organizational entity operating in the core market.

Our analysis indicates that the optimal design of organizational structures should take into account whether agents tend to react extremely or conservatively to a changing environment. In that sense, companies should be sensitive to differences in manager types within the organization when defining hierarchies. In an organization with multiple functional divisions, for instance, natural differences in the conservativeness of the respective business might thus induce the organization to impose different organizational structures, ceteris paribus. However, to the extent that managerial conservatism is rather individual, our results also provide a rationale for contracts to be incomplete. Whether formal decision rights should be granted or not can only be decided after learning the manager's type, hence it might be rational to omit such clauses from the employment contract.

Finally, the present analysis is based on two simplifying assumptions. First, there are no technical or intellectual constraints to communication, potentially inducing some cost of information transmission. Second, we assume that formal decision rights can be allocated to either the principal or the agent, but the allocation of authority cannot be made contingent on the message received by the agent. Allowing for costly communication and optimal delegation mechanisms will work in favor of delegation relative to communication. Hence, the question to what extent centralization of decision rights is an option for the principal in more general situations would be an interesting avenue for future research.

A2 Mathematical Appendix

Proof of Proposition 1. Let $-L = a_0 < a_1 < \ldots < a_{N-1} < a_N = L$ be a partition of $\Theta = [-L, L]$. Given that the agent truthfully signals the interval $I_i := [a_{i-1}, a_i]$ which includes θ, the principal implements $E[\theta|\theta \in I_i] = \frac{a_{i-1}+a_i}{2} =: x_i$ if $m \in I_i$. The agent, in turn, reveals the interval containing θ if $(x_i - b(\theta))^2 \leq (x_j - b(\theta))^2$ for all $j \neq i$, or

$$\begin{aligned} x_i + x_j &\leq 2b(\theta) \quad \forall j < i \quad \text{and} \\ x_i + x_j &\geq 2b(\theta) \quad \forall j > i. \end{aligned}$$

Given that x_i increases in i and $b(\theta)$ increases in θ, the above condition is fulfilled if and only if $b(a_i) = \frac{x_i + x_{i+1}}{2}$ or

$$a_{i+1} = 2a_i - a_{i-1} + 4(b_a + b_m a_i)$$

for all $i \in \{1 \ldots N-1\}$. Let $v := 4b_m + 2$ and $U_i(v, 1)$ denote the Lucas sequence of the first kind, that is, $U_0(v, 1) = 0$, $U_1(v, 1) = 1$, and $U_i(v, 1) = vU_{i-1}(v, 1) - U_{i-2}(v, 1)$. Then an induction shows that a_i is given by

$$a_i = U_i(v, 1)a_1 + U_{i-1}(v, 1)L + 4b_a \sum_{j=1}^{i-1} U_j(v, 1).$$

Note that for $b_a = 0$, this yields $a_i = U_i(v, 1)a_1 + U_{i-1}(v, 1)L$. Furthermore, if $b_m = 0$, we have $v = 2$ and $U_i(2, 1) = i$. Thus, for an additive bias we have $a_i = ia_1 + (i-1)L + 2b_a i(i-1)$, which coincides with the result by [CS] for uniform distributions on $[-L, L]$. For the Lucas sequence, it is known that $U_i(v, 1) = \frac{c^i - d^i}{c-d}$, where

$$\begin{aligned} c &:= \frac{1}{2}[v + \sqrt{v^2 - 4}] = 1 + 2b_m + \sqrt{(1+2b_m)^2 - 1} \quad \text{and} \\ d &:= \frac{1}{2}[v - \sqrt{v^2 - 4}] = 1 - 2b_m + \sqrt{(1+2b_m)^2 - 1}. \end{aligned}$$

Note that $cd = 1$, $c + d = v$ and $c - d = \sqrt{v^2 - 4}$. Since $b_m > 0$, the square root is a real number.

Using the terminal condition $a_N = L$, we can solve for a_1 and get

$$\begin{aligned} a_1 &= \frac{1 - U_{N-1}(v,1)}{U_N(v,1)} L - \frac{4b_a}{U_N(v,1)} \sum_{j=1}^{N-1} U_j(v,1) \\ &= \frac{1}{c^N - d^N} \left[\left[c(1+d^N) - d(c^N+1) \right] L - 4b_a \left[\frac{c^N - 1}{c-1} - \frac{1-d^N}{1-d} \right] \right]. \end{aligned}$$

In turn, we have

$$\begin{aligned} a_i &= \frac{L}{c^N - d^N} \left[c^i(1+d^N) - d^i(1+c^N) \right] \\ &+ \frac{4b_a}{c^N - d^N} \frac{(1-d^i)(1-d^{N-i}) - (c^i - 1)(c^{N-i} - 1)}{(c-1)(1-d)}. \end{aligned}$$

Note that, if $b_a = 0$, we have $a_i = \frac{1}{c^N - d^N} \left[c^i(1+d^N) - d^i(1+c^N) \right] L$, which coincides with [ADM]. In order to ensure that the partition exists in equilibrium, we have to check whether $\{a_i\}_{i=0 \dots N}$ is an ascending sequence. For $i \in \{0 \dots N-1\}$, using the relation $cd = 1$, we have

$$a_{i+1} - a_i = \frac{(c-1)(c^i + c^{N-i-1})}{c^N - 1} L + \frac{4c(c^i - c^{N-i-1})}{(c^N+1)(c-1)} b_a.$$

Note that the factor of L is always positive, whereas this is not necessarily the case for the factor of b_a. We see that $c^i - c^{N-i-1}$ is positive if $i > \frac{N-1}{2}$ and negative if $i < \frac{N-1}{2}$. If N is odd and $i = \frac{N-1}{2}$, the length of the interval does not depend on b_a. In order to ensure that all intervals have a positive measure, we have to distinguish two cases:

- **$b_a > 0$:** The second summand $\frac{4c(c^i - c^{N-i-1})}{(c^N+1)(c-1)} b_a$ is negative if and only if $i < \frac{N}{2}$, hence $a_{i+1} - a_i > 0$ for all $i \in \{0 \dots N-1\}$ if and only if

$$b_a < \frac{(c-1)^2(c^N+1)}{4c(c^N-1)} \frac{c^{N-i-1} + c^i}{c^{N-i-1} - c^i} L$$

for all $i < \frac{N}{2}$. In that case, the right hand side is positive and $\frac{c^{N-i-1} + c^i}{c^{N-i-1} - c^i}$ increases in i, hence the inequality has to be satisfied for $i = 0$, or $b_a < \frac{(c-1)^2(c^N+1)(c^{N-1}+1)}{4c(c^N-1)(c^{N-1}-1)} L$.

- $b_a < 0$: The second summand $\frac{4c(c^i - c^{N-i-1})}{(c^N+1)(c-1)} b_a$ is negative if and only if $i > \frac{N}{2}$, hence we have to ensure that

$$b_a > \frac{(c-1)^2(c^N+1)}{4c(c^N-1)} \frac{c^{N-i-1}+c^i}{c^{N-i-1}-c^i} L$$

for all $i > \frac{N}{2}$. In that case, the right hand side is negative and $\frac{c^{N-i-1}+c^i}{c^{N-i-1}-c^i}$ increases in i, hence the inequality has to be satisfied for $i = i_{max} = N-1$, or $b_a > -\frac{(c-1)^2(c^N+1)(c^{N-1}+1)}{4c(c^N-1)(c^{N-1}-1)} L$.

Summarizing the two cases, a partition with N intervals is possible in equilibrium if

$$|b_a| < \frac{(c^N+1)(c^{N-1}+1)}{(c^N-1)(c^{N-1}-1)} b_m L,$$

using $(c-1)^2 = 4b_m c$. Furthermore, the sequence $y_N := \frac{(c^N+1)(c^{N-1}+1)}{(c^N-1)(c^{N-1}-1)}$ decreases in N, thus the larger the absolute value of the additive part of the bias, $|b_a|$, the smaller is the possible number of intervals. It turns out that $\lim_{N\to\infty} y_N = 1$, thus

$$\lim_{N\to\infty} \frac{(c^N+1)(c^{N-1}+1)}{(c^N-1)(c^{N-1}-1)} b_m L = b_m L.$$

That means that any number of partition intervals is possible if and only if $|b_a| < b_m L$. If $|b_a| = b_m L$, we have

$$\begin{aligned} a_{i+1} - a_i &= \frac{2(c-1)(c^{2N-i-1}+c^i)}{c^{2N}-1} \quad \text{if } b_a < 0 \text{ and} \\ a_{i+1} - a_i &= \frac{2(c-1)(c^{N+i}+c^{N-i-1})}{c^{2N}-1} \quad \text{if } b_a > 0, \end{aligned}$$

which is positive for all $i \in \{0 \dots N-1\}$. Hence, if the preferred projects of the principal and the agent intersect within the range of consideration, $[-L, L]$, there is a communication equilibrium for any $N \in \mathbb{N}$. ■

Geometrical notes. Taking a closer look at whether the length of partition intervals increases or decreases, we have for $i \in \{1 \ldots N-1\}$

$$l(I_{i+1}) - l(I_i) = a_{i+1} - 2a_i + a_{i-1} = \frac{(c-1)^2(c^{i-1} - c^{N-i-1})}{(c^N - 1)} L + \frac{4c(c^{i-1} + c^{N-i-1})}{(c^N + 1)} b_a.$$

If $b_a > 0$, the length of intervals is hence increasing for $i \geq \frac{N}{2}$. It is decreasing if and only if

$$c^{2i} < c^N \frac{(c-1)^2(c^N+1)L - 4c(c^N-1)b_a}{(c-1)^2(c^N+1)L + 4c(c^N-1)b_a} < c^N.$$

Hence, interval lengths decrease up to some interval and then increase again. The smallest interval is some I_i with $i \leq \frac{N}{2}$. It is precisely when $b_a > b_m L$ that the smallest interval is the first, which means that interval lengths increase from left to right and only finitely many intervals are possible. In general, the smallest interval is further to the left the larger b_a is. Finally, $c^{2i} < c^N \frac{(c-1)^2(c^N+1)L-4c(c^N-1)b_a}{(c-1)^2(c^N+1)L+4c(c^N-1)b_a}$ if and only if $a_i < -\frac{b_a}{b_m}$, which means that the agreement state lies in the smallest interval. The analysis for $b_a < 0$ is vice-versa.

If $b_a = 0$, we have $l(I_i) = l(I_{N-i+1})$, that is, interval lengths are symmetric. Further, $\theta = 0$ is either the midpoint of the smallest interval (if N odd) or it is the border of the two smallest intervals (if N even, see [ADM]). In contrast, if $b_a \neq 0$, there is generically no symmetry, there are no two intervals with the same lengths, and the agreement state is not the border of an interval.

Proof of Corollary 1. We determine the threshold for $|b_a|$ in Proposition 1 if b_m approaches zero, or, equivalently, if $c \searrow 1$. Applying L'Hôpital's rule, we find

$$\lim_{c \to 1} \frac{(c-1)^2}{(c^N-1)(c^{N-1}-1)} = \lim_{c \to 1} \frac{2(c-1)}{(2N-1)c^{2N-2} - Nc^{N-1} - (N-1)c^{N-2}}$$
$$= \lim_{c \to 1} \frac{2}{(2N-1)(2N-2)c^{2N-3} - N(N-1)c^{N-2} - (N-1)(N-2)c^{N-3}} = \frac{1}{N(N-1)}.$$

Hence,

$$\lim_{c \to 1} \frac{(c^N+1)(c^{N-1}+1)}{(c^N-1)(c^{N-1}-1)} b_m L = \lim_{c \to 1} \frac{(c^N+1)(c^{N-1}+1)(c-1)^2}{4c(c^N-1)(c^{N-1}-1)} L = \frac{L}{N(N-1)}.$$

■

Proof of Proposition 2. Similar to the Proof of Proposition 2 in [ADM], if there was a profitable deviation for the agent in an equilibrium with infinitely many intervals, this deviation would also be profitable in an equilibrium with N intervals, for some $N < \infty$, contradicting the assumption that communication rules, decision functions and belief functions constitute an equilibrium for all finite N. ∎

Proof of Proposition 3. Let $b_m \in (-1, 0)$. Since $x_A(\theta) = b(\theta)$ is increasing in θ, we obtain the same constraints on the equilibrium partition as in the proof of Proposition 1 and the formulas for $a_i - a_{i-1}$ continue to hold. However, monotonicity is no longer guaranteed since $c \in \mathbb{C} \setminus \mathbb{R}$. Considering a partition with two intervals, $N(b) \geq 2$ if and only if $|a_1| < L$, where a_1 is given by $a_1 = -\frac{2b_a}{2b_m+1}$, using the arbitrage condition (A). Hence, $N(b) \geq 2$ if and only if $|b_a| < \frac{L}{2}|2b_m + 1|$.

Similarly, in an equilibrium with $N \geq 3$ partition intervals, (A) is equivalent to

$$
\begin{aligned}
a_2 &= 2a_1 + L + 4(b_a + b_m a_1) \quad \text{and} \\
L &= 2a_2 - a_1 + 4(b_a + b_m a_2),
\end{aligned}
$$

which is equivalent to

$$
\begin{aligned}
a_1 &= -\frac{L}{4b_m + 3} - \frac{4b_a}{4b_m + 1} \\
a_2 &= \frac{L}{4b_m + 3} - \frac{4b_a}{4b_m + 1}.
\end{aligned}
$$

Hence, $a_1 < a_2$ if and only if $b_m > -\frac{3}{4}$, which we assume in the following. Furthermore, $a_1 > -L$ if and only if $\frac{4b_m+2}{4b_m+3}L > \frac{4}{4b_m+1}b_a$ and $a_2 < L$ if and only if $\frac{4b_m+2}{4b_m+3}L > -\frac{4}{4b_m+1}b_a$. In particular, it has to hold that $\frac{4b_m+2}{4b_m+3} > 0$. Since $b_m > -\frac{3}{4}$, this is equivalent to $b_m > -\frac{1}{2}$.

Next, we show that $N(b) < \infty$ if $b_m \in (-1, 0)$. In that case, we find that $c \in \mathbb{C} \setminus \mathbb{R}$ is a root of unity. Indeed, if $b_m \in (-1, 0)$ we have

$$
c = 1 + 2b_m + i\sqrt{1 - (1 + 2b_m)^2},
$$

where $i = \sqrt{-1}$ denotes the imaginary unit in the upper half plane, and $|c| = (1+2b_m)^2 + 1 - (1+2b_m)^2 = 1$. Hence, we can write

$$c = e^{i\alpha} = \cos\alpha + i\sin\alpha$$

for some $\alpha \in (0, \pi)$. Now we define

$$r_k := \frac{(c^k - 1)^2}{c^k} = 2\left(\cos(k\alpha) - 1\right) \in [-4, 0],$$

and see that r_k is a real number for any $k \in \{0, \dots, N\}$. Furthermore, $r_0 = 0$ and $r_1 = 4b_m$. Thus, for all $k \in \{0, \dots, N\}$, we have

$$a_k = \frac{r_k - r_{N-k}}{r_N} L - \frac{b_a}{b_m} \frac{r_N - r_k - r_{N-k}}{r_N + 4},$$

and accordingly

$$a_k - a_{k-1} = \frac{r_k - r_{k-1} + r_{N+1-k} - r_{N-k}}{r_N} L - \frac{b_a}{b_m} \frac{-r_k + r_{k-1} + r_{N+1-k} - r_{N-k}}{r_N + 4}.$$

Hence, if N is odd, we have

$$a_{\frac{N+1}{2}} - a_{\frac{N-1}{2}} = 2\frac{r_{\frac{N+1}{2}} - r_{\frac{N-1}{2}}}{r_N} L = 2\frac{\cos[\frac{N+1}{2}\alpha] - \cos[\frac{N-1}{2}\alpha]}{\cos[N\alpha] - 1} L = \frac{\sin[\frac{N}{2}\alpha]\sin\alpha}{1 - \cos[\frac{N}{2}\alpha]} L.$$

Let $\alpha \in (0, \pi)$ such that if $\alpha = q\pi$, we have $q \notin \mathbb{Q}$, which means that there is no $s \in \mathbb{N}$ with $\cos(s\alpha) = 1$. Then $\frac{\alpha}{2} \in (0, \frac{\pi}{2})$ and we can find N_0 odd such that $\frac{\alpha}{2} N_0 \in (\pi, 2\pi)$.[10] For this N_0, we have $\sin[\frac{N_0}{2}\alpha] < 0$. Since $1 - \cos[\frac{N}{2}\alpha] > 0$ for any N and $\sin\alpha > 0$ for $\alpha \in (0, \pi)$, we conclude that $a_k - a_{k-1} < 0$ for $k = \frac{N_0+1}{2}$. Hence, for any given $\alpha \in (0, \pi)$ which is no rational multiple of π, there is a N_0 such that there is no equilibrium with N_0 partition intervals.

[10] More precisely, $\frac{\alpha}{2} N_0 \in (\pi, 2\pi)$ is equivalent to $N_0 \in (\frac{2\pi}{\alpha}, \frac{4\pi}{\alpha})$. Since $\frac{2\pi}{\alpha} > 2$, this interval includes two consecutive natural numbers, thus also an odd one.

If $\alpha \in (0, \pi)$ is such that $\cos(s\alpha) = 1$ for some $s \in \mathbb{N}$, we have $r_s = 0$ and the division by r_s is invalid. For $N = s$, the condition for $a_k - a_{k-1}$ is redundant and there is a partition equilibrium with s intervals for any b_a. However, if we consider $s + 1$, the first partition interval has to satisfy

$$\begin{aligned} a_1 - a_0 &= \frac{r_1 - r_0 + r_{s+1} - r_s}{r_{s+1}} L - \frac{b_a}{b_m} \frac{-r_1 + r_0 + r_{s+1} - r_s}{r_{s+1} + 4} \\ &= \frac{2\cos\alpha - 2}{\cos\alpha - 1} L + 2b_a \cdot 0 = 2L. \end{aligned}$$

Hence, an equilibrium with $s + 1 \geq 2$ partition intervals is not possible and $N(b) < \infty$ if $b_m \in (-1, 0)$.

