Three Essays on Communication in Signalling Games

vorgelegt von
MSc Lilo Wagner

von der Fakultät VII – Wirtschaft und Management
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Wirtschaftswissenschaften
– Dr. rer. oec –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Georg Meran
Gutachter:	Prof. Dr. Pio Baake
Gutachterin:	Prof. Dr. Dorothea Kübler

Tag der wissenschaftlichen Aussprache: 15. Dezember 2014

Berlin 2015

Summary

This dissertation is concerned with various economic applications of signalling games. In this type of games, a sender transmits information to a receiver, who interprets it.

Educational institutions – senders – evaluate students based on a fixed grading scale. Full disclosure of information is the best response by schools to a student body that consists of expected utility maximizers. The impact of a deviation from this behaviour is investigated: the human mind typically focuses on salient states – the very best or worst grades here. It is demonstrated how this behaviour brings educational institutions to employ coarse, rather than fine, grading schemes.

While educational institutions presumably disclose information truthfully, this is not necessarily the case for privately organized certifiers. If information manipulation is not possible, a certifier should be interested in fully revealing information; at least if producer investment correlates with product quality. It is shown that if this so called threat of capture exists, information disclosure may be coarse. A reduced transparency may also increase social welfare because it prevents the market from breaking down.

In the market for life insurances, firms which are asymmetrically informed about the risk of applicants, might benefit hereof. Information is implicitly transmitted to applicants and competitors via contract offers. It is shown how the non-ability of an informed firm to persuasively transmit information serves all insurers. Industry profits increase at the expense of consumers. This structure is created by the exchange of information on the bargaining behaviour of applicants.

Keywords: Signalling games, Information disclosure, Grading, Education, Behavioural theory, Certification, Bribery, Insurance markets, Asymmetric information

Zusammenfassung

Diese Dissertation befasst sich mit verschiedenen ökonomischen Anwendungen von Signalspielen. In dieser Art von Spielen übermittelt ein Sender einem Empfänger Information, die dieser interpretiert.

Bildungsinstitutionen – Sender – bewerten Studenten nach einem festgelegten Notensystem. Die vollständige Offenlegung von Information ist die beste Strategie von Institutionen, wenn Studenten ihren erwarteten Nutzen maximieren. Untersucht wird eine Abweichung hiervon: menschliche Wahrnehmung konzentriert sich oftmals auf besonders auffällige Zustände – hier sehr gute oder sehr schlechte Noten. Es wird gezeigt, wie dieses Verhalten Bildungsinstitutionen dazu bringt, Notensysteme gröber zu gestalten.

Während davon ausgegangen werden kann, dass Bildungsinstitutionen Informationen, wenn schon nicht vollständig, so doch wahrheitsgemäß offenlegen, ist dies bei privatwirtschaftlich organisierten Zertifizierern nicht notwendigerweise der Fall. Ist die Manipulation von Information nicht möglich, sollte ein Zertifizierer Interesse an der vollständigen Offenlegung von Informationen haben, wenn Produzenten Einfluss auf die Qualität ihrer Güter nehmen. Hier wird gezeigt, dass, besteht die Möglichkeit zur Falschzertifizierung, dies nicht mehr der Fall ist. Eine reduzierte Transparenz ist nicht nur im Sinne des Zeritifizierers wünschenswert, sondern sichert auch die Funktionsfähigkeit von Märkten.

In dem Markt für Lebensversicherungen können Firmen, welche asymmetrisch über das Risiko von Bewerbern informiert sind, von diesem Zustand profitieren. Information wird implizit über Vertragsangebote an Bewerber und Wettbewerber übermittelt. Es wird gezeigt, wie durch die Tatsache, dass eine informierte Firma nicht in der Lage ist, Information überzeugend zu übermitteln, Industriegewinne zulasten der Konsumenten gesteigert werden können. Diese Marktstruktur wird geschaffen durch den Austausch von Informationen über das Verhandlungsverhalten der Bewerber.

Schlüsselwörter: Signalspiele, Informationsoffenlegung, Notengebung, Bildung, Verhaltensökonomie, Zeritfizierung, Bestechung, Versicherungsmärkte, Asymmetrische Information

Acknowledgements

I am indebted to many people for accompanying me on the path of completing this dissertation. First and foremost, I would like thank my supervisor Pio Baake. He continuously guided and supported my research and shared his profound knowledge on the topic of microeconomic theory with me.

I would also like to express my sincere gratitude to Dorothea Kübler for very kindly agreeing to be my supervisor and lending her expertise from the field of microeconomics.

All chapters of this thesis have immensely benefited from the comments of many people – many thanks especially to Georg Weizsäcker, Ives Breitmoser, Paul Viefers, Daniel Huppmann, Andreas Harasser, Laura Wagner and Alfonso Caiazzo for this. I also owe a huge debt of gratitude to Tobias Schmidt for sharing many insights and his enthusiasm on new developments in decision theory, to Vanessa von Schlippenbach and Balkan Akbaba for introducing me to the DIW, to my coauthors for productive and delightful cooperation and to Alex Gilbert, Kathy Sartori and Adam Lederer for proof-reading.

I wrote this dissertation while being a doctoral student at DIW Berlin. This program's research environment helped tremendously in completing this thesis. Financial support from DIW Berlin is gratefully acknowledged.

Last but certainly not least, I would like to express many thanks to my family and my dear friends without whose support writing this thesis would have been a much more arduous path.

Contents

1 Introduction

1.1 Sender-receiver games

Sender-receiver relationships are prevalent in everyday life. By transmitting signals, a sender often reveals valuable information to a receiver. For instance, good grades in college may give information about a high educational quality, but also about the good productivity of a potential employee (as suggested in the canonical example due to Spence 1973).

Signals are subject to interpretation by the receiver. If a product is labelled bio, its quality may meet a defined minimum standard, but is it possibly even better? Also, what can be inferred from the absence of this label?

These relationships show different properties depending on the natural settings in which information is transmitted.

One important way to classify sender-receiver games – also: signalling games – is sender manipulation. For instance, a job applicant will typically not be suspected to have forged degrees since doing so may entail high fines. By contrast, an insurance broker is not made responsible for having sold inappropriate products to his clients.

In case signals are subject to manipulation, a further distinction is whether signals are costly, or more specifically, whether different signals incur different costs. In the job market signalling example due to Spence (1973), persons incur costs for high degrees, their utilities are hence directly connected to the signal choice. An insurance broker on the other hand receives commission for selling products, but talking is costless (he is 'talking cheap'). His payoff therefore solely depends on the receiver's interpretation.

This dissertation treats different topics and settings of sender-receiver relationships. Its chapters integrate into signalling games as follows.[1]

[1] A more detailed description of the objectives and results of the single chapters is given further below. Also, note that structures are richer than in the basic sender-receiver game. In Chapter 3, this inverts the motivation: while cheap talk games are concerned with how persuasive communication can possibly be, the work is motivated by the question why information is not exactly revealed.

Signals are not subject to manipulation: Chapter 2.
This chapter explores how educational institutions (the senders) should optimally choose their grading scales through which they reveal information about the educational level or quality of their students to potential employers (the receivers). Although universities may have an incentive to make their students appear as brilliant as possible, they repeatedly interact on the market. Attempts to overstate the real market value of students quickly become common knowledge.

Signals are subject to manipulation and signalling is non-costly: Chapter 3.
It is analysed in which way privately organized certifiers (the senders) choose to reveal information at hand. Most of the economic literature presumes honest information disclosure. Albeit rather rare, intentional mis-certification is discovered from time to time– the ADAC and ZDF-scandals are two recent examples. The chapter explicitly focuses on the possibility of information manipulation and in particular on how it affects information revelation.

Signals are subject to manipulation and signalling is costly: Chapter 4.
Here, an insurer signals the personal default risk of an applicant by offering him a contract. Because these offers are binding, information transmission is costly. The chapter's focus is on demonstrating how senders may benefit from not being persuasive (to be defined).

1.2 Methodological approach

A sender S sends a message $m \in M$ to a receiver R. The message is chosen from a message space M. R observes m and responds by choosing an action a from some action space, which affects both players' payoffs.

S possesses some private information $\theta \in \Theta$ (his type), where Θ is the type space. Its cardinality is not larger than that of M, meaning that messages are allowed to be different for each type. R's type space consists of a single element, that is, R is not privately informed. He holds, however, beliefs about θ. Prior to information being sent, these beliefs correspond to the true probability distribution over Θ with probability function $p(\theta)$.

Perfect Bayesian Equilibrium. After m has been transmitted and observed, R updates his beliefs according to Bayes' Rule when possible, meaning that beliefs have to be consistent in equilibrium. More specifically, in equilibrium, R knows S's strategy σ^*, which is a specification of a message m for each type in Θ. Therefore, R is able to infer information from the observed message. Upon observation of an on-path message m, R's believed probability that S's type is θ, $\tilde{p}(\theta|m)$, is given by

$$\tilde{p}(\theta|m) = \frac{p(\theta)\sigma^*(m|\theta)}{\Sigma_{\hat{\theta}\in\Theta}p(\hat{\theta})\sigma^*(m|\hat{\theta}))}.$$

Equilibria in which R updates his beliefs accordingly are called *Perfect Bayesian Equilibria.*

Signal manipulation. If signalling is not subject to manipulation by S, he commits to some previously announced signalling strategy σ. A game without manipulation has the following basic structure.

(1) S announces and commits to some strategy σ,

(2) θ realizes or is observed by S,

(3) S reveals according to σ,

(4) R takes an action a.

If S does not commit to some strategy prior to learning θ, stage (1) is omitted. In Chapter 3, however, the game formulation still incorporates this stage. Instead, S announces σ, from which he can deviate in stage (3). Basically, the formulation does not affect the set of equilibria. Pre-announcements however can be understood as a reference point, for which deviation incentives have to be checked. That is, S 'chooses' an equilibrium. By contrast, if (1) is omitted, it is rather that equilibria arise.

Costly signalling. In both settings – costly or non-costly signalling – equilibria usually contain unused messages in M. The concept of Perfect Bayesian Equilibrium makes a statement only about R's signal interpretation on the path. It says nothing about the interpretation of off path messages, however. Multiple equilibria arise as a result.

If messages do not affect S's payoff, their value is generated only through their use in equilibrium. In that case, every off-path message can just be interpreted like some on-path message, and equilibria persist (Sobel 2009). Stated differently, off-path beliefs are of concern only when signalling is costly, as in Chapter 4.

Several restrictions to off-path beliefs have been proposed, the most famous being the *Intuitive Criterion* by Cho and Kreps (1987). It works as follows.

Consider some candidate equilibrium strategy $\sigma^*(m|\theta)$, the corresponding sender equilibrium payoff $U_S^*(\theta)$ and some message $m' \in M$ which is not sent for any θ according to σ^*. When observing this unexpected message m', R best responds to m' by choosing action a^*. This choice depends on R's off-path belief about the probability that the type is θ, $\tilde{p}^d(\theta|m')$ and possibly also on m'.

The *Intuitive Criterion* then states that if for some θ, the sender's deviation profit $U_S(\theta, m', a^*(\tilde{p}^d(\theta|m'), m'))$ is smaller than $U_S^*(\theta)$ for all possible receiver beliefs, then R must put zero probability on the event θ. A deviation to m' is said not to be *admissible* on θ. All candidate equilibria which required the off-path belief $\tilde{p}^d(\theta|m')$ to be strictly greater than zero are eliminated.

In words, upon seeing an unexpected message that is not admissible on θ, it is as if the receiver is implicitly making the following speech.

'If S's type is θ, he would not have sent me that message (m'). Because if he had done so, he would have been worse off as compared to sending the equilibrium message – no matter what my belief about his type is.'

1.3 Objectives and contributions

Chapter 2: optimal grading in schools.

In this chapter, I seek to provide an explanation for the prevalence of coarse grading schemes in schools. While students' examination results are recorded on a very precise scale, like 1,2,..,99,100, final reporting is much less informative. American colleges for instance often adopt letter grading systems or report on a pass/fail basis. Also, at universities with letter schemes, not all grades are used – especially top schools like Harvard are known to generously award A's, whereas grades below B are virtually non-existent. This phenomenon known as grade inflation makes grades even less informative.

This coarse grading is puzzling, as it generates informational asymmetries between the school/the students and employers. This asymmetry however should not be in the interest of the sender (the school) – it reduces the welfare of all market participants.

To see why, consider a game with the commitment structure as outlined above: the school which is assumed to act in the best interest of its students announces a system (a signalling strategy); students employ effort (and thereby influence the realization of the type, or educational quality); grades are awarded according to the previously announced strategy and students go on the job market, where employers (the receivers) pay higher salaries for higher educational quality.[2]

If different qualities are pooled together and if studying creates costs of some sort to

[2] Note also that this is not a signalling game in the sense of Spence (1973): employers are not awarded talent or costs for studying, but the outcome.

students, their best response to coarse schemes will be to study less. In the extreme case where no information is revealed at all, that is, the same message is sent for all types, students will not study at all. Therefore, the probability for high educational quality is low.[3] Employers update according to Bayes' rule, and pay only low salaries to graduates from the respective school. If by contrast information is fully revealed, students will be awarded for their studying efforts.

It is therefore surprising that most top tier schools employ particularly coarse schemes whereas institutions of lower reputation tend to reveal more information. The explanation I offer is well captured by the following conjecture.

"*Ivy League educational institutions attract a disproportionate share of grade obsessed overachievers. [...] Their compulsion to succeed as others define it and their sheepish failure to prioritize higher-order benefits with their time at college perhaps makes a grading system that is based on obvious grade inflation the best option available.*"[4]

It is long known in social science that people tend to draw special attention to the most salient events. In terms of choice under uncertainty, lottery payoffs that stand out are overweighted at the expense of rather average outcomes. As shown by Bordalo et al. (2012), this phenomenon accounts for several puzzles of decision theory, like the Allais (1953) paradox. Allais has demonstrated how actual observed behaviour contradicts the independence axiom of expected utility. For illustration, consider the famous version of Kahnemann and Tversky (1979). Individuals are asked to make a choice between two lotteries, $L_1(z)$ and $L_2(z)$:

$$L_1(z) = \begin{cases} \$2500 \text{ with prob. } & 0.33 \\ \$0 & 0.01 \\ \$z & 0.66 \end{cases} \qquad L_2(z) = \begin{cases} \$2400 \text{ with prob. } & 0.34 \\ \$z & 0.66 \end{cases}$$

By the independence axiom, expected utility theory predicts that an individual's preference ranking should be stable in variations of the common consequence z. However, when z is 2400, most people prefer lottery $L_2(2400)$ whereas if $z = 0$, people more frequently opt for $L_1(0)$.