If $b_m < -1$, the agent's preference is reversed, i.e. $b(\theta)$ decreases in θ. Hence, the agent has an incentive to truthfully reveal the state of the world (see proof of Proposition 1) if and only if

$$\begin{aligned} x_i + x_{i-1} &\leq 2b(a_i) \quad \forall i \in \{2 \ldots N\} \quad \text{and} \\ x_i + x_{i+1} &\geq 2b(a_{i-1}) \quad \forall i \in \{1 \ldots N-1\}. \end{aligned}$$

Thus, for all $i \in \{0 \ldots N-2\}$ we conclude $b(a_i) \leq b(a_{i+2})$, which contradicts the assumption $b_m < -1$ and a_i strictly increasing. Consequently, communication is not informative for $b_m < -1$.

For $b_m = -1$, we similarly conclude $x_i + x_{i+1} = 2b_a$ for all $i \in \{1 \ldots N-1\}$, which yields $x_1 = x_2 = \ldots = x_N$ if $N > 2$. For $N = 2$, we have $x_1 + x_2 = 2b_a$, which is equivalent to $a_1 = 2b_a$. Hence, an equilibrium with two intervals is possible if $|b_a| < \frac{L}{2} = \frac{L}{2}|2b_m + 1|$. ■

Proof of Remark 1. From Lemma 1, we conclude

$$\sigma_2^2 = \frac{L^2}{12} + \frac{4c^2}{(c^2+1)^2} b_a^2 = \frac{L^2}{12} + \frac{b_a^2}{(2b_m+1)^2}.$$

Since $|b_a| < \frac{|2b_m+1|}{2} L$ if $N(b) \geq 2$, we have $\sigma_2^2 < \frac{L^2}{12} + \frac{L^2}{4} = \frac{L^2}{3} = \sigma^2$. ■

Proof of Lemma 1. The residual variance of θ given a message $m \in \Theta$ in an equilibrium with N partition intervals, is given by

$$\begin{aligned}
\sigma_N^2 &= \frac{1}{2L}\sum_{i=1}^{N}\int_{a_{i-1}}^{a_i}\left[\theta - \frac{a_{i-1}+a_i}{2}\right]^2 d\theta \\
&= \frac{1}{24L}\sum_{i=1}^{N}(a_i - a_{i-1})^3 \\
&= \frac{1}{24}\frac{(c-1)^3}{(c^N-1)^3}L^2\sum_{i=1}^{N}\left(c^{3i-3} + c^{3N-3i} + 3c^{N+i-2} + 3c^{2N-i-1}\right) \\
&+ \frac{8}{3}\frac{c^3}{(c^N+1)^3(c-1)^3}\frac{b_a^3}{L}\sum_{i=1}^{N}\left(c^{3i-3} - c^{3N-3i} - 3c^{N+i-2} + 3c^{2N-i-1}\right) \\
&+ \frac{1}{2}\frac{c(c-1)}{(c^N-1)^2(c^N+1)}Lb_a\sum_{i=1}^{N}\left(c^{3i-3} - c^{3N-3i} + c^{N+i-2} - c^{2N-i-1}\right) \\
&+ 2\frac{c^2}{(c^N+1)^2(c-1)(c^N-1)}b_a^2\sum_{i=1}^{N}\left(c^{3i-3} + c^{3N-3i} - c^{N+i-2} - c^{2N-i-1}\right).
\end{aligned}$$

As $\sum_{i=1}^{N} c^{3i-3} = \frac{c^{3N}-1}{c^3-1} = \sum_{i=1}^{N} c^{3N-3i}$ and $\sum_{i=1}^{N} c^{N+i-2} = c^{N-1}\frac{c^N-1}{c-1} = \sum_{i=1}^{N} c^{2N-i-1}$, the above expression is equivalent to

$$\begin{aligned}
\sigma_N^2 &= \frac{L^2}{12}\frac{(c-1)^2}{(c^N-1)^2}\left(\frac{c^{2N}+c^N+1}{c^2+c+1} + 3c^{N-1}\right) \\
&+ \frac{4b_a^2c^2}{(c^N+1)^2(c-1)^2}\left(\frac{c^{2N}+c^N+1}{c^2+c+1} - c^{N-1}\right).
\end{aligned}$$

Using the definition of r_k as in the proof of Proposition 3, we find that

$$\begin{aligned}
\sigma_N^2 &= \frac{L^2}{12}\frac{(c-1)^2}{c}\frac{c^N}{(c^N-1)^2}\left(\frac{c}{c^2+c+1}\left[\frac{(c^N+1)^2}{c^N} - 1\right] + 3\right) \\
&+ 4b_a^2\frac{c}{(c-1)^2}\frac{c^N}{(c^N+1)^2}\left(\frac{c}{c^2+c+1}\left[\frac{(c^N+1)^2}{c^N} - 1\right] - 1\right) \\
&= \frac{L^2}{3}\frac{b_m}{r_N}\left(\frac{r_N+3}{4b_m+3} + 3\right) + b_a^2\frac{1}{b_m}\frac{1}{r_N+4}\left(\frac{r_N+3}{4b_m+3} - 1\right) \in \mathbb{R}.
\end{aligned}$$

■

Proof of Corollary 4. A vanishing b_m is equivalent to $c \to 1$. Applying L'Hôpital's rule, we obtain

$$\begin{aligned}\lim_{c\to 1}\frac{(c-1)^2}{(c^N-1)^2} &= \lim_{c\to 1}\frac{2(c-1)}{2N(c^N-1)c^{N-1}} \\ &= \lim_{c\to 1}\frac{2}{2N^2c^{2(N-1)}+2N(N-1)(c^N-1)c^{N-2}} = \frac{1}{N^2}.\end{aligned}$$

This yields

$$\lim_{c\to 1}\frac{L^2}{12}\frac{(c-1)^2}{(c^N-1)^2}\left[\frac{c^{2N}+c^N+1}{c^2+c+1}+3c^{N-1}\right] = \frac{L^2}{12}\frac{1}{N^2}\cdot 4 = \frac{L^2}{3N^2}.$$

Again using L'Hôpital's rule, we find

$$\begin{aligned}&\lim_{c\to 1}\frac{1}{(c-1)^2}\left[\frac{c^{2N}+c^N+1}{c^2+c+1}-c^{N-1}\right] \\ =\ & \lim_{c\to 1}\frac{1}{2(c-1)}\left[\frac{2Nc^{2N-1}+Nc^{N-1}}{c^2+c+1}-\frac{(c^{2N}+c^N+1)(2c+1)}{(c^2+c+1)^2}-(N-1)c^{N-2}\right] \\ =\ & \frac{1}{2}\lim_{c\to 1}\frac{2N(2N-1)c^{2N-2}+N(N-1)c^{N-2}}{c^2+c+1} \\ -\ & \lim_{c\to 1}\frac{(2Nc^{2N-1}+Nc^{N-1})(2c+1)+(c^{2N}+c^N+1)}{(c^2+c+1)^2} \\ +\ & \lim_{c\to 1}\frac{(c^{2N}+c^N+1)(2c+1)^2}{(c^2+c+1)^3}-(N-1)(N-2)c^{N-3} \\ =\ & \frac{1}{2}\left[\frac{2N(2N-1)+N(N-1)}{3}-2\frac{3N+1}{3}+2-(N-1)(N-2)\right] = \frac{N^2-1}{3}.\end{aligned}$$

Hence,

$$\lim_{c\to 1}\frac{4b_a^2c^2}{(c^N+1)^2(c-1)^2}\left[\frac{c^{2N}+c^N+1}{c^2+c+1}-c^{N-1}\right] = \frac{N^2-1}{3}b_a^2$$

and

$$\lim_{c\to 1}\sigma_N^2 = \frac{L^2}{3N^2}+\frac{N^2-1}{3}b_a^2.$$

■

Proof of Lemma 2. The residual variance, σ_N^2, decreases from $N-1$ to N if and only if

$$\begin{aligned} & \frac{L^2}{12}(c-1)^2\left(\frac{c^{2N}+c^N+1}{(c^N-1)^2(c^2+c+1)}-\frac{c^{2N-2}+c^{N-1}+1}{(c^{N-1}-1)^2(c^2+c+1)}\right) \\ + & \frac{L^2}{4}(c-1)^2\left(\frac{c^{N-1}}{(c^N-1)^2}-\frac{c^{N-2}}{(c^{N-1}-1)^2}\right) \\ + & \frac{4b_a^2c^2}{(c-1)^2}\left(\frac{c^{2N}+c^N+1}{(c^N+1)^2(c^2+c+1)}-\frac{c^{2N-2}+c^{N-1}+1}{(c^{N-1}+1)^2(c^2+c+1)}\right) \\ - & \frac{4b_a^2c^2}{(c-1)^2}\left(\frac{c^{N-1}}{(c^N+1)^2)}-\frac{c^{N-2}}{(c^{N-1}+1)^2)}\right) < 0. \end{aligned}$$

Some lengthy calculations show that the left-hand side is equal to

$$\frac{(c+1)^2(c^{3N-3}-c^{N-2})}{c^2+c+1}\left[\frac{4b_a^2c^2}{(c-1)(c^N+1)^2(c^{N-1}+1)^2}-\frac{L^2(c-1)^3}{4(c^N-1)^2(c^{N-1}-1)^2}\right].$$

This term is negative if and only if

$$b_a^2 < \frac{L^2}{16}\frac{(c-1)^4}{c^2}\frac{(c^N+1)^2}{(c^N-1)^2}\frac{(c^{N-1}+1)^2}{(c^{N-1}-1)^2},$$

which is exactly the case if an equilibrium with N partition intervals exists. If an equilibrium exists for all $N\in\mathbb{N}$, we know that $c>1$, thus $\lim_{N\to\infty}\frac{(c-1)^2}{(c^N-1)^2}\frac{c^{2N}+c^N+1}{c^2+c+1} = \frac{(c-1)^2}{c^2+c+1}$, $\lim_{N\to\infty}\frac{(c-1)^2}{(c^N-1)^2}c^{N-1} = 0$, $\lim_{N\to\infty}\frac{c^2(c^{2N}+c^N+1)}{(c-1)^2(c^N+1)^2(c^2+c+1)} = \frac{c^2}{(c-1)^2(c^2+c+1)}$ and $\lim_{N\to\infty}\frac{c^{N+1}}{(c-1)^2(c^N+1)^2}=0$. Hence,

$$\lim_{N\to\infty}\sigma_N^2 = \frac{L^2}{12}\frac{(c-1)^2}{c^2+c+1}+4b_a^2\frac{c^2}{(c-1)^2(c^2+c+1)} = \frac{b_m}{4b_m+3}\frac{L^2}{3}+\frac{b_a^2}{b_m(4b_m+3)}.$$

∎

Proof of Proposition 4. If $N=\infty$ (thus $b_m>0$), we have

$$\sigma_N^2 = \frac{b_m}{4b_m+3}\frac{L^2}{3}+\frac{b_a^2}{b_m(4b_m+3)},$$

which is smaller than $\delta^2 := b_a^2 + b_m^2 \frac{L^2}{3}$ if and only if

$$\begin{aligned} & \left[b_m - \frac{1}{4b_m+3}\right] b_m \frac{L^2}{3} + \left[1 - \frac{1}{b_m(4b_m+3)}\right] b_a^2 > 0 \\ \Leftrightarrow \quad & (4b_m^2 + 3b_m - 1)(b_m^2 L^2 + 3b_a^2) > 0 \\ \Leftrightarrow \quad & (4b_m - 1)(b_m + 1) > 0 \Leftrightarrow b_m > \frac{1}{4}, \end{aligned}$$

given that b_m is positive. ■

Proof of Proposition 5. If $b_m < 0$ and $N = 1$, communication dominates delegation if and only if $\frac{L^2}{3} < b_a^2 + b_m^2 \frac{L^2}{3}$, which is equivalent to $b_a^2 > \frac{L^2}{3}(1 - b_m^2)$. In particular, communication always dominates delegation if $b_m < -1$.
In an equilibrium with $N = 2$ intervals, we have $b_m \in (-1, 0)$ and communication dominates delegation if

$$\frac{L^2}{12} + \frac{4c^2 b_a^2}{(c^2+1)^2} < b_a^2 + b_m^2 \frac{L^2}{3} \quad \Leftrightarrow \quad b_a^2 < \frac{L^2}{48} \frac{(1 - 4b_m^2)(1 + 2b_m)^2}{b_m(b_m + 1)}.$$

In particular, communication can only dominate delegation if $1 - 4b_m^2 < 0$ or $b_m < -\frac{1}{2}$ in the area where $N(b) \geq 2$.

If communication improves, i.e. $N \geq 3$, we have $b_m \in (-\frac{1}{2}, 0)$ and $b_a^2 < \frac{L^2}{4}(1 + 2b_m)^2$. We now show that in this parameter range delegation always dominates communication. Indeed, using the definition of r_k as in the proof of Proposition 3, we obtain for the difference between the residual variance and the loss from delegation,

$$\sigma_N^2 - b_a^2 - b_m^2 \frac{L^2}{3} = L^2 \frac{b_m(b_m+1)}{3} \frac{r_N(1 - 4b_m) + 12}{(4b_m+3) r_N} + b_a^2 (b_m + 1) \frac{r_N(1 - 4b_m) - 16b_m}{b_m(4b_m+3)(r_N+4)}.$$

Multiplication with $\frac{3b_m(4b_m+3)r_N(r_N+4)}{b_m+1} > 0$ yields that the above term is positive, and hence delegation performs better than communication, if and only if

$$3b_a^2 \left[r_N^2(1 - 4b_m) - 16 r_N b_m\right] + L^2 b_m^2 \left[r_N^2 (1 - 4b_m) + 16 r_N (1 - b_m) + 48\right] \geq 0. \quad \text{(A2.1)}$$

The second summand in (A2.1) is non-negative given that $r_N \in [-4, 0]$ and $b_m > -\frac{1}{2}$. In order to see this, note that

$$r_N^2(1-4b_m) + 16r_N(1-b_m) + 48 \geq 0 \quad \Leftrightarrow \quad \left| r_N + \frac{8(1-b_m)}{1-4b_m} \right| \geq \frac{4(2b_m+1)}{1-4b_m}.$$

Then, $\frac{8(1-b_m)}{1-4b_m} > 4$ if $b_m > -\frac{1}{2}$, thus $r_N + \frac{8(1-b_m)}{1-4b_m} > 0$ if $r_N \in [-4, 0]$ and the above expression is equivalent to $r_N \geq -4$, which holds for all $N \in \mathbb{N}$. Hence, condition (A2.1) is satisfied if $r_N^2(1-4b_m) - 16r_N b_m \geq 0$.

Now we assume that $r_N^2(1-4b_m) - 16r_N b_m < 0$. In this case, we have

$$\begin{aligned} & 3b_a^2 \left[r_N^2(1-4b_m) - 16r_N b_m \right] + L^2 b_m^2 \left[r_N^2(1-4b_m) + 16r_N(1-b_m) + 48 \right] \\ > \; & \frac{L^2}{4} \left[r_N^2(1-4b_m)(16b_m^2 + 12b_m + 3) - 16r_N b_m(16b_m^2 + 8b_m + 3) + 192b_m^2 \right], \end{aligned}$$

using $b_a^2 < \frac{L^2}{4}(1+2b_m)^2$. This term is non-negative if and only if

$$\begin{aligned} & \left| r_N - \frac{8b_m(16b_m^2 + 8b_m + 3)}{(1-4b_m)(16b_m^2 + 12b_m + 3)} \right| \\ \geq \; & \frac{64b_m^2(16b_m^2 + 8b_m + 3)^2 - 192b_m^2(1-4b_m)(16b_m^2 + 12b_m + 3)}{(1-4b_m)^2(16b_m^2 + 12b_m + 3)^2}. \end{aligned}$$

It turns out that $64b_m^2(16b_m^2+8b_m+3)^2 - 192b_m^2(1-4b_m)(16b_m^2+12b_m+3) < 0$ if $b_m > -\frac{1}{2}$, which completes the proof. ∎

Proof of Lemma 3. An equilibrium with $N = 2$ partition intervals exists if the arbitrage condition

$$x_1 + x_2 = 2b(a_1)$$

is fulfilled. In this case, $x_1 + x_2 = h(a_1)$. Since both $h(a_1)$ and $2b(a_1)$ are strictly increasing in a_1, the functions intersect in some $a \in supp(f)$ if and only if

1. $\lim_{a \to \theta_L} h(a) > 2 \lim_{a \to \theta_L} b(a)$ and $\lim_{a \to \theta_H} h(a) < 2 \lim_{a \to \theta_H} b(a)$, or
2. $\lim_{a \to \theta_L} h(a) < 2 \lim_{a \to \theta_L} b(a)$ and $\lim_{a \to \theta_H} h(a) > 2 \lim_{a \to \theta_H} b(a)$.