As opposed to Kahnemann and Tversky's prospect theory, Bordalo, Gennaioli, and

[3]Inefficiencies from informational asymmetries may arise in the absence of moral hazard (i.e. in adverse selection settings) if higher qualities imply higher reservation utilities. This is Akerlof's famous lemon problem (Akerlof 1970). In this educational setting, this would be the case if good students preferred to not work at all rather than work at a salary which is considered too low.

[4]"In Defense of Grade Inflation at Harvard", *The Atlantic*, December 6, 2013.

Shleifer (2012)'s salience theory suggests that distortions are based on the value of z rather than on the value of the underlying probabilities.[5] If $z = 2400$, the payoff \$0 is eye-catching – or salient – in $L_1(2400)$, and therefore overweighted. On the other hand, when $z = 0$, the payoff of \$2500 in $L_1(0)$ is more salient than the payoff of \$2400 in $L_2(0)$, which lets $L_1(0)$ appear relatively more attractive.

In particular, Bordalo, Gennaioli, and Shleifer (2012) propose that the decision maker considers different payoff states in the state space Λ. For instance, consider $z = 2400$. Then, the state space Λ is given by $\{(2500, 2400), (0, 2400), (2400, 2400)\}$. The salience of each of these states is then described by a salience function which satisfies what is denoted as the ordering property: consider two states s and s'. If the lowest payoff in s is lower than the lowest payoff in s', and the highest payoff in s is higher than the highest payoff in s', then s is more salient than s'. In the example above, this means that $(2500, 2400)$ and $(0, 2400)$ are both more salient than state $(2400, 2400)$, whereas the property says nothing about the salience relation of states $(2500, 2400)$ and $(0, 2400)$.[6]

The individual then weights the states following a weighting function. The degree to which salience distorts perceptions is measured by some variable $\delta \in (0, 1]$. If it is 1, the decision maker is an expected utility maximizer, if it is 0, he considers only the most salient state(s).

Applied to the grading setting, a student's effort choice can be interpreted as a lottery choice. His attention is drawn to the most salient grades or, more precisely, to the most salient payoffs associated with grades. This affects incentives to study. I assume that educational institutions seek to maximize its students' expected utilities.

Clearly, if students are expected utility maximizers themselves, a fully revealing rule is optimal. It is shown that extreme distortions bring students to employ too much effort from the school's perspective. Roughly speaking, this is the case because the true probability for the state *(worst grade when choosing no effort, best grade when choosing effort)* is higher than that of the state *(best grade when choosing no effort, worst grade when choosing effort)*. Therefore, when multiplied by a constant larger than 1, the effect is more important for the first state.

Subsequent to this result, it is then shown that every coarseness category (that is, the number of factually assigned grades in a system) is optimal for some bias measure δ. Moreover, optimal rules exist for all δ for which students employ too much effort.

[5] Prospect theory suggests that very small probabilities are overweighted. Therefore, if $z = 2400$, the probability 0.01 of getting \$0 is overweighted while when $z = 0$, the total probability of getting \$0 is 0.67.

[6] Bordalo, Gennaioli, and Shleifer (2012) also suggest that the function should satisfy diminishing sensitivity, but I do not need this for my results.

Numerical analysis suggests that this is often the case, and that lower δ require coarser rules.

The bias measure δ can be interpreted as students' inherent competitiveness, or school selectivity. Put differently, grade obsession in the above citation is translated into a focus on both the very good and very bad results. In that sense, the results account for both the occurrence of coarse schemes and the different degrees to which information is suppressed.

Chapter 3: information disclosure for product certification.

This article, which is co-authored by Martin Pollrich, treats a related topic. Here, we seek to give an explanation to why privately organized certifiers do not fully reveal information. Like in educational institutions, information is usually recorded very precisely, whereas revelation is not. For instance, biofood-labelling is often based on a pass/fail decision, where only 'passes' receive certification. The German organization Peta differentiates two passing animal-rights categories (one or two stars), the cosmetic label Natrue offers three passing certificates.

The basic market structure differs from the grading setting in Chapter 2: privately organized certifiers are profit-maximizers, their incomes are generated through fees which typically have to be paid upon application for certification. Producers of goods typically have an idea of its quality, which also determines their decision for application. In that sense, producers are also senders here. If some qualities remain uncertified, consumers (the receivers) also hold beliefs about these types and they have to be consistent in equilibrium (see above).

The following is the basic game structure: in a first step, the certifier (the sender) announces a disclosure rule (a signalling strategy) and a fee which has to be paid by producers who apply for certification. Sellers make an investment which relates to the realization of the quality (the type). Quality is observed by sellers who then decide whether to apply or not. If so, quality is revealed according to the rule. On the market, consumers are willing to pay more for higher qualities.

Despite the different setting, as in the previous chapter, full disclosure can be shown to maximize certifier utility – her profits – and should therefore be observed more often.[7]

As an explanation for why this is not the case, we propose the following: certifiers may be tempted to accept bribes for releasing favourable certificates. This behaviour, which we call *capture*, enables the certifier to extract payments other than the certification fee. Since consumers are aware of this threat of capture, the certifier must find a way to credibly commit to honesty. The best way to do so is to employ a coarse disclosure rule.

[7] Lizzeri (1999) has shown in a seminal paper that coarseness is, in a similar setting, optimal when quality is exogenously given. This should however not often be the case.

The basic intuition for this is that coarseness reduces a seller's willingness to pay for bribery, since it lowers differences in the market values of certificates.

Methodically, as mentioned above, off-path behaviour is not much of a concern here. It does not make much sense to consider messages outside the previously announced rule – a certifier whose rule it is to rate on a one to five star scale and who then certifies a product '12.7' is assumed to have rated a worthless good. Instead, deviations are possible on the path, i.e. a two-star product may receive five stars instead.

Certifier incentives are shaped by consumer beliefs about his actual behaviour. We analyse whether equilibria exist in which consumers are confident, meaning that they faithfully trust certificates. In order for non-babbling equilibria to exist at all then, the certifier must be punished for deviations.[8] This is done by a reputation mechanism in which consumers adopt a trigger strategy: they trust certificates as long as no deviation is detected (after purchase), and never buy again otherwise.[9] This is the hardest possible punishment. If such an equilibrium does not exist, then neither does it for other punishment strategies.

Our conclusion that coarseness reduces the threat of capture then implies that equilibria can be sustained for lower discount factors than if information is fully revealed. Therefore, informational asymmetries are socially desirable in our setting – they might help to prevent market breakdowns, and should therefore not be objected per se.

Chapter 4: non-persuasive insurers.

This paper is a joint work with Julian Baumann. Like the previous chapter and unlike Chapter 2, it treats a rational choice model in which the sender cannot commit to a signalling strategy.