∎

Proof of Corollary 5. For a general symmetric distribution $F(\theta)$ supported on $[-L, L]$, we have $E[\theta] = 0$. Hence, $h(L) = L$, $h(-L) = -L$ and $2b(L) = 2b_a + 2(b_m + 1)L$, $2b(-L) = 2b_a - 2(b_m + 1)L$. Considering both cases in Lemma 3 together is thus equivalent to

$$|b_a| < \frac{|2b_m + 1|}{2} L.$$

■

Proof of Lemma 4. Let $a_1 \in supp(f)$ be given. If we consider $x_1 + x_2 = 2b(a_1)$ as a continuous function in a_2 for a given a_1, there exists an $a_2 \in supp(f)$ with $a_2 > a_1$ and $x_1 + x_2 = 2b(a_1)$ if and only if

$$\begin{aligned} x_1 + a_1 &< 2b(a_1) \quad \text{and} \\ x_1 + E[\theta|\theta > a_1] &> 2b(a_1). \end{aligned} \tag{A2.2}$$

Since $h(a_1) = x_1 + E[\theta|\theta > a_1]$, an equilibrium with $N = 3$ partition intervals is only possible if $h(a_1) > 2b(a_1)$. Analogously, if $a_2 \in supp(f)$ is given, there exists $a_1 \in supp(f)$ with $a_1 < a_2$ and $x_2 + x_3 = 2b(a_2)$ if and only if

$$\begin{aligned} E[\theta|\theta < a_2] + x_3 &< 2b(a_2) \quad \text{and} \\ a_2 + x_3 &> 2b(a_2). \end{aligned} \tag{A2.3}$$

Hence, an equilibrium with $N = 3$ partition intervals requires $h(a_2) < 2b(a_2)$. Since $a_1 < a_2$ and $h(a)$ is strictly increasing in $a \in supp(f)$, (A2.2) and (A2.3) can only be satisfied if

$$\lim_{a \to \theta_L} h(a) > \lim_{a \to \theta_L} 2b(a) \text{ and } \lim_{a \to \theta_H} h(a) < \lim_{a \to \theta_H} 2b(a),$$

representing case 1 in the description. ■

Proof of Proposition 6. Let $a_1 < \ldots < a_i < \theta_0$ be a sequence in $supp(f)$ satisfying the arbitrage condition. Then there exists exactly one a_{i+1} with $a_i < a_{i+1} < \theta_0$, if $x_i + x_{i+1}$ (which increases in a_{i+1}) lies below $2b(a_i)$ for $a_{i+1} = a_i$, and above $2b(a_i)$ for $a_i = \theta_0$.

Formally, this yields

$$\begin{aligned} x_i + a_i &< 2b(a_i) \quad \text{and} \\ x_i + E[\theta|\theta \in (a_i, \theta_0)] &> 2b(a_i). \end{aligned}$$

The argument for a sequence $A_1 > \ldots > A_i > \theta_0$ follows analogously. ■

Proof of Corollary 6. Let f be the uniform distribution on $[-L, L]$ for some $L > 0$. Then the conditions for an ascending sequence in Proposition 6 are equivalent to

$$\begin{aligned} 4b_a &< a_{i-1} - (4b_m + 2)a_i + \theta_0 \quad \text{and} \\ 4b_a &> a_{i-1} - (4b_m + 1)a_i. \end{aligned}$$

Using the recursive formula

$$a_i = U_i(v,1)a_1 + U_{i-1}(v,1)L + 4b_a \sum_{j=1}^{i-1} U_j(v,1),$$

as in the proof of Proposition 1, this yields

$$\begin{aligned} U_{i+1}(v,1)a_1 + U_i(v,1)L + 4b_a \sum_{j=1}^{i} U_j(v,1) &< \theta_0 \quad \text{and} \\ (U_{i+1}(v,1) - U_i(v,1))a_1 + (U_i(v,1) - U_{i-1}(v,1))L + 4b_a U_i(v,1) &> 0. \end{aligned}$$

Substituting $U_i(v,1) = \frac{c^i - d^i}{c-d}$ and $\theta_0 = -\frac{b_a}{b_m} = -\frac{4b_a c}{(c-1)^2}$ yields

$$\begin{aligned} b_a &< -\frac{c-1}{4c}\left[\frac{c(c^{2i}-1)}{c^{2i+1}+1}L + \frac{c^{2i+2}-1}{c^{2i+1}+1}a_1\right] \quad \text{and} \\ b_a &> -\frac{c-1}{4c}\left[\frac{c(c^{2i-1}+1)}{c^{2i}-1}L + \frac{c^{2i+1}+1}{c^{2i}-1}a_1\right]. \end{aligned}$$

In the limit, as $i \to \infty$, the right-hand sides of both inequality conditions converge to $-\frac{c-1}{4c}\left[L + ca_1\right]$, using $c > 1$. As the above conditions have to hold for some $a_1 \in (-L, \theta_0)$

in the limit, this yields on the one hand

$$b_a < -\frac{c-1}{4c}\left[L + ca_1\right] < \frac{(c-1)^2}{4c}L = b_m L,$$

since $a_1 > -L$. On the other hand, we have $b_a > -\frac{c-1}{4c}\left[L + ca_1\right] > -\frac{c-1}{4c}L + \frac{b_a c}{c-1}$, since $a_1 < \theta_0 = -\frac{4b_a c}{(c-1)^2}$. This is equivalent to

$$b_a < \frac{(c-1)^2}{4c}L = b_m L.$$

The respective conditions for a descending sequence yield $b_a > -b_m L$, hence $|b_a| < b_m L$ has to be satisfied in order to obtain infinitely many partition intervals. ∎

Proof of Lemma 5. From $N(b_a, b_m) = \infty$ we conclude that $b_m > 0$. In the following, we show that for any $R > 0$, there exists a $\delta_R > 0$ such that for any (b_m, b_a) with $\|(b_m, b_a)\| = \sqrt{b_m^2 + b_a^2} < \delta_R$ we have

$$\frac{L^2 b_m^2 + 3b_a^2}{(4b_m + 3)b_m(b_a^2 + L^2 b_m^2)} > R.$$

Let thus $R > 0$ be given. We choose $\delta_R > 0$ small enough such that $(4\delta_R + 3)\delta_R < \frac{1}{R}$. Then $b_m^2 \leq b_a^2 + b_m^2 < \delta_R^2$, thus $b_m < \delta_R$, and we have

$$\begin{aligned}\frac{L^2 b_m^2 + 3b_a^2}{(4b_m + 3)b_m(b_a^2 + L^2 b_m^2)} &= \frac{2b_a^2}{(4b_m + 3)b_m(b_a^2 + L^2 b_m^2)} + \frac{1}{(4b_m + 3)b_m} \\ &> 0 + \frac{1}{(4\delta_R + 3)\delta_R} > R.\end{aligned}$$

∎

Proof of Proposition 7. The proof of Proposition 7 proceeds largely analogous to the proof of Proposition 3 in [D].

We focus on the most informative partition equilibrium given $b = (b_m, b_a)$. Let $h_i := a_i - a_{i-1}$ be the length of the ith partition element and let $\bar{h}$ be the largest partition element. The proof of Proposition 7 follows directly from the following three lemmata.

Lemma A2.1. *As $\bar{h}$ tends to zero,*

$$h_{i+1} = h_i + 4(b_a + b_m a_i) - [h_i^2 + h_{i+1}^2]\frac{f'(a_i)}{6f(a_i)}$$

where we have neglected all terms in the 3rd or higher order of h_i and h_{i+1}.

Proof. The proof proceeds as in [D]. In our case, the agent's indifference condition is given by

$$x_{i+1} - (b_a + (b_m + 1)a_i) = b_a + (b_m + 1)a_i - x_i,$$

which yields the result. ■

In the following, we assume that the convergence path of $b = (b_m, b_a)$ stays in one quadrant if $\|b\|$ is small enough.

Lemma A2.2. *If* $\lim_{b\to 0} \bar{h} = 0$*, then*

$$\lim_{b\to 0} \frac{E[(x^*(\theta) - \theta)^2]}{b_a^2 + b_m^2\sigma^2} = +\infty.$$

Proof. If $\theta_0 = -\frac{b_a}{b_m} < 0$, we fix two arbitrary points $\sigma < y_1 < y_2$ in $supp(f)$, and if $\theta_0 > 0$, we choose $[y_1, y_2] \in supp(f)$ with $-\sigma < y_1 < y_2$. Since $\theta_0 \notin [y_1, y_2]$, $[y_1, y_2]$ contains only finitely many partition intervals in equilibrium for all $b = (b_m, b_a)$. Moreover, $E[(x^*(\theta) - \theta)^2 | \theta \in (y_1, y_2)] \leq \frac{1}{F(y_2) - F(y_1)} E[(x^*(\theta) - \theta)^2]$ and $0 < F(y_2) - F(y_1) < \infty$, hence it is sufficient to show that

$$\lim_{b\to 0} \frac{E[(x^*(\theta) - \theta)^2 | \theta \in (y_1, y_2)]}{b_a^2 + b_m^2\sigma^2} = +\infty \tag{A2.4}$$

in order to prove Lemma A2.2. Slightly abusing notation, we denote by $N(b) < \infty$ the number of partition elements fully contained in $[y_1, y_2]$ for a given $b = (b_m, b_a)$ and by h_i the length of the ith partition interval fully included in $[y_1, y_2]$.

We define $\overline{\mu} := \max_{\theta\in(y_1,y_2)} \frac{f'(\theta)}{6f(\theta)}$ and $\underline{\mu} := \min_{\theta\in(y_1,y_2)} \frac{f'(\theta)}{6f(\theta)}$. Furthermore, we denote

$$\mu := \begin{cases} \overline{\mu}, & \text{if } b_a > 0 \\ -\underline{\mu}, & \text{if } b_a < 0. \end{cases}$$

In the following, we first show that

$$\lim_{b\to 0} \frac{E[h_i^2]}{b_a^2 + b_m^2\sigma^2} = +\infty \tag{A2.5}$$

holds (Part A and B), and then that the latter implies (A2.4) (Part C).

Let $q(b)$ denote the number of partition intervals in $[y_1, y_2]$ with $\mu h_i^2 \leq |b_a| + |b_m|\sigma$ and $Q(b)$ the number of partition intervals in $[y_1, y_2]$ with $\mu h_i^2 > |b_a| + |b_m|\sigma$. Then we have $q(b) + Q(b) = N(b)$. By assumption, if $\|b\|$ goes to zero, $\bar{h}$ vanishes and $q(b) + Q(b)$ goes to infinity. We consider two cases:

(A) First, assume that

$$\lim_{b\to 0} \frac{Q(b)}{N(b)} =: \Phi > 0.$$

This implies that $Q(b) > 0$ for some b and thus $\mu > 0$. Then we have

$$\begin{aligned} \lim_{b\to 0} \frac{E[h_i^2]}{b_a^2 + b_m^2\sigma^2} &\geq \lim_{b\to 0} \frac{1}{b_a^2 + b_m^2\sigma^2} \frac{Q(b)E[h_i^2|\mu h_i^2 > |b_a| + |b_m|\sigma]}{N(b)} \\ &= \lim_{b\to 0} \Phi \frac{E[h_i^2|\mu h_i^2 > |b_a| + |b_m|\sigma]}{b_a^2 + b_m^2\sigma^2} \\ &\geq \lim_{b\to 0} \Phi \frac{1}{b_a^2 + b_m^2\sigma^2} \frac{|b_a| + |b_m|\sigma}{\mu} = +\infty, \end{aligned}$$

since $\lim_{b_a\to 0} \frac{|b_a|}{b_a^2+b_m^2\sigma^2} = +\infty$ for all $b_m \in \mathbb{R}$, and $\lim_{b_m\to 0} \frac{|b_m|}{b_a^2+b_m^2\sigma^2} = +\infty$ for all $b_a \in \mathbb{R}$.

(B) Second, assume that

$$\lim_{b\to 0} \frac{Q(b)}{N(b)} = 0.$$

Let us denote by $\bar{n}(b) < \infty$ the average number of partition intervals of a series of adjacent partition elements for which $\mu h_i^2 \leq |b_a| + |b_m|\sigma$. For any two adjacent partition elements h_i and h_{i+1} in such a series, from Lemma A2.1 we estimate the minimal increase in interval lengths if $b_a > 0$, and the maximal decrease if $b_a < 0$, respectively. For $b_a > 0$ it follows that

$$\begin{aligned} h_{i+1} &\geq h_i + 4(b_a + b_m a_i) - \overline{\mu}[h_i^2 + h_{i+1}^2] \\ &\geq h_i + 4(b_a + b_m a_i) - 2(b_a + |b_m|\sigma) \\ &\geq h_i + 2(b_a + |b_m|\sigma), \end{aligned}$$

since $b_m a_i > \sigma|b_m|$ by the choice of the interval $[y_1, y_2]$. If $b_a < 0$, we have

$$\begin{aligned} h_{i+1} &\leq h_i + 4(b_a + b_m a_i) - \underline{\mu}[h_i^2 + h_{i+1}^2] \\ &\leq h_i + 4(b_a + b_m a_i) + 2(-b_a + |b_m|\sigma) \\ &\leq h_i + 2(b_a - |b_m|\sigma), \end{aligned}$$

since $b_m a_i < -\sigma|b_m|$ by the choice of the interval $[y_1, y_2]$. It follows for $b_a < 0$ that $h_i \geq h_{i+1} - 2(b_a - |b_m|\sigma)$, thus in general, the minimal increase from left to right ($b_a > 0$), or from right to left ($b_a < 0$), is given by $2(|b_a| + |b_m|\sigma)$. Hence, we have

$$\sum_{i=1}^{\bar{n}(b)} h_i^2 \geq \sum_{i=1}^{\bar{n}(b)-1} [2(|b_a| + |b_m|\sigma)i]^2 = \frac{2}{3}(|b_a| + |b_m|\sigma)^2 \bar{n}(b)[\bar{n}(b) - 1][2\bar{n}(b) - 1],$$

and consequently,

$$\begin{aligned} E[h_i^2|\mu h_i^2 \leq |b_a| + |b_m|\sigma] &\geq \frac{1}{\bar{n}(b)} \sum_{i=1}^{\bar{n}(b)-1} [2(|b_a| + |b_m|\sigma)i]^2 \\ &= \frac{2}{3}(|b_a| + |b_m|\sigma)^2[\bar{n}(b) - 1][2\bar{n}(b) - 1]. \quad \text{(A2.6)} \end{aligned}$$

If $\lim_{b\to 0} \frac{Q(b)}{N(b)} = 0$ holds, then $\lim_{b\to 0} E[h_i^2] = \lim_{b\to 0} E[h_i^2|\mu h_i^2 \leq |b_a| + |b_m|\sigma]$. Second, we have $\lim_{b\to 0} \bar{n}(b) = \infty$. In order to see that, note that the number of series of adjacent partition elements for which $\mu h_i^2 \leq |b_a| + |b_m|\sigma$ is at most $Q(b)$. Hence,

$N(b) = Q(b) + q(b) \leq Q(b) + Q(b)\bar{n}(b)$. Thus, if we suppose $\lim_{b\to 0} \bar{n}(b) = n^* < \infty$, then $\lim_{b\to 0} \frac{Q(b)}{N(b)} \geq \lim_{b\to 0} \frac{Q(b)}{Q(b)+Q(b)\bar{n}(b)} = \frac{1}{1+n^*} > 0$, a contradiction. Finally, we have

$$\lim_{b\to 0} \frac{\frac{2}{3}(|b_a| + |b_m|\sigma)^2}{b_a^2 + b_m^2\sigma^2} = \lim_{b\to 0} \frac{2}{3} + \frac{4}{3} \underbrace{\frac{|b_a||b_m|\sigma}{b_a^2 + b_m^2\sigma^2}}_{\geq 0} \in \left[\frac{2}{3}, \frac{4}{3}\right],$$

since $(|b_a| - |b_m|\sigma)^2 = b_a^2 - 2|b_a||b_m|\sigma + b_m^2\sigma^2 \geq 0$ and hence $\frac{|b_a||b_m|\sigma}{b_a^2+b_m^2\sigma^2} \leq \frac{1}{2}$. Taking all these facts together, it follows from (A2.6) that

$$\begin{aligned} \lim_{b\to 0} \frac{E[h_i^2]}{b_a^2 + b_m^2\sigma^2} &= \lim_{b\to 0} \frac{E[h_i^2|\mu h_i^2 \leq |b_a| + |b_m|\sigma]}{b_a^2 + b_m^2\sigma^2} \\ &\geq \lim_{b\to 0} [\bar{n}(b) - 1][2\bar{n}(b) - 1] \frac{2(|b_a| + |b_m|\sigma)^2}{3(b_a^2 + b_m^2\sigma^2)} = +\infty. \end{aligned}$$

(C) Completely analogous to [D], we find that (A2.5) implies (A2.4).

■

Lemma A2.3. *If* $\lim_{b\to 0} \bar{h} > 0$, *then* $\lim_{b\to 0} \frac{E[(x^*(\theta)-\theta)^2]}{b_a^2+b_m^2\sigma^2} = +\infty$.

Proof. The proof is analogous to the proof of Lemma A3 in [D]. ■

■

Proof of Lemma 6. Communication performs at least as good as if the principal takes an uninformed decision. In this case, the principal chooses $x = E[\theta]$ and his loss is given by $E[E[\theta] - \theta)^2] = \sigma^2$. On the other hand, if he delegates the project choice to the agent, the principal incurs a loss of $E[(b_a + b_m\theta)^2] = b_a^2 + b_m^2(E[\theta]^2 + \sigma^2)$. Thus, communication dominates delegation at least if

$$\begin{aligned} &\sigma^2 < b_a^2 + b_m^2(E[\theta]^2 + \sigma^2) \\ \Leftrightarrow\ & b_a^2 + b_m^2 E[\theta]^2 + (b_m^2 - 1)\sigma^2 > 0, \end{aligned}$$

which is at least the case if $|b_m| > 1$. ■

Chapter 3

Promises and Image Concerns*

Cooperation among interacting partners is essential for economic success in many situations, as joint value creation often exceeds individual achievements. These situations become challenging as soon as cooperation cannot be contractually enforced, but relies on mutual trust by the interacting partners. Among a large literature focusing on how to improve cooperation, various experimental studies show that communication can be an effective tool to enhance it (see, e.g. Cooper et al., 1992; Ellingsen and Johannesson, 2004; Bochet and Putterman, 2009). While several articles analyze whether cheap talk can be effective and how this depends on the communication protocol and the game structure (see, for instance, Blume and Ortmann, 2007; Mohlin and Johanneson, 2008; Ellingsen and Östling, 2010; Camera et. al., 2011; Kriss et al., 2011), we contribute to the literature focusing on *why* individuals stick to a commitment, given that rationality predicts a deviating behavior. In particular, we analyze whether and to what extent social image concerns motivate people to stick to a promise. More precisely, as reneging on a promise is deemed negatively in society, avoiding the image of being a promise breaker might induce individuals to keep their word. Consequently, we study whether an individual is more likely to deliver on a promise if its violation is more obvious to its receiver.