Providers of life or disability insurances offer contracts on the basis of the outcome of a screening. More specifically, a person who wishes to apply for insurance is required to provide detailed information on his medical record. Because firms dispose of advanced technological means to estimate the statistical risk of a loss, they may gain an informational advantage versus the applicant. In this chapter, we consider this possibility. By offering an individual contract to some person, the insurer (the sender) may reveal information about the person's default risk. Clearly, an insurer would always like to signal

[8] Babbling equilibria are equilibria in which no information is transferred. In non-costly signalling games, these equilibria always exist. The qualitative difference to classic cheap talk models à la Crawford and Sobel (1982) in which non-babbling equilibria may exist without reputation mechanisms is sender utility: here, the certifier would always like to signal high quality (against the bribery payment).

[9] More precisely, we require that deviations be detected with certainty. This restricts the set of feasible rules because if a rule is announced for which deviations are not unambiguously detected, equilibria cease to exist. In that sense, the paper does not make general statements on desirability of coarse rules, but provides a possible explanation.

high risk because an applicant's willingness to pay increases in his default probability.

In this market, insurance premiums are paid constantly – but only until default occurs. This implies that the insuree has no possibility to punish sender misbehaviour. This forces, unlike in the previous chapter, receivers to be suspicious against signals.

More precisely, all equilibria require the following receiver belief. *'Whenever accepted, I believe that my risk is low. Only when rejected, I am willing to accept that my risk is high.'*

This has been shown by Villeneuve (2005). The reason is that if a contract offer existed that could make the applicant believe that he is a high risk, an insurer would always offer him this contract – even if he is a low risk.[10]

The market for life or disability insurances has another special feature: applicants are not only required to provide information on their medical data, but also on their application history. In particular, they are asked whether they have applied for insurance before and whether they have been rejected or accepted.

From this observation, we conclude that at least some applicants should be bargainers, on the search for the best offer they can possibly get. Price takers on the other hand are assumed to consider only one offer.

We analyse a game in which insurers compete in two stages. In particular, the game has three steps: first, two insurers compete in list prices. Second, the firm that makes the best offer wins all applicants, i.e. price takers and bargainers. This firm can either accept the applicant and offer the previously announced list price or reject him. Only bargainers continue their search, and firms compete for them in a third step.

We allow firms to decide whether to collect, upon receiving an application, information on the default risk of each single applicant. This decision is not observable by other players (applicants and competitors). Every equilibrium proves that firms are asymmetrically informed on each single applicant (the winner of the list price competition is informed).

This game formulation generates a special effect, which we call *persuasiveness*, or *non-persuasiveness* respectively. What is understood by this is the following. Consider some candidate equilibrium where applicants with low default probability are accepted in the first place. The informed firm may consider to deviate from this equilibrium by rejecting an applicant of this type. By doing so, she trades off loosing low risk price takers against being able to ask high risk premiums from the low risk bargaining type – who consequently believes to be a high risk.

If the informed insurer is tempted to deviate in this way from an equilibrium where

[10] As opposed to Spence (1973), signalling is not costly, and therefore not possible. However, pooling equilibria may not exist if the low risk applicant is not willing to pay at least the fair price for the high risk type.

all bargainers are offered their fair prices, she is said to be non-persuasive. Otherwise, she is said to be persuasive.

Whether an informed firm is persuasive determines market outcomes: if the firm is persuasive, all bargainers are accepted in the competition stage and receive fair offers. If not, by contrast, high risk bargainers accept the offer from the uninformed firm – even if it exceeds the fair price.

The following belief of an applicant who has been rejected previously forms an equilibrium when the firm is not persuasive.

'Whenever in the second round of competition, the best offer comes from the uninformed insurer, believe that the risk is high. If the best offer comes from the informed firm, believe that the risk is low.'

An equilibrium with this belief structure survives the Intuitive Criterion: the informed firm would like to deviate, upon having observed a high risk type, in the second round of competition, to a price which exceeds the fair price for the high risk type but undercuts her competitor's offer. Suppose he does. If the applicant believes that his risk is high, he accepts the deviation offer. But if the applicant is in reality a low risk, the insurers' profits are large if the person can be convinced to be a high risk. On the other hand, the regular equilibrium profits on low risks are lower – by definition of non-persuasiveness. Following the logic of the Intuitive Criterion, the deviation is then admissible on the low risk type, and the equilibrium is not eliminated.

As a consequence, when the informed firm is not persuasive, the uninformed firm benefits. This impacts list price competition: knowing that not winning this competition earns this firm positive profits mitigates competition in the first stage. Put differently, all firms benefit from the effect. In fact, it can be shown that firms manage to always coordinate on this non-persuasiveness outcome. Perhaps surprisingly, industry profits increase in the share of bargainers in the market.

Finally, if firms were prohibited from collecting information on the bargaining behaviour of applicants, they would be symmetrically informed, and effective competition would be restored.

2 Pass/Fail, A-F or 0-100? Optimal grading of eager students

Chapter Abstract

Most schools grade their students using a very coarse grading scheme, making grades essentially less informative. Especially highly selective schools suppress more information than schools with lower reputation. We formalize an explanation which is intuitively described as the 'excessive preoccupation with grades' in top schools. Coarsening is the best response by schools to a student body that consists of individuals who overweight grades that stand out on their school's scale. Such overweighting of salient grades leads students to exert more effort than expected utility maximizers. Schools which attempt to maximize an EU-welfare criterion counterbalance this bias through the use of broad grading schemes.

2.1 Introduction

Students' examination results are typically recorded on a very accurate scale, like $1, 2, ..., 99, 100$. Yet, educational institutions pool different scores together in broad categories, for instance by adopting a letter grading system. In fact, perfectly accurate final reporting can rarely be found. This prevalence of coarse grading systems is puzzling as it reduces the informativeness of diplomas to employers. Even more surprisingly, selectivity seems to be a good indicator for the degree to which information is suppressed. Yale Law school for instance employs a broad system on the basis of fail/low pass/pass/honours. Other top law schools like Harvard and Stanford have adapted similar systems while less selective schools tend to stick to the finer traditional (and usually curved) letter grading system.

A more subtle method to install a coarse system is to factually abolish lower marks by just never or very rarely assigning them, a phenomenon recently discussed in the news under the headline of grade inflation. Stuart Rojstaczer, one of the initiators of the debate states in a *Washington Post* article: *"We recently handed in our grades for an undergraduate course we teach at Duke University. They were a very limited assortment: A, A-minus, B-plus, B and B-minus. There were no C's of any flavor and certainly no D's or F's. It was a good class, but even when classes aren't very good, we just drop down slightly, to grades that range from A-minus to B-minus."*[1]

At Harvard, the most frequently given grade is now an A, the median an $A-$.[2] When the *Business Insider* lists *13 schools where it's almost impossible to fail*[3], the list comprises almost exclusively elite private schools. More selective institutions assign higher GPA's to their students than do their peers for the same education level (Healy and Rojstaczer, 2012), presumably corresponding to a lower support in the scale.[4]

The purpose of the present paper is to give an explanation for the prevalence of such coarse grading and for the different degrees to which information is pooled.

The factual non-existence of lower grades at the elite schools resulting from grade inflation is commonly considered a problem, the main arguments against it being both low informativeness and reduced effort incentives. This view provokes calls on school administrations to take action. Their behaviour however mostly suggests that this phe-

[1]"Where All Grades Are Above Average", *Washington Post*, January 28, 2003

[2]"Substantiating Fears of Grade Inflation, Dean Says Median Grade at Harvard College Is A-, Most Common Grade Is A", *The Harvard Crimson*, December 3, 2013

[3]"13 schools where it's almost impossible to fail", *The Business Insider*, May 29, 2013

[4]Albeit not perfectly correlated, the fact that the share of D's at private colleges has now dropped to less than 3% (from the same report) leads us to conclude that. Also, many reports indicate that this phenomenon is prevalent most and forall at the elite schools.

nomenon is wilfully neglected, if not actively initiated.[5] One might be led to believe that this coarsening is the equilibrium outcome of a game in which employers can be mislead about educational quality (as for instance investigated in Chan et al. 2007), it is however widely accepted that grading standards are common knowledge among institutions, students and prospective employers.

We call attention to a different explanation. Information suppression may be the best response of schools to a student body who consists of individuals who place too much weight on salient grades.

Conjectures pointing to the same direction are well captured in the following statements: *"Ivy League educational institutions attract a disproportionate share of grade-obsessed overachievers. [...] Their compulsion to succeed as others define it and their sheepish failure to prioritize higher-order benefits with their time at college perhaps makes a grading system based on obvious inflation the best option available."*[6]

Similarly, when Yale Law School changed its system from numerical grading to the broader system in 1967, it was reported that *"it is believed by those members of the faculty who voted in favor the plan that it will offer some relief from what one educator described as 'the excessive preoccupation with number or letter grades' ".*[7]

To formalize what is understood as 'grade obsession' or 'excessive preoccupation' in the above citations, we posit that students' attentions are drawn to salient payoffs according to Bordalo, Gennaioli, and Shleifer (2012) – henceforth BGS. We further claim that the degree to which this is done is determined by students' inherent competitiveness, or school selectivity.

Thus, grade obsession is translated into a focus on both the very good and the very bad outcomes at the expense of average results.[8] Since students can never be certain about the success of their efforts, they are faced with the choice between independent lotteries, with payoffs being determined on the job market. Due to the salience of extreme outcomes, students replace the true outcome probabilities by salience-distorted weights. More specifically, the salience of a payoff is context-dependent, meaning that outcomes are compared to those of other lotteries which realize in the same state of the world.

We take it for granted that all grades of all lotteries are taken into account, however small their occurrence probability may be. With states being evaluated according to a salience function that exhibits ordering, meaning that for a consideration set with two

[5] In fact, when the issue came up in 2002, the only institution that would revise its grading policies was Princeton University.

[6] "In Defense of Grade Inflation at Harvard", *The Atlantic*, December 6, 2013

[7] "Yale College inaugurates pass-fail marking system", *The Heights*, November 13, 1967

[8] The hypothesis that focusing equates to focussing on large differences is also explored by Bordalo et al. (2013), Kőszegi and Szeidl (2013) and Tversky (1969).

lotteries, if the payoffs associated with state s are nested within the payoffs associated to some other state s', then s' is more salient, the most salient states are {lowest payoff in lottery 1, highest payoff in lottery 2} and {highest payoff in lottery 1, lowest payoff in lottery 2}.

Consider the following setting: an institution which seeks to maximize the expected utility of its students commits to and publicly announces a grading system. Based on this, students of different cost types[9] make an unobserved decision between employing costly high effort, and costless low effort. The associated probability mass functions are standardly assumed to satisfy the monotone likelihood ratio property. Scores then realize, grades are assigned, and students go on the job market where risk neutral employers are willing to pay higher salaries for better expected scores.

Clearly, when students maximize expected utility, a grading system which fully reveals scores is socially efficient, and even uniquely optimal.

A focus on the most salient states however brings students to exert inefficiently high effort, from the school's perspective: roughly speaking, there are two most salient states, but a student who considers employing effort overweights the good grades relatively more than the bad grades. Similarly, when considering to exert low effort, the student overweights bad grades relatively more than good grades. This effect arises due to the monotone likelihood ratio property.

The share of hard working students is a continuous function of the degree of salience, implying that coarse rules which mimic full disclosure at the extreme have an intersection with the expected utility maximizing cut-off and are therefore optimal for some distortion level. It can further be shown that optimal coarse rules exist whenever full disclosure induces too many students to choose the effort lottery, from the school's perspective. Numerical analysis suggests that this is typically the case for all levels of distortion. Further: (a) given a coarseness category (e.g. Pass/Fail or A,B,C), optimal rules exist for a closed interval of the bias measure, (b) stricter biases require coarser rules, (c) within one coarseness category, extreme grading (e.g. almost all pass when the grading system is Pass/Fail or almost all receive an A when grading category has multiple cuts) is optimal for high distortions. Therefore, our results account for both the occurrence of coarse grading schemes and for the different degrees to which information is suppressed.

[9]This may the opportunity costs for what is named higher order benefit in the above citation: it is a well established thesis among psychologists that while extrinsic motivation responds well to incentives, it comes at the cost of a reduced intrinsic motivation (for an overview and critical assessment: Cameron and Eisenberger (1996)). Similarly, institutions may be concerned about the health of their students: at Harvard Law School, the system change reportedly served to reduce stress and anxiety levels.

Related literature. Coarse grading may be advantageous in contest games. In particular, Dubey and Geanakoplos (2010) show that if students care about their rank in class and grading schemes are chosen such that students are best motivated, coarse schemes may be the best choice and are always so when students are disparate. The basic idea is that when disparate, students of lower ability should be given the chance to outperform their high ability classmates. Similarly, these should not feel too comfortable. Their conclusion is qualitatively different from ours as coarse schemes motivate students to work more, not less.[10] Further, Ostrovsky and Schwarz (2010) analyse best disclosure policies when employers rank students according to their exogenously given abilities. Partial disclosure can be a market equilibrium if student bodies exhibit different exogenously given talent levels. In particular, an average school, if faced with low quality competitors, may take into account non-revelation of highly talented students in order to attain a good placement for less able persons. A similar effect is present in Boleslavsky and Cotton (2014).[11]

In a non-contest setting, Costrell (1994) analyses optimal Pass/Fail schemes under different welfare conceptions. Rayo and Segal (2010) show that pooling can be optimal if grading is based not solely on the abilities or educational achievement of students but also on their profitability to the sender (the paid tuition fees).

In industrial organization, a stream of literature starting with Lizzeri (1999) shows why profit maximizing certifiers may find it optimal to employ broad disclosure policies. In Lizzeri (1999), a certifier who is bound to offer his service at a fixed non-discriminatory fee to sellers with different quality goods considers disclosure on a Pass/Fail basis optimal. In comparison to an educational setting, participation is voluntary with the decision being based on a previously observed quality level. Similarly, intermediaries who seek maximize not their own profits but the information being provided to the public, partial disclosure may outrule fully revealing rules if a selection process is at work and certification is costly (Harbaugh and Rasmusen, 2013).