In order to test whether social image concerns influence promise keeping behavior, we conduct a controlled laboratory experiment. Here, subjects are randomly matched in

*This chapter is based on joint work with Carmen Thoma.

pairs of two and play a one-shot sequential trust game similar to the one used in Charness and Dufwenberg (2006). A first mover (A) decides whether to enter the game or to opt out, the latter choice inducing a low outside option for both players. If A enters the game, a second mover (B) chooses between a selfish option, yielding a payoff of zero for A, and cooperation, in which case a chance move determines whether A gets a positive payoff or 0.[1] Prior to the strategic decisions, B sends one out of three pre-defined messages to A, one of which is a promise to cooperate. In order to test for social image concerns, we vary the ex-post observability of the second mover's action. While in condition *Rev* A learns B's action choice, in condition *NoRev* she cannot infer whether a payoff of zero is due to B behaving selfish or just to bad luck.[2] We hypothesize that a higher share of Bs cooperate if B's action is revealed to A (*Rev*) than if it is concealed (*NoRev*), assuming that a fraction of Bs has a preference for avoiding the image to be a promise breaker.

By the choice of our experimental design, we attempt to differentiate social image concerns from other possible reasons for promise keeping by second movers. Up to now, the literature mainly provides two motivations why individuals might stick to their promises. First, Charness and Dufwenberg (2006) explain promise keeping by simple guilt, i.e. the aversion to disappoint other people's expectations, as introduced by Batigalli and Dufwenberg (2007). If B promises cooperation, A expects a higher payoff, which increases B's guilt in case he refuses to cooperate. However, in our experiment only the game structure and the payoffs are common knowledge, but Bs are privately informed about the revelation condition. In contrast, As are not even aware that different conditions exist. Thus, As' first-order beliefs, and consequently Bs' second-order beliefs should not vary across conditions, inducing the same amount of guilt for non-cooperation in both conditions.[3] Second, Vanberg (2008) claims that subjects have a preference for keeping their promises per se, independent of others' expectations. This assumption cannot explain a difference in Bs'

[1]While rational behavior predicts the second mover to behave selfish, and therefore the first mover not to enter the game, mutual cooperation is the unique Pareto-optimal outcome, which generates the highest joint payoff.

[2]Conditions are assigned randomly to pairs.

[3]Otherwise A might expect B to choose *Roll* with a higher probability if his choice is revealed, inducing higher simple guilt in *Rev* than in *NoRev* (if we assume consistent beliefs).

behavior across conditions either, as the preference for keeping a promise should be independent of As' ex-post information.

Yet, as the revelation of B's action choice might also induce a concern of being perceived as selfish (Tadelis, 2011), we conduct a control treatment without communication (*NoCom*). We claim that the effect of revelation on behavior in treatment *Com* is larger than the respective effect in *NoCom*, indicating that the differential effect is due to the mere aversion of being perceived as a promise breaker, additional to the aversion to an egoistic image.

With pre-play communication, we observe marginally significantly more cooperation in *Rev* than in *NoRev*. This effect does not seem to be driven by shame to be selfish alone, as without communication revelation even marginally decreases cooperation rates in *Rev* compared to *NoRev*. However, although conditions are identical at the pre-play communication stage, the number of promises sent is significantly higher in *Rev* than in *NoRev*. Thus, the higher *Roll* rate in *Rev* might only be driven by a higher number of promises and not by image concerns of being perceived as a promise breaker. When comparing the share of promises kept, we do observe a slightly higher rate in *Rev* (85%) than in *NoRev* (81%), however the difference is not significant. Thus, we fail to prove our hypothesis that avoiding the image of being perceived as a promise breaker plays a significant role in the individual decision to keep a given promise.

It is worth noting that the high promise keeping rate without revelation (81%) limits the scope for further increase. In treatment *Com1*, where Bs can choose between a promise to cooperate, a statement of intent, and an empty message, this high promise keeping rate might be partly due to the fact that Bs who attempt to influence their interaction partner without planning to cooperate have the possibility to send a statement of intent. In order to reduce the promise keeping rate without revelation by forcing this type of subjects to either break a promise or refrain from influencing the interaction partner, we exclude the opportunity of stating an intention in a further treatment, *Com2*. However, we do not observe a significant effect of revelation in *Com2* either.

Still, this design variation provides another interesting finding. The menu of messages available to B seems to play a significant role for the effectiveness of communication as Bs are significantly more likely to keep a promise than to stick to a statement of intent. Hence, intentions seem to be less costly to break than promises. In contrast, As who are unaware of the available messages, seem to trust intentions to the same amount as promises.

Literature.

This chapter is mainly related to two strands of the economic literature. First, there is an expanding literature analyzing the effect of non-binding communication on behavior. Experimental studies show that communication can increase coordination (Blume and Ortmann, 2007; Ellingsen and Östling, 2010; Kriss et al., 2011), generosity in a dictator game (Mohlin and Johannesson, 2008; Andreoni and Rao, 2011), and, most relevant for our study, cooperation (Cooper et al., 1992; Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006; Vanberg, 2008; Bochet and Putterman, 2009; Charness and Dufwenberg, 2010). So far, mainly two reasons for the effectiveness of communication have been identified. On the one hand, guilt aversion in the sense of Batigalli and Dufwenberg (2007) has been found to induce promise keeping (see, for instance, Charness and Dufwenberg, 2006, or Beck et al., 2013).[4] On the other hand, individuals can exhibit a preference for promise keeping per se, that is, promises have a commitment value (Vanberg, 2008; Ismayilov and Potters, 2012). Likewise, individuals might face costs of lying (Fischbacher and Föllmi-Heusi, 2013; Mazar et al., 2008; Hurkens and Kartik, 2009; Lundquist et al., 2009). However, none of these papers consider social image concerns as a reason why people stick to their promises.

A second related strand in the behavioral economics literature studies the effect of social image concerns on behavior. Following Ariely et al. (2009), "image motivation [...] refers to an individual's tendency to be motivated partly by others' perceptions." That is, individuals dislike to publicly violate a social norm, such as altruism or modesty. Cor-

[4]However, the effect of guilt aversion has been found to be relatively small (Ellingsen et al., 2010).

respondingly, evidence for individuals behaving more selfishly or greedily if their action is less likely to be observed, has for example been found in experimental dictator games (Dana et al., 2006; Broberg et al. 2007; Dana et al., 2007; Koch and Normann, 2008; Larson and Capra, 2009; Andreoni and Bernheim, 2009; Grossman, 2010a, 2010b), and in the context of volunteering (e.g. Linardi and McConnell, 2008; Carpenter and Myers, 2010) or donations (Ariely et al., 2009; Lacetera and Macis, 2010; DellaVigna et al., 2012).

Similar to our experimental design, Tadelis (2011) builds on the framework by Charness and Dufwenberg (2006) and varies the ex post information of the first mover. He shows that image concerns to appear selfish (the "shame" effect) exist and increase cooperation, especially if anonymity is lifted by announcing the second mover's action choice to all participants in the room. However, subjects in his setting are not able to communicate.

In our study, we combine these two strands of literature and investigate whether social image concerns are even more pronounced with communication, due to the aversion of being perceived as a promise breaker. Bracht and Regner (2011) also analyze social image concerns in a similar trust game with communication, however, they focus on the correlation of behavior to proneness to shame and guilt, which they elicit via a psychological test.[5] While Bracht and Regner (2011) analyze the effect of transparency and communication separately, we focus on how communication interacts with the effect of revelation on behavior. To the best of our knowledge, social image concerns have rarely been analyzed in the context of communication.

The remainder of this chapter is structured as follows. Section 3.1 introduces the experimental design and the leading hypotheses. In Section 3.2, we analyze and discuss the experimental results. We compare our results to previous research in Section 3.3 and Section 3.4 concludes.

[5]Bracht and Regner (2011) find that disposition to guilt predicts behavior, but not disposition to shame.

3.1 Experimental Design and Hypotheses

3.1.1 Experimental Design

At the beginning of the experiment, role A is assigned to half of the subjects while the other half is assigned role B. One subject with role A and one subject with role B are randomly matched to form a pair.[6] Each pair subsequently plays the one-shot trust game depicted in Figure 3.1, which is akin to the one used by Charness and Dufwenberg (2006), henceforth CD (2006).[7]

A ("she") decides whether to enter the game (*In*) or not (*Out*). Without learning A's decision, B ("he") decides whether to keep a payoff of 30 tokens for himself while A receives nothing (*Don't Roll*), or to let a die decide over A's payoff (*Roll*). In this case, A receives a payoff of 24 tokens with probability $\frac{5}{6}$ and a payoff of 0 with probability $\frac{1}{6}$, while B earns a payoff of 20 tokens in any case. In order to elicit Bs' action choice we use the strategy method, i.e. B decides on his action independent of whether A enters the game or not. At the end of the experiment, one token is converted into 0.25 euros.

We conduct three treatments, called *Com1*, *Com2* and *NoCom*. In *Com1* and *Com2* B sends one out of three predefined messages to A, prior to playing the trust game. *Com1* and *Com2* differ only in the type of messages that can be sent. In *Com1* B can choose between a promise ("I promise to choose *Roll*."), an intention ("I will choose *Roll*.") and an empty message ("Hello, how are you? I'm fine."). In *Com2* B can choose between the same promise and two empty messages ("Hello!" and "How are you?"), i.e. B cannot send an intention in *Com2*. As the design of *Com1* and *Com2* is the same except for the message choices, we sometimes refer to the pooling of both communication treatments as *Com*. *NoCom* is a control treatment, which is identical to the other two treatments, but without pre-play communication.

[6] In the following, we refer to the player with role A (B) as A (B).

[7] In comparison to CD (2006), stakes are lower in our set-up, as one session consists of two separate experiments, which are both paid out (see Section 3.1.3). However, the proportions of the payoffs resulting from different strategies are similar.

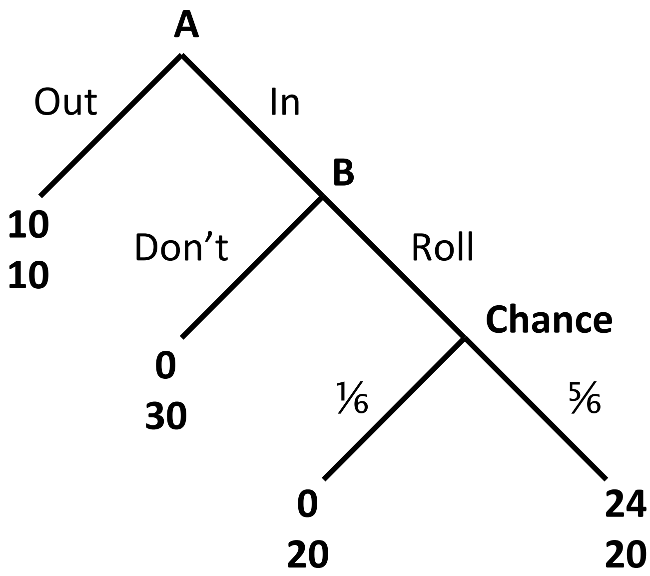

Figure 3.1: Trust game played by each pair of subjects. The upper number refers to A's payoff, the lower one to B's.

Without any further information, A cannot infer whether B has chosen *Roll* or *Don't Roll* whenever she experiences a payoff of 0. However, we are interested in the influence of social image concerns on B's cooperative behavior (see Section 3.1.2), that is, whether B cares about how he is perceived by A. Consequently, we vary whether A can observe B's action choice at the end of the experiment or not, which yields two conditions within each treatment. Before playing the trust game, half of the pairs is randomly assigned to condition "Revelation" *(Rev)*, the other half plays condition "No Revelation" *(NoRev)*. In condition *Rev*, B's choice will be revealed to A at the end of the experiment, whereas A does not learn B's behavior in condition *NoRev*.

B is informed about the condition he plays before choosing between *Roll* and *Don't Roll*, but after having sent a message to A. Thereby, we ensure that only the action choice and not the type of message sent is affected by the condition. In other words, when B chooses the message to be sent, both conditions are exactly equal and Bs' communication behavior should not differ across conditions. Hence, any difference in *Roll* rates across conditions is then due to the variation of the observability of Bs' action choices and not to a difference in messages across conditions.

A neither learns the condition she is playing in, nor is she aware that two different conditions exist until the end of the experiment. The instructions are the same for As and Bs and inform the participants only about the course and the payoffs of the game, without commenting on information structures.[8] Bs receive private information about the condition they play via their screen during the experiment. By not informing A, we ensure that A's first-order belief about B's behavior is constant across conditions. Furthermore, B is explicitly informed about A's unawareness that two conditions exist, thus his second-order belief about A's expectations should not vary across conditions. Therefore, guilt aversion, i.e. the aversion to disappoint A's expectations cannot cause a difference in B's behavior across conditions. We explain the concept of guilt aversion in more detail in Section 3.1.2.

As' first-order and Bs' second-order beliefs are elicited after the trust game, but before subjects learn their payoffs. As were asked: *"What do you think, how many of the x Bs in the room have chosen Roll?"*, where x was substituted by the number of Bs in the session. For Bs, eliciting beliefs is a bit more involved. In a sequential game like the one we consider, B's choice only becomes relevant for those As who choose *In*, thus only the first-order beliefs of those As should matter for B's behavior and his second-order belief. Hence, we asked all Bs: *"We asked all As: "What do you think, how many of the x Bs in the room have chosen Roll?" Consider only the As who chose In. What do you think is the average guess of those As?"*[9] Subjects earn a supplement of 6 tokens for a guess deviating by at most +/-1 from the correct answer. This way, we elicit an interim second-order belief conditional on the event of A choosing *In*.[10] Figure 3.2 provides an overview of the course of the experiment.

[8]However, the instructions emphasize that all Bs throw a die such that Bs' decisions can not be inferred, which is likely to induce a prior of getting no information among As.

[9]This procedure is analogous to the one in CD (2006).

[10]One could argue that observing the actual choice of the A-player is far more influential for beliefs than a hypothetical choice. However, we think that this effect is negligible given that the results show a high correlation between second-order beliefs and actual strategy choices.

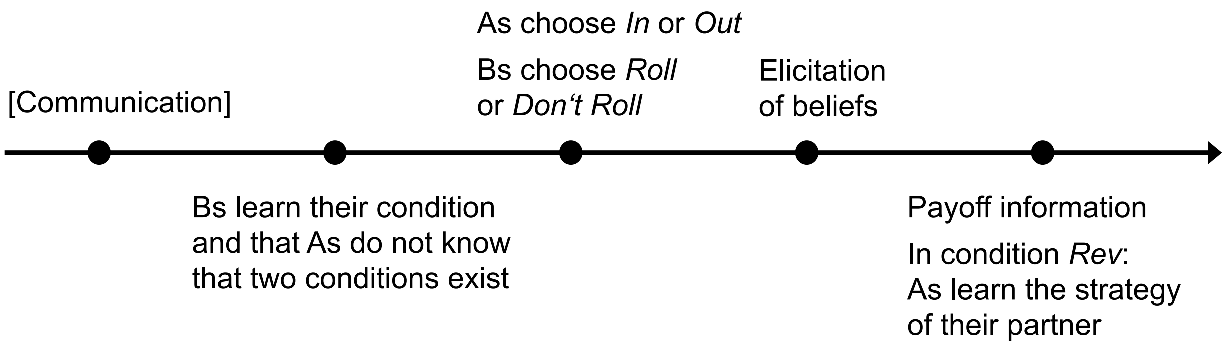

Figure 3.2: The sequence of the experiment.

3.1.2 Hypotheses

In the following, we derive our hypotheses from a notion of social image concerns and subsequently exclude other possible behavioral explanations for our hypotheses.

Hypotheses

Assuming selfish and risk-neutral players, the unique subgame-perfect equilibrium in the trust game illustrated in Figure 3.1 is (*Out, Don't Roll*). However, while the classical game theory claims that non-binding communication cannot influence the players' strategies if information is symmetric, it has been observed in the laboratory that communication indeed enhances cooperation in trust games – promises are made, taken as credible and frequently kept. While CD (2006) argue that subjects keep their promises due to guilt aversion, that is, to not disappoint the increased expectations of the truster, Vanberg (2008) claims that people have a tendency to keep their promise per se, independent of the truster's expectations. Still, in their experiments a considerable share of trustees break a given promise.[11] We analyze whether a change in the set-up, i.e. introducing transparency about the trustee's action induces more trustees to be true on their word. More precisely, we investigate whether social image concerns of being perceived as a

[11]CD (2006) observe that 25% of promisors break their promise without revelation, Vanberg 2008 observes a share of 27% (*no switch* condition).

promise breaker exist and induce individuals to stick to their word. This yields our main hypothesis, which we break down to testable hypotheses in the following.