The article proceeds as follows. Section 3.2 describes the model. In Section 2.3, we derive two general insights, which are independent of the focus on salient payoffs but relate only to the game structure. The salience bias is characterized in Section 2.4, and optimal rules are analysed in Section 2.5. The last section discusses extensions and concludes. All proofs are relegated to the appendix.

[10] Unfortunately, empirical studies exploring this relationship in educational settings do not exist. Some works investigate the effect of raising a Pass/Fail standard (e.g. Figlio and Lucas (2004), Betts and Grogger (2003)) and others find that effort is chosen strategically, i.e. such that scores closely above a threshold become probable (Oettinger, 2002), but none systematically explores the effect of coarsening grades in relation to finer schemes.

[11] Many more examples of how information suppression takes place is presented in both Ostrovsky and Schwarz (2010) and Boleslavsky and Cotton (2014).

2.2 The model

Students are assumed to have a cost for studying of $\theta \in [0,1]$. θ is uniformly distributed across students. A student's choice set is $\{\mathcal{L}_0, \mathcal{L}_e\}$ where the choice of lottery $\mathcal{L}_e$ costs θ whereas that of $\mathcal{L}_0$ is free. Without loss of generality, students are assumed to be risk neutral. An expected utility maximizing student of type θ then chooses lottery $\mathcal{L}_e$ if and only if the expected value from doing so, minus the costs is at least as high as the expected value of choosing $\mathcal{L}_0$.

The lottery choice determines the probability distribution of exam scores. In particular, an exam score is an element of the ordered set $\Omega = \{q^1, q^2, ..., q^N\}$ with $N \geq 3$. The probability that a score in Ω is q^n is given by p_e^n if the student chooses lottery $\mathcal{L}_e$ and it is p_0^n otherwise. p_e and p_0 are the respective probability vectors. All elements are assumed to be strictly positive.

The values of scores to students are determined by the salaries paid on the job market. There, high scores are rewarded. In particular, it is assumed that $q^n = n/N$ for every $q^n \in \Omega$. Hence, when the market exhibits symmetric information, the expected valuations for both lotteries are $E[q^n|\mathcal{L}_e] = \sum_{n=1}^{N} q^n p_e^n$ or $E[q^n|\mathcal{L}_0] = \sum_{n=1}^{N} q^n p_0^n$ respectively.

Teachers convert scores into grades using a grading rule that maps every possible score q^n into a grade. Denote a grading rule by $\Gamma^H = \{\gamma_1, \gamma_2, ..., \gamma_H\}$ with $\gamma_H = q^N$ and the associated grades by $C = \{C_1, C_2, ..., C_H\}$. For instance, γ_1 could be 30 out of 100 points and C_1 would be grade D. Further, let Ψ be the set of all possible grading rules. In particular, a grading rule in Ψ is a partition of Ω into consecutive intervals. It can be expressed as a subset of Ω: $\Gamma^H \subseteq \Omega$.

The elements of a rule Γ^H are defined such that all scores q^n that are equal to or worse than γ_1 are granted grade C_1, all scores that are strictly better than γ_1 but equal to or worse than γ_2 are assigned grade C_2 and so forth. The best grade C_H is awarded to scores greater than γ_{H-1}. An example for a rule Γ^3 is depicted in figure 2.1: if $\gamma_1 = q^m, \gamma_2 = q^n$, then the lowest grade C_1 is awarded to all scores lower or equal to q^m, C_2 is awarded to scores q^{m+1} to q^n and the best grade to all scores strictly greater than q^n. From this follows that $\gamma_H \equiv q^N = 1$ and $1 \leq H \leq N$.

If $H = 1$, no information is disclosed, if $H = N$, information is fully revealed. We use Γ to describe a general rule and Γ^H to describe some rule with $H - 1$ cuts. The same holds for Ψ and Ψ_H respectively: $\Psi_H \subset \Psi$ is the set of disclosure rules that are described by some Γ^H. For instance, all possible Pass/Fail schemes are contained in the set Ψ_2, where γ_1 is an element of $\Omega/\{q^N\}$.

Note that describing grading rules accordingly has two implications: first, probabilis-

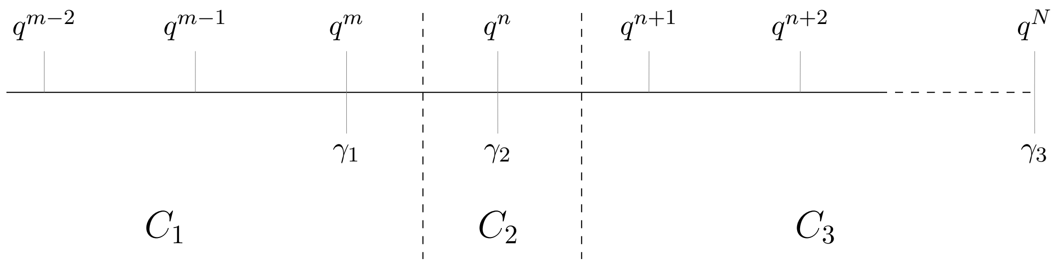

Figure 2.1: A grading rule $\Gamma^3 \subset \Psi_3$.

tic rules are not accounted for (i.e. every score is deterministically assigned a grade)[12] and second, every grade occurs with strictly positive probability.[13]

Employers are informed about the grading rule, on the basis of which they form expectations about the value of a grade C_h, denoted by v_h. Also, the distribution of costs and the probability distributions are assumed to be common knowledge while the students' lottery choices cannot be observed. The game then proceeds as follows:

(1) The grading rule is publicly announced,

(2) students make a choice between $\mathcal{L}_e$ and $\mathcal{L}_0$,

(3) scores are disclosed according to the grading rule,

(4) students sell their education on the job market.

2.3 Optimal grading with expected utility

An expected utility maximizing student θ chooses lottery $\mathcal{L}_e$ if and only if $V_e^\Gamma - \theta \geq V_0^\Gamma$, where V_e^Γ and V_0^Γ are the expected valuations from choosing the respective lottery for a given disclosure rule. Define $\gamma_0 = 0$. Then, V_e^Γ and V_0^Γ are given by

$$\begin{aligned} V_e^\Gamma =& E[v_h|\mathcal{L}_e] = \sum_{h=1}^{H} v_h \sum_{i=\gamma_{h-1}N+1}^{\gamma_h N} p_e^i, \\ V_0^\Gamma =& E[v_h|\mathcal{L}_0] = \sum_{h=1}^{H} v_h \sum_{i=\gamma_{h-1}N+1}^{\gamma_h N} p_0^i. \end{aligned} \tag{2.1}$$

Note that the second sum in both expressions of equation (2.1) is the probability for grade h to occur given the lottery choice and the grading rule: if for instance $\gamma_{h-1} = q^m$

[12]That is, for the moment. Simple probabilistic rules are used to derive an existence condition of optimality for given bias measures, see below.

[13]Similarly, this assumption is not crucial as shortly discussed in the last section.

and $\gamma_h = q^n$, then the probability that a student choosing lottery $\mathcal{L}_e$ is assigned grade C_h is given by $p_e^{m+1} + p_e^{m+2} + ... + p_e^n$.