Main Hypothesis. *The aversion of being perceived as a promise breaker exists and is one reason for why people keep a promise.*

Indeed, it is frequently observed by economists, sociologists and psychologists that people care for how they are perceived by others (e.g. Apsler, 1975; Scheff, 1988; Lewis, 1995; Tangney, 1995; Smith et al., 2002; Grossman, 2010a; Lacetera and Macis, 2010). Applied to our setting, we hypothesize that the trustee is more likely to cooperate if his action choice is revealed than if it is concealed, in a situation where communication is possible.[12]

Hypothesis 1. *The revelation of* `B`*s' action choices induces more cooperation among* `B`*s. In our setting, the Roll rate in* $[Com|Rev]$ *is higher than in* $[Com|NoRev]$.

However, the presence of social image concerns does not necessarily rely on the possibility to communicate. In fact, even without communication evidence for social image concerns has been found, such as the aversion of being perceived as egoistic or greedy (e.g. Güth et al., 1996; Dana et al., 2006, 2007; Koch and Normann, 2008; Ariely et al., 2009; Tadelis, 2011). From a theoretical point of view, Tadelis (2011) proposes a model of "shame" inducing disutility of being perceived as a non-cooperator, in order to explain the effect he observes.[13]

Besides the social disapproval of egoism, we are interested in another social norm which condemns promise breaking, thereby inducing additional social image concerns. Accordingly, we hypothesize that the effect of revelation on *Roll* rates is larger if subjects can communicate than without communication, indicating that the differential effect has to be due to an aversion to be regarded as a promise breaker. Hence, we compare the results of *Com* to the control treatment *NoCom* and state the following hypothesis.[14]

[12] We are aware that the experimenter always observes whether a promise is kept or not and that this can also evoke some social image concerns. However, the presence of the experimenter does not vary across conditions.

[13] "Guilt from blame" (Batigalli and Dufwenberg, 2007) also accounts for more cooperation in the *Rev* condition, based on `B` facing disutility from `A` blaming him for a bad outcome.

[14] We consider the *Roll* rates of all trustees in *Com* rather than focusing on those of the promising

Hypothesis 2. *The effect of revelation on cooperation is larger if pre-play communication takes place. In our setting, the difference between* $[Com|Rev]$ *and* $[Com|NoRev]$ *is larger than the difference between* $[NoCom|Rev]$ *and* $[NoCom|NoRev]$.

Yet, communication might enhance cooperative behavior of `B`s independent of the observability of `B`s' action choice (CD, 2006; Vanberg, 2008). In order to contribute the hypothesized higher *Roll* rate in $[Com|Rev]$ compared to $[Com|NoRev]$ to revelation only, the share of promises has to be equal in both conditions.

Hypothesis 3. *The share of promises among all messages in* $[Com|Rev]$ *is not statistically different from the share in* $[Com|NoRev]$.

Given that `B`s do not know the condition they play at the pre-play communication stage, and `B`s are randomly assigned to both conditions, promising behavior should not differ across conditions. Still, if and only if Hypothesis 3 holds, we can conclude our main hypothesis from Hypotheses 1 and 2.

Elimination of Alternative Explanations

In the following, we argue that, given that we observe the effect in Hypothesis 1, it can neither be due to simple guilt nor to promise-keeping per se.

Simple Guilt. If `B` is subject to *simple guilt*, in the sense of Batigalli and Dufwenberg (2007), he is reluctant to cause a lower payoff for `A` compared to what he believes she expects to earn. Let thus $\alpha_{\mathtt{A}} := \Pr_{\mathtt{A}}(Roll)$ denote `A`'s belief about the probability that `B` cooperates. Then `A` expects to earn a payoff of $\frac{5}{6} \cdot \alpha_{\mathtt{A}} \cdot 24 = 20\alpha_{\mathtt{A}}$ upon entering the game. In turn, `B` forms a belief about `A`'s belief about his action choice, given that `A` chooses *In*. This results in `B`'s interim second-order belief $\beta_{\mathtt{B}} := E[\alpha_{\mathtt{A}}|In]$. By choosing *Don't Roll* conditional on `A` choosing *In*, `B` experiences simple guilt proportional to $20\beta_{\mathtt{B}}$, his belief about the difference between `A`'s payoff expectation and her experienced payoff.

trustees only, as this allows for a comparison of *Roll* rates to the behavior in the control treatment, *NoCom*.

In contrast, if B cooperates, any deception by A cannot be due to B's behavior, thus he doesn't feel guilty. Assuming that B's utility is additively separable in his material payoff and his experienced simple guilt, this yields

$$\begin{aligned} u_{\mathtt{B}}(In, Roll) &= 20 \\ u_{\mathtt{B}}(In, Don'tRoll) &= 30 - \theta^{SG} \cdot 20\beta_{\mathtt{B}}, \end{aligned}$$

where θ^{SG} denotes B's sensitivity to simple guilt.

Simple guilt can explain why communication is able to influence behavior. If B makes a promise, he believes that he influences A's belief about his behavior, i.e. $\beta_{\mathtt{B}}$ increases. Ceteris paribus, this induces a lower payoff for choosing *Don't Roll*, hence a larger share of Bs chooses *Roll* after sending a promise.

While simple guilt delivers an explanation for why communication fosters cooperation, it cannot explain the effect in Hypothesis 1. As As do not learn the condition they play, their first-order beliefs cannot depend on whether Bs' behavior is revealed or not. Bs know about the unawareness among As and thus their second-order beliefs cannot depend on the condition either. Thus, ceteris paribus, guilt aversion predicts the same *Roll* rates for conditions *Rev* and *NoRev*. As the condition is not known to both players at the time communication takes place, the amount of promises should be the same in both conditions (Hypothesis 3). Given that Hypothesis 3 holds, guilt aversion predicts the same *Roll* rates whether B's decision is revealed or not, which contradicts Hypothesis 1.

Self-Image Concerns ("Promise Keeping Per Se"). Vanberg (2008) argues that there exists a preference for promise keeping per se independent of the truster's expectations. He shows that in case an individual faces a different player than the one he made a promise to, his action choice does not depend on whether the new partner has received a promise by another player before or not. In a similar vein, Ellingsen and Johannesson (2004) introduce the notion of "lying cost". They propose a model where inequity averse players suffer from a fixed personal cost of being inconsistent, $l \geq 0$, which in turn leads to a higher commitment power and credibility of promises.

However, whether B's action in the trust game is revealed to A in the end or not does not make a difference to B if he is a "promise-keeper per se". Thus, given an equal number of promises in both conditions (Hypothesis 3), Hypothesis 1 can not be solely induced by promise-keeping per se.

3.1.3 Experimental Procedure

The experiment was conducted in the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA). Subjects were recruited using the online recruitment system ORSEE (Greiner, 2004), and the 406 participants in 17 sessions consisted mainly of students. Upon entering the laboratory, subjects were randomly assigned to 24 visually isolated computer terminals. The instructions were distributed and read out loud by one of the experimenters. Questions were answered individually at the subjects' seats. Before the experiment started, subjects filled out a short questionnaire ensuring the comprehension of the rules.

The experiment was the first of two independent experiments conducted in one session. Before the experiment started, participants were informed that two independent experiments would be conducted, without any further information about the second experiment. Both experiments were paid out at the end of the session, where the average earning was 12.6 EUR, including a fixed show-up fee of 4 EUR. In the first experiment, which is reported in this chapter, As received 3.5 EUR on average, while the mean among Bs was 5.2 EUR. The experiment was programmed and conducted with the software z-tree (Fischbacher, 2007). Each session ended with a detailed questionnaire on demographics and social preferences and lasted about 50 minutes.

3.2 Experimental Results

In this section we first analyze the effect of revelation on Bs' behavior (Section 3.2.1), followed by an investigation of the effects of communication (Section 3.2.2).

3.2.1 The Effect of Revelation on Bs' Behavior

In the following, we pool the data of *Com1* and *Com2* to *Com* in order to analyze the differences between $[Com|Rev]$ and $[Com|NoRev]$. This procedure is justified as the effect of revelation on Bs' behavior does not differ between the two communication treatments. Data considering each treatment separately is gathered in Appendix A3. Our first result provides some evidence for Hypothesis 1.

Result 1. *The Roll rate in* $[Com|Rev]$ *is higher than in* $[Com|NoRev]$*, with the difference being marginally significant.*

Indeed, while 63% of Bs choose *Roll* in $[Com|Rev]$, this share amounts to only 51% in $[Com|NoRev]$ (test of proportions, one-tailed, Z=1.450, p=0.074).[15] Thus, Hypothesis 1 is confirmed on a marginally significant level, indicating that subjects in a situation where communication is possible behave more cooperatively when their action is revealed than when it is not.

The next step is to take a closer look at the source of this marginally significant effect. We claim that Bs behave more cooperatively in $[Com|Rev]$ than in $[Com|NoRev]$ as they do not want to be perceived as a promise breaker. In order to confirm this claim, we have to verify that the higher *Roll* rate in $[Com|Rev]$ is not only caused by an image concern of being perceived as selfish, but rather induced by the combination of communication and revelation (Hypothesis 2). Therefore, we compare the observed effect of revelation in

[15] If we consider *Com1* and *Com2* separately, the effect goes in the same direction, but is no longer significant (*Com1*: 54% vs. 42%, Z=1.062, p=0.144; *Com2*: 72% vs. 61%, Z=1.000, p=0.159, one-tailed test). Throughout this chapter, the Z-Statistics reflect the test of proportions (see Glasnapp and Poggio, 1985) and p-values are on one-tailed tests, because we use our underlying hypotheses, except when reported otherwise.

Com to the one in *NoCom*. Figure 3.3 illustrates the shares of Bs choosing *Roll* in *Com* and *NoCom* separated by condition.

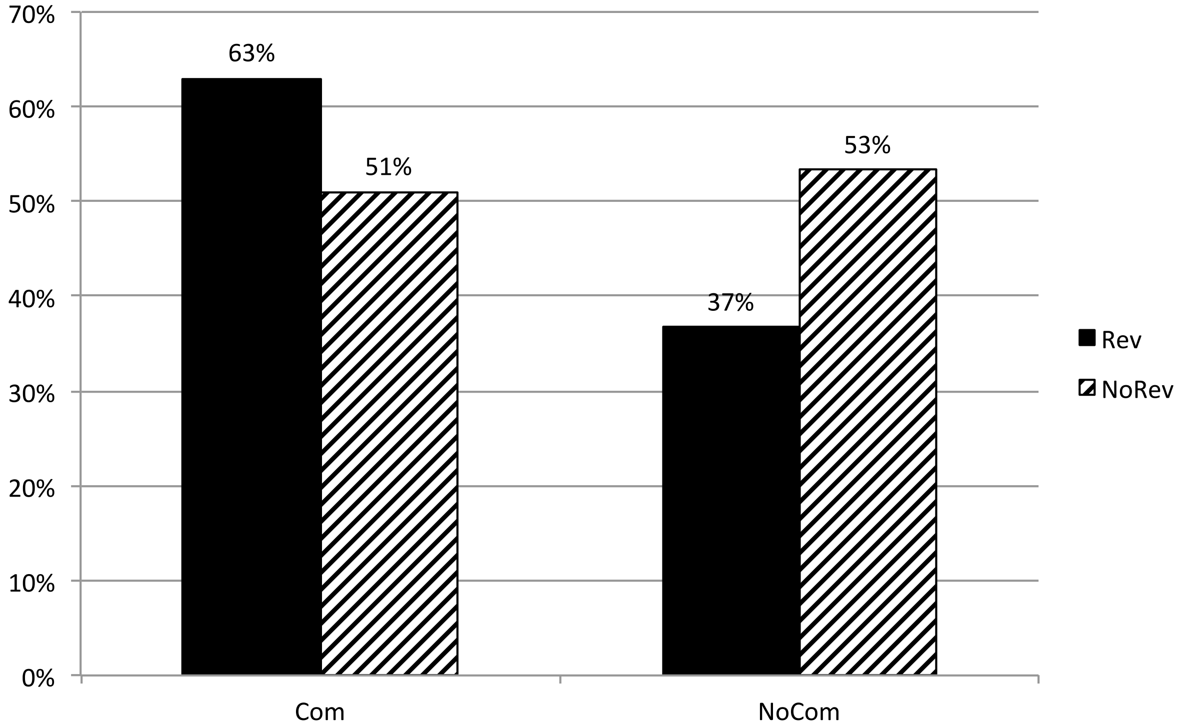

Figure 3.3: *Roll* rates of Bs.

First, we consider the *NoCom* treatment separately and find no evidence for an image concern of being perceived as selfish.

Result 2. *The Roll rate in [NoCom|Rev] is marginally significantly lower than the one in [NoCom|NoRev]. Hence, there is no evidence for the existence of image concerns of being perceived as selfish.*

Indeed, only 37% of Bs choose *Roll* in [*NoCom*|*Rev*] whereas 53% cooperate in [*NoCom*|*NoRev*]. This difference is marginally significant (test of proportions, one-tailed, Z=1.292, p=0.098). Thus, if revelation changes Bs' behavior in *NoCom*, it rather decreases cooperative behavior.[16] This rather unexpected result is unlikely to be a demand effect as Bs are only informed about their own condition, i.e. that their behavior is revealed or not revealed to A, but not about the existence of the other condition. The low *Roll* rate

[16] This result is in contrast to Tadelis (2011).

in [$NoCom|Rev$] might be a sullen behavior due to the sudden announcement that `B`'s action will be revealed to `A`, which was not mentioned in the instructions.

While Result 1 and Result 2 already suggest the confirmation of Hypothesis 2, i.e. that the effect of revelation on cooperation is larger in *Com* than in *NoCom*, we conduct a probit regression to compare the differences across conditions in *Com* and *NoCom*, delivering the following result (Hypothesis 2).

Result 3. *The difference in Roll rates between* [$Com|Rev$] *and* [$Com|NoRev$] *is significantly larger than the one between* [$NoCom|Rev$] *and* [$NoCom|NoRev$]. *Hence, Hypothesis 2 is confirmed.*

The results of the probit regression are reported in Table 3.1. The dependent variable is 1 if `B` chooses *Roll* and 0 otherwise. The independent variables are a dummy for *Com*, a dummy for *Rev* and an interaction of the two.

Table 3.1: Regression of choosing *Roll*

	PROBIT Coefficient (p-value)	OLS Coefficient (p-value)
Com	-0.049	-0.019
	(0.770)	(0.775)
Rev	-0.424	-0.167
	(0.105)	(0.108)
*Com*Rev*	0.731	0.287
	(0.030)	(0.040)
Constant	0.084	0.533
	(0.279)	(0.000)

We cluster standard errors on sessions (17 sessions). Number of observations is 203. In the probit regression Pseudo R-squared is 0.023 and log Pseudo Likelihood is -136.961. In the OLS regression R-squared is 0.031.

We observe that the only significant coefficient is the one of the interaction term *Com*Rev* (p=0.030), which is positive, showing that cooperation among Bs is increased by revelation only in treatment *Com*. The negative and almost marginally significant coefficient of *Rev* indicates the negative effect of revelation on cooperation without communication. The results are robust to an OLS regression, which is also reported in Table 3.1. Thus, Hypothesis 2 is confirmed.

Yet, it remains to show that Bs' communication behavior does not differ between [$Com|Rev$] and [$Com|NoRev$] (Hypothesis 3). For no apparent reason, we are not able to confirm Hypothesis 3.

Result 4. *The share of promises among messages in* [$Com|Rev$] *is significantly higher than the one in* [$Com|NoRev$]. *Hence, Hypothesis 3 is violated and we are not able to conclude the main hypothesis about the existence of an image concern of being perceived as a promise breaker.*

Table 3.2 provides an overview of the messages sent from B to A in *Com*.

Table 3.2: An overview of messages sent in *Com*

		Promise	Intention	Empty
Com	*Rev*	48/71	6/71	17/71
		68%	9%	24%
	NoRev	37/72	13/72	22/72
		51%	18%	31%
	Z stat.	1.975	-1.692	0.888
	(p-value)	(0.024)	(0.045)	(0.187)

The Z Stat reflects the test of proportions for the two treatments or conditions (see Glasnapp and Poggio, 1985). The p-value is on one-tailed tests.

There is neither a difference in the design nor in the instructions of the two conditions. B does not even know that two different conditions exist when sending his message. Still, we observe a significantly higher share of Bs sending a promise in [$Com|Rev$] than in

[*Com*|*NoRev*] (68% vs. 51%, one-tailed test of proportions, Z=1.975, p=0.024).[17] On the other hand, we also observe a significantly smaller share of intentions in [*Com*|*Rev*] than in [*Com*|*NoRev*] (Z=1.692, p=0.045), yielding a similar share of intentions and promises (pooled) in both conditions (76% in [*Com*|*Rev*] vs. 70% in [*Com*|*NoRev*], p=0.448, two-tailed test). However, as further analyzed in Section 3.2.2 and already reported in Table 3.3, subjects sending a promise choose *Roll* significantly more often than subjects sending an intention or an empty message (in *Rev* Z=5.568, p=0.000, in *NoRev* Z=5.183, p=0.000). Thus, we cannot pool intentions and promises, and the communication behavior has to be considered as largely different in both conditions, indicated by a significantly higher share of promises in *Rev* than in *NoRev*.

Therefore, we cannot confirm our main hypothesis via Hypotheses 1 and 2. In order to further investigate what drives the higher *Roll* rate in [*Com*|*Rev*] in comparison to [*Com*|*NoRev*], we examine the behavior of subjects sending a promise separately and compare it between conditions, thereby accounting for the different number of promises.

If the combination of revelation and communication drives the higher *Roll* rate in [*Com*|*Rev*], the share of promise keepers should be higher in condition [*Com*|*Rev*] than in [*Com*|*NoRev*]. As shown in Result 2, revelation itself does not lead to a higher *Roll* rate in comparison to no revelation, hence image concerns of being perceived as selfish play a negligible role in our setting. This allows us to conduct a separate analysis on the set of Bs having sent a promise and attribute a difference in *Roll* rates among promising Bs across conditions to the image concern of being perceived as a promise breaker.[18]

Result 5. *The share of* Bs *keeping their promise among* Bs *who send a promise is slightly higher in* [*Com*|*Rev*] *than in* [*Com*|*NoRev*]*, however the difference is not significant.*

From Result 5, we conclude that the higher *Roll* rate in [*Com*|*Rev*] in comparison to [*Com*|*NoRev*] is mostly driven by the higher number of promises, and not by social

[17] This difference is not driven by one or two sessions, but occurs in all sessions of both communication treatments. It is only marginally significant if we consider *Com1* and *Com2* separately (see Appendix A3).