In order to ensure inner solutions, it is standardly assumed that probability functions satisfy the monotone likelihood ratio property. Define $\tau_n \equiv p_e^n / p_0^n$.

Assumption 2.1. *(*Monotone likelihood ratio property*)* $\tau_n > \tau_m \quad \forall n > m$

The condition guarantees that students who choose the costly lottery $\mathcal{L}_e$ will attain higher expected scores than the ones whose choice is $\mathcal{L}_0$. Therefore, it holds that $\theta \leq V_e^\Gamma - V_0^\Gamma$ for at least some θ if information is revealed at all, that is, if $H \geq 2$.

The principal/the school seeks to maximize his conception of student welfare, which is overall expected utility. Since it is never optimal that high types choose $\mathcal{L}_e$ when lower types do not, it is given by

$$W^\Gamma(\hat{\theta}) = \int_0^{\hat{\theta}} (V_e^\Gamma - \theta) d\theta + \int_{\hat{\theta}}^1 V_0^\Gamma d\theta \tag{2.2}$$

for a given rule Γ. Transforming equation (2.2) and maximizing with respect to $\hat{\theta}$ gives the first best critical value of θ such that all students with lower costs should choose $\mathcal{L}_e$ whereas higher types should choose $\mathcal{L}_0$. It is given by $\hat{\theta}^r = V_e^\Gamma - V_0^\Gamma$. Therefore, expected utility maximizing students share the principal's conception of welfare. The market exhibits symmetric information if the education level is fully revealed to the public. Then, $\hat{\theta}^r$ is the difference in expected quality for a given lottery choice $\mathcal{L}_e$ or $\mathcal{L}_0$.

$$\hat{\theta}^{FB} = E[q^n|\mathcal{L}_e] - E[q^n|\mathcal{L}_0] = \sum_{n=1}^{N} q^n (p_e^n - p_0^n).$$

Denote $W^{FD}(\hat{\theta})$ the $W^\Gamma(\hat{\theta})$ when information is fully revealed. In words, $W^{FD}(\hat{\theta})$ is the expected utility with full disclosure as a function of the student behaviour $\hat{\theta}$. The first best expected utility level is then given at $W^{FD}(\hat{\theta}^{FB})$, that is, if students are expected utility maximizers and scores are fully revealed.

When information is not fully revealed, employers form, for a given disclosure rule, a belief ϕ about the students' lottery choices, $\hat{\theta}$ (i.e. about the critical type). The market values are then a function of this belief. To see why, consider some rule such that grade C_H is coarse: if many students are believed to have chosen $\mathcal{L}_e$, that is, if ϕ is large, then the value of this grade is high as compared to a lower ϕ because the likelihood of a student with an awarded grade C_H having attained a top score rather than a still good score is high whereas it is lower for lower ϕ. If information is fully revealed however, a grade is simply worth the score. The value of a grade C_h is given by

$$v_h(\Gamma|\phi) = \frac{\sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} q^i(\phi p_e^i + (1-\phi)p_0^i)}{\sum_{i=1+\gamma_{h-1}N}^{\gamma_h N}(\phi p_e^j + (1-\phi)p_0^j)}. \tag{2.3}$$

This given, the true expected valuations from choosing lotteries $\mathcal{L}_e$ and $\mathcal{L}_0$ can be derived by inserting equations (2.3) into equations (2.1), $V_e^\Gamma(\phi)$ and $V_0^\Gamma(\phi)$. Students' actions are a (not necessarily rational) reaction to the market values and therefore to ϕ, hence $\hat{\theta}(\Gamma|\phi)$. The used equilibrium concept is the Perfect Bayesian Equilibrium. This implies that on the end market, students' lottery choices can be foreseen and beliefs are correct. Denote $\mathring{\theta}(\Gamma)$ the ϕ that solves $\phi = \hat{\theta}(\Gamma|\phi)$.
For a given disclosure rule Γ, the school maximizes[14]

$$W^\Gamma(\mathring{\theta}(\Gamma)) = \int_0^{\mathring{\theta}(\Gamma)} (V_e^\Gamma(\mathring{\theta}(\Gamma)) - \theta)d\theta + \int_{\mathring{\theta}(\Gamma)}^1 V_0^\Gamma(\mathring{\theta}(\Gamma))d\theta.$$

This is to say, even if students are not expected utility maximizers, what counts for the school is still the expected valuation of the lotteries, albeit their values depend on the equilibrium behaviour. If information is fully revealed, V_e^Γ and V_0^Γ equal the expected scores for a given lottery choice, and thereby do not depend on student behaviour. Hence, $W^{FD}(\mathring{\theta}(\Gamma))$ is a function of the bounds of integration. The following then holds:

Lemma 2.1. *For a given disclosure rule $\Gamma \subset \Psi$, for any $\mathring{\theta}(\Gamma) \in [0,1]$, it holds that $W^\Gamma(\mathring{\theta}(\Gamma)) = W^{FD}(\mathring{\theta}(\Gamma))$.*

Lemma 2.1 states that expected utility is exclusively determined by the student behaviour, the lottery choice. Put differently, with rational actors on the end market, coarseness itself does not affect the expected market value, only the effect on the effort choice does. This holds although $V_e^\Gamma(\mathring{\theta}(\Gamma))$ and $V_0^\Gamma(\mathring{\theta}(\Gamma))$ are typically different from $E[q^n|\mathcal{L}_e]$ and $E[q^n|\mathcal{L}_0]$ respectively. As a result, the first-best outcome is restored whenever $\mathring{\theta}(\Gamma) = \hat{\theta}^{FB}$, in which case we call student behaviour 'optimal'. Similarly, a rule that implements $\hat{\theta}^{FB}$ is referred to as 'first-best' or 'optimal'. It can easily be verified that $W^{FD}(\hat{\theta})$ is concave and symmetric around $\hat{\theta}^{FB}$. Hence, welfare is improved as $|\mathring{\theta}(\Gamma) - \hat{\theta}^{FB}|$ becomes smaller.

It is this study's main purpose to demonstrate why schools may find it optimal to employ a coarse grading scheme rather than fully revealing examination results. If students

[14] We implicitly assume a minimum degree of rationality which is that there exists a critical type such that all higher types choose $\mathcal{L}_0$ and all lower types choose $\mathcal{L}_0$. When students focus on salient payoffs, this is given as will be shown below.

share the school's conception of welfare, a rule is optimal if and only if $\hat{\theta}^r(\Gamma|\phi) = \phi = \hat{\theta}^{FB}$. The following can be shown:

Proposition 2.1. *For all rules in Ψ/Ψ_N, it holds that for any belief ϕ, $\hat{\theta}^r(\Gamma|\phi) < \hat{\theta}^{FB}$.*

Therefore, when students are expected utility maximizers, no other than the full disclosure rule is optimal. More precisely, any rule that reveals less than full information induces also too less effort from the school's point of view, a result which does not depend on the belief ϕ. Here, the intuition that incentives to work are lowered if better scores are not fully rewarded, applies due to the simple structure of the game.