[18] If there was a higher *Roll* rate in [*NoCom*|*Rev*] than in [*NoCom*|*NoRev*], this analysis would not be meaningful since we cannot compare the effect of revelation among Bs sending a promise in *Com* to the overall effect in *NoCom*. Therefore, we started off with considering overall *Roll* rates in *Com*.

image concerns of being perceived as a promise breaker. Table 3.3 reports the *Roll* rates for each type of message sent in both conditions.

Table 3.3: *Roll* rates by type of message sent

		Promise	Intention	Empty
Com	*Rev*	41/48	1/6	3/17
		85%	17%	18%
	NoRev	30/37	4/13	3/22
		81%	31%	14%
	Z stat.	0.534	-0.650	0.344
	(p-value)	(0.270)	(0.258)	(0.365)

The Z Stat reflects the test of proportions for the two treatments or conditions (see Glasnapp and Poggio, 1985). The p-value is on one-tailed tests.

In *NoRev* already 81% of promising Bs stick to their word, which leaves little scope for further increase by revelation. Still, in *Rev* the share is even higher with 85%. Although the effect goes in the predicted direction, the difference is not large enough to be significant (Z=0.534, p=0.270).

Result 5 is further supported by probit regressions of the decision to choose *Roll*, which are reported in Table 3.4. Here, we categorize messages into promises and no promises, where we categorize intentions as “no promise”, as Bs’ behavior after having sent an intention is not significantly different from the behavior after having sent an empty message (see Table 3.3 and Section 3.2.2). Column 1 of Table 3.4 reports the results of *Com*, Column 2 of *NoCom* and Column 3 (4) reports the results of a regression including both treatments with (without) controls.[19]

In all 4 regressions the dependent variable is B’s decision, represented by a dummy variable which takes the value 1 if B chooses *Roll* and 0 otherwise. Promise (NoPromise) is a dummy variable for sending a (no) promise in *Com*, *Rev* is a dummy for the condition

[19]Due to a lack of controls we excluded one session of *Com2*. Results for *Com* and *NoCom* do not change when excluding controls and/or including the excluded session. Results in Column 4 do not change when including this session either. Moreover, the results are robust to OLS regressions.

Table 3.4: Probit of Bs' decision to choose *Roll*

	Coefficient (p-value)			
	Com	*NoCom*	All	All
Promise	1.742		0.808	0.715
	(0.000)		(0.000)	(0.000)
Rev	-0.059	-0.322	-0.382	-0.424
	(0.846)	(0.157)	(0.142)	(0.105)
Promise* *Rev*	0.216		0.527	0.623
	(0.561)		(0.158)	(0.085)
NoPromise			-0.639	-0.535
			(0.116)	(0.233)
NoPromise* *Rev*			-0.264	-0.347
			(0.464)	(0.346)
Risk	0.097	0.205	0.132	
	(0.064)	(0.039)	(0.003)	
Female	0.486	-0.235	0.259	
	(0.022)	(0.339)	(0.150)	
# of observations	131	60	191	191
# of sessions	11	5	16	16
Pseudo R-squared	0.314	0.106	0.240	0.212
Log Pseudo Likelihood	-61.519	-36.908	-100.433	-104.052

The regressions cluster on sessions. The reference category is *NoRev*, or [*NoCom*|*NoRev*] respectively. The sample consists of all Bs in all sessions, except of one session of *Com2*, which we exclude due to a lack of controls. Results (in column 4) do not change if we include the session. Results for *Com* and *NoCom* (columns 1 and 2) do not change when excluding the controls.

Rev and Promise**Rev* (NoPromise**Rev*) is an interaction dummy of the two. In Column 1-3, we also include two controls, a measurement of risk and a female dummy.[20]

We observe that the probability to choose *Roll* is significantly higher if B sends a promise (p=0.000) than if he sends another message or does not communicate.[21] However, the coefficient of the interaction dummy Promise**Rev* is far away from being significant (p=0.561), indicating that the probability to choose *Roll* when having sent a promise is not further increased by revelation. Moreover, *Rev* does neither have a general significant effect in *Com* (p=0.846), nor in *NoCom* (p=0.157). As shown by the non-parametric test, if anything revelation without communication even leads to less cooperative behavior as the coefficient of *Rev* in Column 2 is negative and the p-value not far from being marginally significant.

In column 3 we report the results of the probit regression including both treatments. The reference category is a subject in $[NoCom|NoRev]$. In order to account for the different number of promises in the two conditions, we separate the subjects in *Com* into promisors and non-promisors and include a dummy for each group. Thus, NoPromise only takes the value 1 for Bs not promising in *Com* and it is 0 for subjects in *NoCom*. Altogether, we have 6 categories, with $[NoCom|NoRev]$ being the base case including dummies for all other cases. Similar to *Com*, we observe that Bs sending a promise have a higher probability to choose *Roll* (p=0.000). The coefficients of *Rev* and and NoPromise**Rev* are not significant, showing that revelation does not change behavior when no promise has been sent. The coefficient of Promise**Rev* is positive, but not significant (p=0.158). Still, it becomes marginally significant (p=0.085) when excluding the two control variables.[22] It seems that *Rev* marginally increases the probability of choosing *Roll* conditional on sending a promise, yet, the effect is very small and not robust. Thus, we are not able to prove our main hypothesis.

[20] We elicited risk preferences based on subjects' self-assessment on a scale from 0 to 10, with 0 indicating that a subject has a very weak willingness to take risks, while a score of 10 means that a subject has a strong willingness to take risks. Dohmen et al. (2011) show that this general risk question is a good predictor of actual risk-taking behavior.

[21] However, the causality is not clear. B might send a promise as he knows he will choose *Roll*, or he might choose *Roll* due to the promise sent.

[22] Results for column 1 and 2 do not change when excluding risk and female.

3.2.2 The Effect of Communication on Cooperation

One major reason for our effects being only marginally significant might be that in [*NoCom*|*NoRev*] already 81% of all Bs sending a promise stick to it, restricting the scope for further increase in promise keeping with revelation.

We started off with conducting treatment *Com1*, where Bs choose between a promise, a statement of intent, and an empty message, and observe a very high promise keeping rate even without revelation. In order to achieve more promise breaking in the baseline without revelation, we conducted a second communication treatment, *Com2*, allowing Bs to choose only between the same promise and two empty messages. Thereby, Bs attempting to influence As while planning to take the non-cooperative decision are forced to break a promise.[23] However, this change in the set of messages failed to generate a higher rate of promise breaking in the baseline, such that we are not able to confirm our main hypothesis in *Com2* either.[24] Yet, the comparison of *Com1* and *Com2* reveals some interesting findings about the the effect of communication on cooperation and the differences between promises and intentions, which we will address in this section.

Bs' Behavior and the Choice of Messages

Considering Bs' behavior, it turns out that the set of messages available to B highly influences the effectiveness of communication on cooperation. In the following analysis, we pool the data of *Rev* and *NoRev*, as there is no significant difference in Bs' behavior between both conditions.

Result 6. *While the share of Bs choosing Roll in Com2 is significantly higher than in NoCom, the share in Com1 is not.*

We conclude that the possibility to send an intention in *Com1* constrains the effectiveness of communication on cooperation. Figure 3.4 illustrates the shares of Bs choosing *Roll* in all three treatments.

[23] As are only informed that there are three messages to choose from, but they are unaware of the type of messages or the wording. This was complete information.

[24] Note that the unchanged communication behavior of Bs allows us to pool *Com1* and *Com2* to *Com* in the analysis of Section 3.2.1.

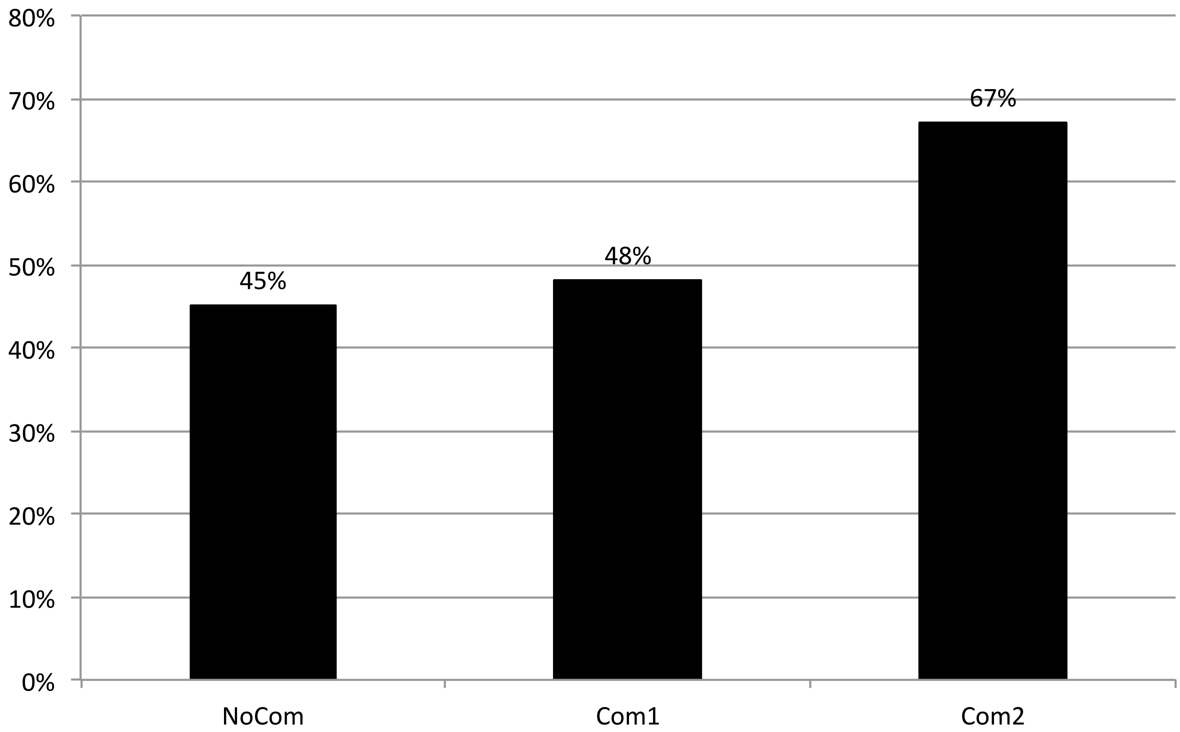

Figure 3.4: *Roll* rates of Bs.

In *Com2* 67% of Bs choose *Roll*, which is significantly higher than the share of 45% in *NoCom* (Z= 2.502, p=0.006), and than the share of 48% in *Com1* (Z=2.270, p=0.012). In contrast, Bs in *Com1* are as likely to cooperate as Bs in *NoCom* (Z=0.330, p=0.371).

In order to identify the driving forces behind these effects, we analyze the data separated by types of messages. Table 3.5 reports the shares of Bs sending each of the three types of message separately and the corresponding shares of Bs choosing *Roll*.

Result 7. *The share of Bs sending a promise is significantly higher in Com2 than in Com1, where roughly one quarter of Bs choose to send an intention. While the majority of promises is kept, the majority of intentions is broken.*

While 71% of Bs send a promise in *Com2*, only 48% do so in *Com1*, with the difference being highly significant (Z=2.794, p=0.003). In *Com1* 26% of Bs send an intention, which is not possible in *Com2*. In both treatments the majority of promises are kept, in *Com1* 77%, in *Com2* even 88%. In contrast, only 26% of Bs sending an intention stick to it. This share is significantly smaller than the share of Bs keeping their promise

Table 3.5: Overview of messages sent and subsequent behavior

Messages sent				
	Promise	Intention	Empty	Total
Com1	34/71 (48%)	19/71 (27%)	18/71 (25%)	71 (100%)
Com2	51/72 (71%)	–	21/72 (29%)	72 (100%)

Shares choosing *Roll*				
	Promise	Intention	Empty	Total
Com1	26/34 (77%)	5/19 (26%)	3/18 (17%)	34/71 (48%)
Com2	45/51 (88%)	–	3/21 (14%)	48/72 (67%)

The sample consists of all B-Persons in *Com1* and *Com2*.

(*Com1*: Z=3.554, p=0.000; *Com*: Z=5.083, p=0.000), but not significantly different from the share of cooperating Bs conditional on sending an empty message (*Com1*: Z=0.713, p=0.238; *Com*: Z=0.997, p=0.160).

Note that the share of promises being kept in *Com2* (88%) is even marginally significantly higher than in *Com1* (77%) (Z=1.433, p=0.076), although more Bs send a promise in *Com2* than in *Com1*. This observation, together with the fact that most intentions are broken, yields the following result.

Result 8. *Not sticking to an intention seems to be less costly than breaking a promise.*[25]

Result 8 can either be caused by the diction of the message itself or by the comparison to message alternatives, indicating that breaking an intention is not the strongest lie. However, such a difference in behavior when sending a promise compared to sending an intention does not seem to occur in CD (2006), who use free-form messages. Therefore, we suggest that the latter reason is more likely to explain the observed phenomenon. Subjects who send an intention might not think "I am indicating to my partner that I

[25] This result does not follow from the mere observation that promises are kept and intentions are broken in *Com1* as this might be caused by selection into messages (altruistic subjects send a promise and selfish subjects send an intention). However, the fact that a higher share of promises is sent *and* kept in *Com2* than in *Com1* yields the result.

will choose *Roll*", but more likely "I did not promise anything". Thus, unused alternatives seem to play a role, not as a signal to others, but as a self-justification device to behave selfish.[26] To conclude, the set of messages available to subjects in an experimental setting seem to play a crucial role for their behavior.

Finally, the fact that *Roll* rates after sending an intention are not significantly different from *Roll* rates after sending an empty message, but are significantly different from *Roll* rates after sending a promise in *Com1* speaks against the relevance of guilt aversion in our setting. B knows that A is not aware of the different available messages. Therefore, he should anticipate that both, a promise and an intention message, increase A's first-order belief in comparison to receiving an empty message. Guilt aversion would predict a more cooperative behavior upon sending an intention than upon sending an empty message and a similar behavior upon sending a promise or an intention.[27] In turn, promise keeping per se, as suggested by Vanberg (2008), is likely to play a role in our experiment, given that a high share of Bs stick to a given promise even if their action choice is not observable.

As' Behavior

As As do not know which messages can be sent by Bs, we pool *Com1* and *Com2* to *Com* for the analysis of As' behavior. We do not differentiate between conditions either, as As do neither know about the existence of two conditions nor does the experimental design vary across conditions from A's point of view.

Table 3.6 gives an overview of As' behavior in *Com* and in *NoCom*. For *Com* we report the overall behavior (total) and separated by the message received.

Result 9. *The share of As choosing In is increased by communication for all kinds of messages. Furthermore, As are equally more likely to cooperate after receiving a promise or an intention than after receiving an empty message.*

[26] It would be interesting to test whether the share of people sticking to an intention, if the only options are an intention or two empty messages, is similar to the share keeping their promise in *Com2*.

[27] We cannot directly test for a difference in second-order beliefs as we ask for averages. We can only compare second-order beliefs across *Com1* and *Com2*. These are not significantly different (MWU, 2-sided, p=0.995).

Table 3.6: As' behavior

	NoCom	*Com*			
		Total	Promise	Intention	Empty
In	21 / 35%	102 / 71%	67 / 79%	14 / 74%	21 / 54%
Out	39 / 65%	41 / 29%	18 / 21%	5 / 26%	18 / 46%
	60 / 100%	143 / 100%	85 / 100%	19 / 100%	39 / 100%

The sample consists of all As.

We observe a large effect of communication on As' behavior. The share of As choosing *In* increases significantly from only 35% in *NoCom* to 71% in *Com* (Z=4.833, p=0.000). This effect is driven by both promises and intentions. After receiving a promise, 79% of As choose *In*, and after receiving an intention 74% do so (Z=0.488, p=0.313). These two shares are (marginally) significantly higher than the share of As choosing *In* after receiving an empty message, which amounts to 54% (empty vs. intention: Z=1.450, p=0.074; empty vs. promise: Z=2.845 p=0.002).

Interestingly, the share of As choosing *In* after receiving an empty message is significantly different from the respective share in *NoCom* (54% vs. 35%, Z=1.854, p=0.032). It seems as if As receiving an empty message might not have considered the possibility of a promise or an intention, and react to a friendly, though meaningless message.

The difference in As' behavior across treatments is reflected by their first-order beliefs about Bs' behavior. While without communication As believe that on average 45% of Bs choose *Roll*, this belief amounts to 58% with communication (MWU, 2-sided, p=0.001).[28] Hence, similar to CD (2006), we observe that communication increases As' first-order beliefs, thus enhances trust among As.

[28]In particular, the average first-order belief is 63% conditional on receiving a promise, 57% conditional on receiving an intention and 47% conditional on receiving an empty message. The first-order belief is significantly higher after receiving a promise than after receiving an empty message (MWU, 2-sided, p=0.001), however not significantly higher than after receiving an intention (MWU, 2-sided, p=0.179).

Does Communication Enhance Mutual Cooperation?

We observe a significant increase of mutual cooperation in *Com* compared to *NoCom*, represented by the share of pairs choosing (*In, Roll*) in each treatment (45% vs. 13%, Z=4.270, p=0.000). These shares, reported by type of message sent, are stated in Table 3.7.

Table 3.7: Shares of pairs choosing (*In, Roll*)

NoCom	*Com*			
	Total	Promise	Intention	Empty
8/60	64/143	57/85	3/19	4/39
13.3%	44.8%	67.1%	15.8%	10.3%

Result 10. *Communication increases mutual cooperation. However, while promises increase the share of pairs choosing (In, Roll), intentions do not.*

While promises lead to a very high cooperation rate, intentions do not. This difference is mainly driven by the fact that Bs keep their promises, but break their intentions, while As trust both.[29] We conclude that the set of messages available to B plays a crucial role for the effectiveness of communication in experimental settings.

3.3 Comparison to Previous Research

The present experimental design is based on the work by Charness and Dufwenberg (2006), who analyze the effect of free-form communication on cooperation. While their design informs A only about her payoffs, we vary the revelation of B's action choice in order to test for social image concerns. However, if we restrict our data to the *NoRev* condition,

[29] This might have been different if As had been aware of the messages available to B (compare Charness and Dufwenberg, 2010).

we find largely different results compared to CD (2006). In this section, we therefore analyze these discrepancies to the work by Charness and Dufwenberg,[30] incorporating their follow-up treatment with predefined messages (Charness and Dufwenberg, 2008 and 2010; henceforth CD (2010)). In CD (2010), Bs could choose between sending a sheet saying "I promise to choose *Roll*." or an empty sheet, which is closest to our *Com2* treatment without revelation. Apart from the communication protocol, our design differs from CD (2006) and CD (2010) only in B's relative payoff for choosing *Don't Roll*, which we slightly increased in order to reduce *Roll* rates without communication (see Section 3.2.2).[31]

Considering Bs' *Roll* rates, we do not find any difference between *NoCom* (53%) and *Com* (51%) in the *NoRev* condition (Z=0.179, p=0.429) in our sample. Though slightly more Bs cooperate if we restrict the sample to *Com2* (61%), there is still no significant difference to *NoCom* (Z=0.637, p=0.262).[32] Similarly, communication fails to significantly influence Bs' behavior in CD (2010) either. While in their experiment the average *Roll* rate increases from 44% without communication to 58% allowing for predefined messages, this difference is only marginally significant on a one-tailed test (Z=1.339, p=0.090).[33]

In contrast, Bs in CD (2006) are significantly more likely to choose *Roll* after free-form communication than without communication (44% vs. 67%, Z=2.083, p=0.019). At first glance, this indicates that Bs feel more committed to a free-form promise than to a predefined one, yielding an increase in *Roll* rates in CD (2006). However, while in CD (2006) 57% of Bs send a promise in the communication treatment, we only observe 51% in *Com* and 39% in *Com1*, the latter difference being almost marginally significant on a two-tailed test (*Com*: 51% vs. 57%, Z=0.594, p=0.552; *Com1*: 39% vs. 57%, Z=1.608, p=0.108,

[30]More precisely, we only use the (5,5) treatment for comparison as it reflects our payoff structure.

[31]Furthermore, Charness and Dufwenberg (2006 and 2010) conduct a pen-and-pencil experiment in the classroom while we use the laboratory and computer screens. However, as we can not identify any idiosyncratic effect of this design feature, we neglect it in the following analysis.

[32]Note that the difference was significant pooling *Rev* and *NoRev* (Section 3.2.2), but we restrict the sample to *NoRev* here.

[33]Results considering the whole sample in the communication treatment are only reported in CD (2008).

two-tailed tests).[34] Hence, the higher promise rate in CD (2006) might also account for part of the increased effect on cooperation.

It is striking that, though CD (2010) find that 85% of all Bs send a promise, which differs statistically from our promise rate in *Com2* (64%, Z=2.293, p=0.022, two-tailed test), their effect of communication on cooperation is only marginally significant. Compared to our result, it seems that promises in CD (2010) induce less commitment among Bs. Indeed, while in *Com2* 87% of all Bs who send a promise keep it, this share is significantly lower in CD (2010) (61%, Z=2.183, p=0.029, two-tailed test).[35] This might be due to the fact that the messages available to Bs are common knowledge in their design, while we leave As unaware of message choices, yielding many Bs to send a promise just in order to avoid the mistrusting signal of an empty sheet.

As to As' behavior, we do not find any evidence that predefined messages in our design dampen cooperation compared to free-form communication. In fact, *In* rates among As in our experiment achieve similar levels as in CD (2006) with free-form messages (71% in *Com* vs. 74% in CD (2006), Z=0.341). Furthermore, as cooperation among As is relatively low without communication in our setting (33%),[36] we observe a highly significant effect of communication on As' behavior (71% in *Com*, Z=3.520, p=0.000), exceeding the effect with free-form messages in CD (2006) (56% without communication vs. 74% with communication, Z=1.777, p=0.038). In contrast, predefined messages in CD (2010) do not induce As to choose *In* more often, if at all, *In* rates decrease (56% without communication vs. 52% with communication, Z=0.336).[37]

While this finding seems to be unintuitive at first sight, it shows that besides differentiating between free-form and predefined messages, subtle design differences can account for huge changes in the credibility of messages. First, while Bs in our experiment choose an empty

[34] Though in CD (2006), also intentions were classified as promises, we exclude intentions in *Com1* from the comparison. This is reasonable as Bs in our experiment break intentions more often than promises, and behave similarly after sending an intention as after sending an empty message (see Section 3.2.2).

[35] In contrast, the promise keeping rate in *Com* (81%) is similar to the one in CD (2006) (75%, Z=0.567, p=0.571, two-tailed test).

[36] Note that we restrict the sample to *NoRev* only.

[37] There is no effect despite the higher promise rate in CD (2010).

message if no promise is made (and can not refuse to send a message), the only alternative to a promise in CD (2010) is an empty sheet. It might thus be the case that empty talk in our experiment, though through predefined messages, contains some general pleasantry, thus inducing As to cooperate more often in the present setting compared to CD (2010). Second, the explicit announcement in CD (2010) that promises are not binding might create a social norm reducing both self and social image concerns for non-cooperation among Bs, which in turn might be anticipated by As. Finally, As in our experiment are not aware of the kind of possible messages, while the exact wording and procedure is common knowledge in CD (2010). As an empty message thus signals uncooperative behavior by B in their setting and might induce As to opt out of the game, it is likely that some Bs in CD (2010) send a promise who would not have done so in other circumstances. If As anticipate this cheap-talk nature of promises, the credibility of a promise is reduced, which is why As seem to trust less in CD (2010) than in our setting. The fact that communication has a larger influence on Bs in our setting than with free-form messages in CD (2006) can only be explained by the strong wording of our predefined promise, as compared to the diverse statements of intent in CD (2006).

To summarize, Bs in our experiment as well as in CD (2010) do not seem to be influenced by communication, while in CD (2006), free-form messages increase *Roll* rates. In contrast, *In* rates in our setting highly increase with communication, with this effect being even stronger than in CD (2006), while messages do not influence As' behavior in CD (2010). Hence, starting from a slightly lower cooperation level without communication than CD (2006), we obtain a similar effect of communication on (*In, Roll*) rates, which is also highly significant (13% in *NoCom* vs. 40% in *Com*, 50% in *Com2*, $p<0.01$ in both cases, two-tailed test). In general, while messages are most influential when they are free-form, predefined messages have a larger impact in our experiment than in CD (2010). This might be due to very subtle changes in the communication protocol, such as A's unawareness of message wording or the possibility of empty talk. We conclude that the effect of communication is not robust to slight changes in the experimental design.

3.4 Conclusion

Non-binding communication is at the heart of many economic interactions, especially if cooperation cannot be contractually enforced, for example because writing fully contingent contracts is impossible or too costly, or because cooperation is not verifiable. Hence, we contribute to the literature exploring why and in which environments "cheap talk" can be influential in two-player trust games.

In this chapter, we experimentally analyze whether individuals stick to their promised action, in contrast to the rational prediction, due to the aversion of being perceived as a promise breaker. While we observe slightly more cooperation of the promising party if the receiver of the promise can observe its compliance, the results are not significant. We find that 81% of subjects stick to their promise even if their action is not observable to their interaction partners.[38] On the one hand, this result limits the scope for a further increase in cooperation with revelation. On the other hand, it highlights subjects' preference for promise keeping per se (Vanberg, 2008), which in our experiment seems to play a more important role than social image concerns. Still, we find that the preference for sticking to one's word does only exist for promises and not for statements of intent. While most of the promises are kept, statements of intent tend to be broken. In line with this result, we find that the set of available predefined messages yields different results regarding cooperation by the communicating party, the second mover. While the possibility to communicate increases cooperation by second movers if they can only choose between sending a promise or an empty message, communication has no effect on second movers' behavior if they have the additional option of sending a statement of intent. However, the receivers of messages trust both a promise and a statement of intent in the same way. This finding allows us to exclude guilt aversion as an explanation for promise keeping, as the communicating party seems to be aware that a statement of intent does influence his partner the same way as a promise, but still does not stick to it.

[38]This even exceeds the shares reported in CD (2006) and Vanberg (2008).

To the best of our knowledge, our study belongs to one of the first economic studies analyzing the combined effect of communication and social image concerns on cooperation, suggesting a high potential and the need for further research. While we fail to prove the existence of social image concerns in our anonymous experimental set-up, one should not transfer this finding to other settings. We rather want to point out the crucial role of the design of the experiment, when trying to identify such subtle behavioral patterns. Lifting anonymity (see e.g. Tadelis, 2011) might increase the relevance of social image concerns, just like repeating the game and allowing for reputation building.

A3 Appendix

Results separated for ***Com1*** and ***Com2***

In Table A3.1 we report Bs' *Roll* rates in *Com1*, *Com2*, *Com* and *NoCom*, as well as the corresponding Z-Statistics and p-values. Table A3.2 provides an overview of the messages sent in *Com1*, *Com2* and *Com*. Table A3.3 reports the *Roll* rates for each type of message sent in both conditions.

Table A3.1: Bs' average *Roll* rate by treatment and condition

		Treatment				Z Stat.
		Com1	*Com2*	***Com***	***NoCom***	(p-value)
Condition	***Rev***	19/35	26/36	**45/71**	**11/30**	2.468
		54%	72%	**63%**	**37%**	(0.007)
	NoRev	15/36	22/36	**37/72**	**16/30**	-0.179
		42%	61%	**51%**	**53%**	(0.429)
Z Stat.		1.062	1.000	1.450	-1.292	
(p-value)		(0.144)	(0.159)	(0.074)	(0.098)	

The Z Stat. reflects the test of proportions (see Glasnapp and Poggio, 1985). The p-value is on one-tailed tests. The statistics in the last column test for the difference between *Com* treatment and *NoCom*.

Table A3.2: An overview of messages sent, detailed

		Promise	Intention	Empty
Com1	*Rev*	20/35	6/35	9/35
		57%	17%	26%
	NoRev	14/36	13/36	9/36
		39%	36%	25%
	Z stat.	1.540	-1.805	0.069
	(p-value)	(0.062)	(0.036)	(0.472)
Com2	*Rev*	28/36	–	8/36
		78%	–	22%
	NoRev	23/36	–	13/36
		64%	–	36%
	Z stat.	1.296	–	1.296
	(p-value)	(0.097)	–	(0.097)
Com	*Rev*	48/71	6/71	17/71
		68%	9%	24%
	NoRev	37/72	13/72	22/72
		51%	18%	31%
	Z stat.	1.975	-1.692	0.888
	(p-value)	(0.024)	(0.045)	(0.187)

The Z Stat reflects the test of proportions for the two treatments or conditions (see Glasnapp and Poggio, 1985). The p-value is on one-tailed tests.

Table A3.3: *Roll* rates by type of message sent, detailed

		Promise	Intention	Empty
Com1	*Rev*	16/20	1/6	2/9
		80%	17%	22%
	NoRev	10/14	4/13	1/9
		71%	31%	11%
	Z stat.	0.580	-0.650	0.633
	(p-value)	(0.281)	(0.258)	(0.264)
Com2	*Rev*	25/28	–	1/8
		89%	–	13%
	NoRev	20/23	–	2/13
		87%	–	15%
	Z stat.	0.257	–	-0.183
	(p-value)	(0.400)	–	(0.427)
Com	*Rev*	41/48	1/6	3/17
		85%	17%	18%
	NoRev	30/37	4/13	3/22
		81%	31%	14%
	Z stat.	0.534	-0.650	0.344
	(p-value)	(0.270)	(0.258)	(0.365)

The Z Stat reflects the test of proportions for the two treatments or conditions (see Glasnapp and Poggio, 1985). The p-value is on one-tailed tests.

General Instructions[39]

We welcome you to this experiment. Please read these instructions carefully and follow the instructions on your screen after the start of the experiment.

At the end of the experiment you will get paid according to your decisions and the decisions of the other participants, as described below. In addition, you will get a fixed payment of 4 euros for your attendance.

During the whole experiment you are not allowed to talk to other participants, to use mobile phones, or to start other programs on your computer. If you disobey these rules, we have to exclude you from the experiment and from all payments. If you have any questions, please raise your hand. An experimenter will come to your seat to answer your questions.

During the experiment, we are not talking about euros but about points. Your payment will be calculated in points. At the end of the experiment your overall score will be converted to euro, where

1 Point = 25 euro cents.

The experiment consists of two parts and a questionnaire. Part 1 will be explained below. Once all participants have finished Part 1, you will get the instructions for Part 2. A questionnaire follows after Part 2.

[39]Original instructions were in German and are available upon request. Passages occurring only in the communication treatments are indicated by [...].

Instructions Part 1

At the start of the experiment, either role A or role B will be assigned randomly to each participant. You will be informed on your screen which role was assigned to you. One person A and one person B, respectively, form an interaction pair. The allocation is random and anonymous. No participant will get to know the identity of his partner during or after the experiment. Your payment in Part 1 depends on the decisions made within your interaction pair.

Decisions:

Each person A chooses between IN and OUT. If A chooses OUT, A and B get 10 points each. If person A chooses IN, the payments depend on B's decision. Every person B chooses between ROLL THE DIE and DON'T ROLL THE DIE. At the time of decision, Person B doesn't know whether A has chosen IN or OUT. But as B's decision is only relevant if A chose IN, every person B should make her decision under the assumption that A has chosen IN.

If A chose IN and B chooses DON'T ROLL THE DIE, B gets 30 points and A 0 points. If A chose IN and B chooses ROLL THE DIE, B gets 20 points and rolls a die at the end of the experiment in order to determine A's payoff. If the die shows 1, A gets 0 points, if the die shows 2,3,4,5 or 6, A gets 24 points.

The following table summarizes the payments, depending on the decisions made within an interaction pair and the result of rolling the die.

Decisions	**Payoff A**	**Payoff B**
A chooses OUT	10	10
A chooses IN, B chooses DON'T ROLL THE DIE	0	30
A chooses IN, B chooses ROLL THE DIE, Die=1	0	20
A chooses IN, B chooses ROLL THE DIE, Die=2,3,4,5,6	24	20

Please note: Every participant with role B, regardless if she chose ROLL THE DIE or DON'T ROLL THE DIE, will roll a die at the end of the experiment, such that rolling the die won't reveal the decision made by B. The result of rolling the die however is only relevant for those interaction pairs where A chose IN and B chose ROLL THE DIE.

[**Message:**

Before A and B make their decision, B has the opportunity to choose one of three predefined messages and send it to A.]

Bonus questions:

During the experiment every participant has the opportunity to earn extra points by answering bonus questions correctly. The earnings out of these bonus questions will be displayed separately at the end of the experiment. You will get more detailed information during the experiment.

Control questions:

Before the start of the experiment control questions will appear on your screen to check that you understood the instructions. When all participants have answered these questions correctly, Part 1 of the experiment starts.

Bibliography

ACEMOGLU, D., PISCHKE, J.-S., 1999. The Structure of Wages and Investment in General Training. *Journal of Political Economy* 107(3), pp. 539-572.

AGHION, P., REY, P., DEWATRIPONT, M., 2004. Transferable Control. *Journal of the European Economic Association* 2(1), pp. 115-138.

AGHION, P., TIROLE, J. 1997. Formal and Real Authority in Organizations. *Journal of Political Economy* 105(1), pp. 1-29.

ALONSO, R., DESSEIN, W., MATOUSCHEK, N., 2008. When Does Coordination Require Centralization? *The American Economic Review* 98(1), pp. 145-179.

ALONSO, R., DESSEIN, W., MATOUSCHEK, N., 2010. Strategic Communication: Prices versus Quantities. *Journal of the European Economic Association* 8 (2/3), pp. 365-376.

ALONSO, R., MATOUSCHEK, N., 2007. Relational Delegation. *The RAND Journal of Economics* 38(4), pp. 1070-1089.

ALONSO, R., MATOUSCHEK, N., 2008. Optimal Delegation. *The Review of Economic Studies* 75(1), pp. 259-293.

ANDREONI, J., BERNHEIM, B., 2009. Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica* 77(5), pp. 1607-1636.

Bibliography

Andreoni, J., Rao, J. M., 2011. The Power of Asking: How Communication Affects Selfishness, Empathy, and Altruism. *Journal of Public Economics* 95(7-8), pp. 513-520.

Apsler, R., 1975. Effects of Embarrassment on Behavior Toward Others. *Journal of Personality and Social Psychology* 32(1), pp. 145-153.

Ariely, D., Bracha, A., Meier, S., 2009. Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *The American Economic Review* 99(1), pp. 544-555.

Armstrong, M., 1995. Delegation and Discretion. MPRA Paper No. 17069. Available at: http://mpra.ub.uni-muenchen.de/17069/.

Arrow, K.J., 1974. The Limits of Organization. *New York: W.W. Norton.*

Baker, G., Gibbons, R., Murphy, K.J., 1994. Subjective Performance Measures in Optimal Incentive Contracts. *The Quarterly Journal of Economics* 109(4), pp. 1125-1156.

Batigalli, P., Dufwenberg, M., 2007. Guilt in Games. *The American Economic Review* 97(2), pp. 170-176.

Battaglini, M., 2002. Multiple Referrals and Multidimensional Cheap Talk. *Econometrica* 70(4), pp. 1379-1401.

Beck A., Kerschbamer, R., Qiu, J., Sutter, M., 2013. Shaping Beliefs in Experimental Markets for Expert Services: Guilt Aversion and the Impact of Promises and Money-Burning Options. *Games and Economic Behavior* 81, pp. 145-164.

Benabou, R., Laroque, G., 1992. Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility. *The Quarterly Journal of Economics* 107(3), pp. 921-958.

Bester, H., Krähmer, D., 2008. Delegation and Incentives. *The RAND Journal of Economics* 39(3), pp. 664-682.

Blanes i Vidal, J., 2007. Delegation of Decision Rights and the Winner's Curse. *Economics Letters* 94(2), pp. 163-169.

Blume, A., Board, O.J., Kawamura, K., 2007. Noisy Talk. *Theoretical Economics* 2(4), pp. 395-440.

Blume, A., Ortmann, A., 2007. The Effects of Costless Pre-Play Communication: Experimental Evidence from Games with Pareto-Ranked Equilibria. *Journal of Economic Theory* 132(1), pp. 274-290.

Bochet, O., Putterman, L., 2009. Not Just Babble: Opening the Black Box of Communication in a Voluntary Contribution Experiment. *European Economic Review* 53(3), pp. 309-326.

Bracht, J., Regner, T., 2011. Moral Emotions and Partnership. Jena economic research papers No. 2011-028. Available at: http://hdl.handle.net/10419/56830.

Broberg, T., Ellingsen, T., Johannesson, M., 2007. Is Generosity Involuntary? *Economics Letters* 94(1), pp. 32-37.

Camera, G., Casari, M., Bigoni, M., 2011. Communication, Commitment, and Deception in Social Dilemmas: Experimental Evidence. Quaderni DSE Working Paper No. 751. Available at SSRN: http://ssrn.com/abstract=1854132 or http://dx.doi.org/10.2139/ssrn.1854132.

Carpenter, J., Myers, C.K., 2010. Why Volunteer? Evidence on the Role of Altruism, Image, and Incentives. *Journal of Public Economics* 94(11), pp. 911-920.

Chakraborty, A., Harbaugh, R., 2007. Comparative Cheap Talk. *Journal of Economic Theory* 132(1), pp. 70-94.

Chakraborty, A., Harbaugh, R., 2010. Persuasion by Cheap Talk. *The American Economic Review* 100(5), pp. 2361-2382.

Charness, G., Dufwenberg, M., 2006. Promises and Partnership. *Econometrica* 74(6), pp. 1579-1601.

Charness, G., Dufwenberg, M., 2008. Broken Promises: An Experiment. UC Santa Barbara, Economics Working Paper Series No. qt6836m74q. Available at: http://econ.econ.ucsb.edu/papers/wp10-08.pdf.

Charness, G., Dufwenberg, M., 2010. Bare Promises: An Experiment. *Economics Letters* 107(2), pp. 281-283.

Chen, C.X., Lu, H., Sougiannis, T., 2008. Managerial Empire Building, Corporate Governance, and the Asymmetrical Behavior of Selling, General, and Administrative Costs. AAA 2008 Financial Accounting and Reporting Section (FARS) Paper; (CAAA) 2008 Annual Conference Paper. Available at SSRN: http://ssrn.com/abstract=1014088 or http://dx.doi.org/10.2139/ssrn.1014088.

Cho, I.-K., Kreps, D., 1987. Signaling Games and Stable Equilibria. *The Quarterly Journal of Economics* 102(2), pp. 179-221.

Cleveland, J.N., Murphy, K.R., Williams, R.E., 1989. Multiple Uses of Performance Appraisal: Prevalence and Correlates. *Journal of Applied Psychology* 74(1), pp. 130-135.

Coase, R.H., 1937. The Nature of the Firm. *Economica* 4(16), pp. 386-405.

Cooper, R., DeJong, D.V., Forsythe, R., Ross, T.W., 1992. Communication in Coordination Games. *The Quarterly Journal of Economics* 107(2), pp. 739-771.

Crawford, V., Sobel, J., 1982. Strategic Information Transmission. *Econometrica* 50(6), pp. 1431-1451.

Cripps, M.W., Mailath, G.J., Samuelson, L., 2004. Imperfect Monitoring and Impermanent Reputations. *Econometrica* 72(2), pp. 407-432.

Cripps, M.W., Mailath, G.J., Samuelson, L., 2007. Disappearing Private Reputations in Long-Run Relationships. *Journal of Economic Theory* 134(1), pp. 287-316.

Dana, J., Cain, D., Dawes, R., 2006. What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games. *Organizational Behavior and Human Decision Processes* 100(2), pp. 193-201.

Dana, J., Weber, R., Kuang, J., 2007. Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness. *Economic Theory* 33(1), pp. 67-80.

DellaVigna, S., List, J.A., Malmendier, U., 2012. Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics* 127(1), pp. 1-56.

Dessein, W., 2002. Authority and Communication in Organizations. *The Review of Economic Studies* 69(4), pp. 811-838.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences. *Journal of the European Economic Association* 9(3), pp. 522-550.

Du, J., Leung, C.K.Y., Chu, D., 2013. Return Enhancing, Cash-rich or Simply Empire-building? An Empirical Investigation of Corporate Real Estate Holdings. MPRA Paper No. 44253. Available at http://mpra.ub.uni-muenchen.de/44253/.

Ellingsen, T., Johannesson, M., 2004. Promises, Threats and Fairness. *The Economic Journal* 114(495), pp. 397-420.

Ellingsen, T., Johannesson, M., Tjøtta, S., Torsvik, G., 2010. Testing Guilt Aversion. *Games and Economic Behavior* 68(1), pp. 95-107.

Ellingsen, T., Östling, R., 2010. When Does Communication Improve Coordination? *The American Economic Review* 100(4), pp. 1695-1724.

Englmaier, F., Filipi, A., Singh, R., 2010. Incentives, Reputation and the Allocation of Authority. *Journal of Economic Behavior & Organization* 76(2), pp. 413-427.

Farrell, J., Gibbons, R., 1989. Cheap Talk Can Matter in Bargaining. *Journal of Economic Theory* 48(1), pp. 221-237.

Fischbacher, U., 2007. Z-Tree: Zurich Toolbox for Readymade Economic Experiments. *Experimental Economics* 10(2), pp. 171-178.

Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in Disguise. An Experimental Study on Cheating. *Journal of the European Economic Association* 11(3), pp. 525-547.

Frankel, A., 2010. Aligned Delegation. NYU Working Paper. Available at: http://www. econ.as.nyu.edu/docs/IO/18572/Frankel_20110126.pdf.

Fuchs, W., 2007. Contracting with Repeated Moral Hazard and Private Evaluations. *The American Economic Review* 97(4), pp. 1432-1448.

Fudenberg, D., Levine, D.K., 1989. Reputation and Equilibrium Selection in Games with a Patient Player. *Econometrica* 57(4), pp. 759-778.

Fudenberg, D., Levine, D.K., 1992. Maintaining a Reputation when Strategies Are Imperfectly Observed. *The Review of Economic Studies* 59(3), pp. 561-579.

Glasnapp, D., Poggio, J., 1985. *Essentials of Statistical Analysis for the Behavioral Sciences.* CE Merrill Publishing Company.

Goltsman, M., Hörner, J., Pavlov, G., Squintani, F., 2009. Mediation, Arbitration and Negotiation. *Journal of Economic Theory* 144(4), pp. 1397-1420.

Goltsman, M., Mukherjee, A., 2011. Interim Performance Feedback in Multistage Tournaments: The Optimality of Partial Disclosure. *Journal of Labour Economics* 29(2), pp. 229-265.

Greiner, B., 2004. An Online Recruitment System for Economic Experiments. *Forschung und wissenschaftliches Rechnen 2003*, GWDG Bericht 63. Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen, pp. 79-93.

Grossman, Z., 2010a. Self-Signaling Versus Social-Signaling in Giving. UC Santa Barbara, Department of Economics. Available at: http://escholarship.org/uc/item/7320x2cp.

Grossman, Z., 2010b. Strategic Ignorance and the Robustness of Social Preferences. UC Santa Barbara, Department of Economics. Available at: http://escholarship.org/uc/item/60b93868.

Grossman, S.J., Hart, O.D., 1986. The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *The Journal of Political Economy* 94(4), pp. 691-719.

Güth, W., Huck, S., Ockenfels, P., 1996. Two-Level Ultimatum Bargaining with Incomplete Information: An Experimental Study. *The Economic Journal* 106(436), pp. 593-604.

Hart, O., Moore, J., 1990. Property Rights and the Nature of the Firm. *Journal of Political Economy* 98(6), pp. 1119-1158.

Hayek, F.A., 1945. The Use of Knowledge in Society. *The American Economic Review* 35(4), pp. 519-530.

Holmström, B., 1977. On Incentives and Control in Organizations. Ph.D. Dissertation, Graduate School of Business, Stanford University.

Holmström, B., 1984. On the Theory of Delegation. In M. Boyer and R. Kihlstrom, eds., *Bayesian Models in Economic Theory.* New York: North-Holland.

Holmström, B., 1999. Managerial Incentive Problems: A Dynamic Perspective. *The Review of Economic Studies* 66(1), pp. 169-182.

Hope, O.-K., Thomas, W.B., 2008. Managerial Empire Building and Firm Disclosure. *Journal of Accounting Research* 46(3), pp. 591-626.

Hurkens, S., Kartik, N., 2009. Would I Lie to You? On Social Preferences and Lying Aversion. *Experimental Economics* 12(2), pp. 180-192.

Ismayilov, H., Potters, J.M., 2012. Promises as Commitments. Netspar Discussion Paper No. 07/2012-057. Available at SSRN: http://ssrn.com/abstract=2209662.

Jensen, M., 1986. Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *The American Economic Review* 76(2), pp. 323-329.

Jensen, M., Meckling, W., 2012. Specific and General Knowledge, and Organizational Structure. In: Paul S. Myers, *Knowledge Management and Organisational Design.*. Routledge, pp. 17-38.

Kartik, N., Ottaviani, M., Squintani, F., 2007. Credulity, Lies, and Costly Talk. *Journal of Economic Theory* 134(1), pp. 93-116.

Kerr, N.L., Kaufman-Gilliland, M.K., 1994. Communication, Commitment, and Cooperation in Social Dilemmas. *Journal of Personality and Social Psychology* 66(3), pp. 513-529.

Knight, F.H., 1921. Risk, Uncertainty and Profit. *New York: Hart, Schaffner and Marx.*

Koch, A.K., Normann, H.-T., 2008. Giving in Dictator Games: Regard for Others or Regard by Others? *Southern Economic Journal* 75(1), pp. 223-231.

Koessler, F., Martimort, D., 2012. Optimal Delegation with Multi-Dimensional Decisions. *Journal of Economic Theory* 147(5), pp. 1850-1881.

Kováč, E., Krähmer, D., 2012. Optimal Sequential Delegation. Working Paper. Available at: http://www.wiwi.uni-bonn.de/kraehmer/2013-04-11_EK_SeqDel.pdf.

Kováč, E., Mylovanov, T., 2009. Stochastic Mechanisms in Settings Without Monetary Transfers: The Regular Case. *Journal of Economic Theory* 144(4), pp. 1373-1395.

Kreps, D.M., Wilson, R., 1982. Reputation and Imperfect Information. *Journal of Economic Theory* 27(2), pp. 253-279.

Krishna, V., Morgan, J., 2004. The Art of Conversation: Eliciting Information from Experts through Multi-Stage Communication. *Journal of Economic Theory* 117(2), pp. 147-179.

Kriss, P.H., Blume, A., Weber, R.A., 2011. Coordination, Efficiency and Pre-Play Communication with Forgone Costly Messages. University of Zurich Department of Economics Working Paper No. 34. Available at SSRN: http://ssrn.com/abstract=1957083 or http://dx.doi.org/10.2139/ssrn.1957083.

Lacetera, N., Macis, M., 2010. Social Image Concerns and Prosocial Behavior: Field Evidence from a Nonlinear Incentive Scheme. *Journal of Economic Behavior & Organization* 76(2), pp. 225-237.

Larson, T., Capra, C.M., 2009. Exploiting Moral Wiggle Room: Illusory Preference for Fairness? A Comment. *Judgment and Decision Making* 4(6), pp. 467-474.

Levinson, D.J., 2005. Empire-Building Government in Constitutional Law. *Harvard Law Review* 118(3), pp. 915-972.

Levy, G., 2005. Careerist Judges and the Appeals Process. *The RAND Journal of Economics* 36(2), pp. 275-297.

Lewis, M., 1995. Embarrassment: The emotion of self-exposure and evaluation. Tangney, June Price (Ed); Fischer, Kurt W. (Ed). *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride.* New York, NY, US: Guilford Press, xvii, pp. 198-218.

Linardi, S., McConnell, M.A., 2008. Volunteering and Image Concerns. California Institute of Technology, Social Science Working Paper No. 1282. Available at: http://www.wordsmatter.caltech.edu/SSPapers/sswp1282.pdf.

Loomis, J.L., 1959. Communication, the Development of Trust, and Cooperative Behavior. *Human Relations* 12(4), pp. 305-315.

Lundquist, T., Ellingsen, T., Gribbe, E., Johannesson, M., 2009. The Aversion to Lying. *Journal of Economic Behavior & Organization* 70(1), pp. 81-92.

MacLeod, W.B., Malcomson, J.M., 1989. Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment. *Econometrica* 57(2), pp. 447-480.

Mailath, G.J., Samuelson, L., 2001. Who Wants a Good Reputation? *The Review of Economic Studies* 68(2), pp. 415-441.

Mailath, G.J., Samuelson, L., 2006. *Repeated Games and Reputations: Long-Run Relationships.* Oxford University Press.

Mas-Colell, A., Whinston, M., Green, J., 1995. *Microeconomic Theory.* Oxford University Press.

Mazar, N., Amir, O., Ariely, D., 2008. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* 45(6), pp. 633-644.

Melumad, N.D., Shibano, T., 1991. Communication in Settings with No Transfers. *The RAND Journal of Economics* 22(2), pp. 173-198.

Milgrom, P., Roberts, J., 1982. Predation, Reputation, and Entry Deterrence. *Journal of Economic Theory* 27(2), pp. 280-312.

Mohlin, E., Johannesson, M., 2008. Communication: Content or Relationship? *Journal of Economic Behavior & Organization* 65(3), pp. 409-419.

Morris, S., 2001. Political Correctness. *Journal of Political Economy* 109(2), pp. 231-265.

Mortal, S., Reisel, N., 2012. Capital Allocation by Public and Private Firms. *Journal of Financial and Quantitative Analysis*, forthcoming.

Mukherjee, A., 2008. Career Concerns, Matching, and Optimal Disclosure Policy. *International Economic Review* 49(4), pp. 1211-1250.

Olszewski, W., 2004. Informal Communication. *Journal of Economic Theory* 117(2), pp. 180-200.

Ottaviani, M., Sørensen, P., 2006a. Professional Advice. *Journal of Economic Theory* 126(1), pp 120-142.

Ottaviani, M., Sørensen, P., 2006b. Reputational Cheap Talk. *The RAND Journal of Economics* 37(1), pp. 155-175.

Prat, A., 2005. The Wrong Kind of Transparency. *The American Economic Review* 95(3), pp. 862-877.

Prendergast, C., Topel, R.H., 1996. Favoritism in Organizations. *Journal of Political Economy* 104(5), pp. 958-978.

Sally, D., 1995. Conversation and Cooperation in Social Dilemmas. A Meta-Analysis of Experiments from 1958 to 1992. *Rationality and society* 7(1), pp. 58-92.

Scheff, T.J., 1988. Shame and Conformity: The Deference-Emotion System. *The American Sociological Review* 53(3), pp. 395-406.

Schmidt, K.M., 1993. Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests. *Econometrica* 61(2), pp. 325-351.

Shin, H.-H., Kim, Y.H., 2002. Agency Costs and Efficiency of Business Capital Investment: Evidence from Quarterly Capital Expenditures. *Journal of Corporate Finance* 8(2), pp. 139-158.

Simon, H.A., 1951. A Formal Theory of the Employment Relationship. *Econometrica* 19(3), pp. 293-305.

Smith, R.H., Webster, J.M., Parrott, W.G., Eyre, H.L., 2002. The Role of Public Exposure in Moral and Nonmoral Shame and Guilt. *Journal of Personality and Social Psychology* 83(1), pp. 138-159.

Sobel, J., 1985. A Theory of Credibility. *The Review of Economic Studies* 52(4), pp. 557-573.

Spence, M., 1973. Job Market Signaling. *The Quarterly Journal of Economics* 87(3), pp. 355-374.

Tadelis, S., 2011. The Power of Shame and the Rationality of Trust. Working Paper. Available at: http://faculty.haas.berkeley.edu/stadelis/working.htm.

Tangney, J.P., 1995. Shame and Guilt in Interpersonal Relationships. Tangney, June Price (Ed); Fischer, Kurt W. (Ed). *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride.* New York, NY, US: Guilford Press, xvii, pp. 114-139.

Vanberg, C., 2008. Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica* 76(6), pp. 1467-1480.

Williamson, O., 1975. Markets and Hierarchies: Analysis and Antitrust Implications. *New York: Free Press.*

Zabojnik, J., 2011. Subjective Evaluations with Performance Feedback. Queen's Economics Department Working Paper No. 1283. Available at: http://hdl.handle.net/10419/67781.