2.4 Salience of extreme grades

Following BGS, students evaluate H^2 different payoff states. Each state $s_{gh} \in S$ occurs with a commonly known probability, given by

$$\pi_{s_{gh}} = \sum_{i=\gamma_{h-1}N+1}^{\gamma_h N} p_0^i \sum_{i=\gamma_{g-1}N+1}^{\gamma_g N} p_e^i \quad \forall g, h \in \{1, 2, ..., H\}.$$

Accordingly, it holds that $\sum_{s_{gh} \in S} \pi_{s_{gh}} = 1$.

The salience of a state $s_{gh} \in S$ is described by a continuous, non-negative and symmetric (states s_{gh} and s_{hg} are equally salient) salience function $\sigma_{s_{gh}} \equiv \sigma(v_g, v_h)$. As proposed by BGS, $\sigma_{s_{gh}}$ satisfies the ordering property.[15] Formally:

Assumption 2.2. (Ordering)
If $\min\{v_{g'}, v_{h'}\} < \min\{v_{g''}, v_{h''}\}$ *and* $\max\{v_{g'}, v_{h'}\} > \max\{v_{g''}, v_{h''}\}$, *then* $\sigma_{s_{g'h'}} > \sigma_{s_{g''h''}}$.

The ordering property says that whenever one state's payoffs are nested within the payoffs of some other state, the latter is more salient. Here, it implies that s_{1H} and s_{H1} are the most salient states, whereas some state s_{gg}, $g \in \{1, 2, ..., H\}$, exhibits the least salience.

To obtain the decision weights attached to some state s_{gh}, define the weighting function as

$$\omega_{s_{gh}} = \frac{\delta^{-\sigma_{s_{gh}}}}{\sum_{s_{st} \in S} \delta^{-\sigma_{s_{st}}} \pi_{s_{st}}}. \tag{2.4}$$

[15] BGS also propose that the salience function should exhibit diminishing sensitivity, that is, for any two states $s_{g'h'}, s_{g''h''}$, where $v_{g''} = v_{g'} + \varepsilon$ and $v_{h''} = v_{h'} + \varepsilon$, for any $\varepsilon > 0$, $\sigma_{s_{g'h'}} \geq \sigma_{s_{g''h''}}$. We do not need this property to derive the general result, we will however, use a function that satisfies it when conducting the numerical analysis.

$\delta \in (0, 1]$ measures the degree of distortion, where $\delta = 1$ means that students are expected utility maximizers. The perceived probability for a state s_{gh} is given by

$$\tilde{\pi}_{s_{gh}} = \omega_{s_{gh}} \pi_{s_{gh}}$$

where $\sum_{s_{gh} \in S} \tilde{\pi}_{s_{gh}} = 1$. Then, what a student perceives to be the value of lottery $\mathcal{L}_e$ for some rule in Ψ_H is given by $\sum_{h=1}^{H} v_h \sum_{i=1}^{H} \tilde{\pi}_{ih}$. Accordingly, lottery $\mathcal{L}_0$ is valued $\sum_{h=1}^{H} v_h \sum_{i=1}^{H} \tilde{\pi}_{hi}$ and the critical cost type is given by:

$$\hat{\theta}(\Gamma^H|\phi, \delta) = \sum_{g=1}^{H} \sum_{h=g+1}^{H} (v_h(\Gamma^H|\phi) - v_g(\Gamma^H|\phi))\omega_{s_{gh}}(\Gamma^H|\phi, \delta)(\pi_{s_{gh}}(\Gamma^H|\phi) - \pi_{s_{hg}}(\Gamma^H|\phi)), \tag{2.5}$$

where $\hat{\theta}(\Gamma^H|\phi, 1) = \hat{\theta}^r(\Gamma^H|\phi)$.

Biased students prefer lottery $\mathcal{L}_e$ over $\mathcal{L}_0$ if their cost type is smaller than what they perceive to be the value difference for a given belief ϕ. Whatever this belief is, $\hat{\theta}(\Gamma|\phi, \delta)$ is bound from above by $q^N - q^1 < 1$ and from below by $q^1 > 0$. Further, the function is continuous in ϕ, and defined on the interval $[0, 1]$, which completes Lemma 2.1: a $\mathring{\theta}(\Gamma|\delta) \in [0, 1]$ exists for any Γ and δ. In other words, a Perfect Bayesian Equilibrium always exists.

2.5 Optimal grading with salience

Our focus lies on rules that induce optimal student behaviour for a given bias δ. Therefore, we restrict our analysis to exploring whether $\hat{\theta}(\Gamma|\hat{\theta}^{FB}, \delta) = \hat{\theta}^{FB}$ has a solution. In words, given that employers believe students to behave optimally, is there a scheme which induces students to indeed do so?

First, consider the case where students are extremely biased. For $\delta \to 0$, students take into account only the most salient states, s_{1H} and s_{H1}, all payoffs not contained in these states are ignored. To see this, write

$$\omega_{s_{gh}} = \frac{\delta^{\sigma_{s_{1H}} - \sigma_{s_{gh}}}}{\sum_{s_{st} \in S} \delta^{\sigma_{s_{1H}} - \sigma_{s_{st}}} \pi_{st}} \quad \forall g, h \in \{1, 2, ..., H\}.$$

By the ordering property, $\omega_{s_{1H}} = \omega_{s_{H1}} = (\pi_{s_{1H}} + \pi_{s_{H1}})^{-1}$ and $\omega_{s_{gh}} = 0$ for all other states, and the critical type under a fully transparent grading scheme is given by

$$\lim_{\delta \to 0} \hat{\theta}(\Gamma^N|\phi, \delta) = (q^N - q^1) \frac{\tau_N - \tau_1}{\tau_N + \tau_1}.$$

The following can be shown:

Lemma 2.2. $\lim_{\delta \to 0} \hat{\theta}(\Gamma^N | \phi, \delta) > \hat{\theta}^{FB}$.

Lemma 2.2 states that if results are fully revealed, and students are extremely biased, effort levels are higher than what is considered desirable by the school.

To give an intuition for the result, consider the overweighting of $\mathcal{L}_e$ relative to $\mathcal{L}_0$. The choice of $\mathcal{L}_e$ attaches too much weight to the lowest grade in state s_{N1} and to the highest grade in state s_{1N}. The opposite holds true for $\mathcal{L}_0$. The reason why the effect that is created by s_{1N} overweights, is that although the degree by which initial weights are distorted is equal, the initial weights themselves are not, that is, $\pi_{1N} > \pi_{N1}$ by Assumption 2.1.

The lemma has an important implication: Since students focus only on the most salient states s_{1H} and s_{H1}, the only grades that are taken into account are v_1 and v_H. In the case of full disclosure, $v_1 = q^1$ and $v_H = q^N$. Therefore, any other rule with $\gamma_1 = q^1$ and $\gamma_{H-1} = q^{N-1}$ yields the same limit result. From Proposition 2.1, we know that for $\delta = 1$, every rule different from the full disclosure rule induces too few students to choose $\mathcal{L}_e$ for any belief. As a result, since $\hat{\theta}(\Gamma | \phi, \delta)$ is continuous in δ, there exists some $\delta < 1$ for which a rule of this form – $\gamma_1 = q^1$ and $\gamma_{H-1} = q^{N-1}$ – is optimal. This insight provides an explanation for the occurrence of coarse grades and shall therefore be stated separately:

Proposition 2.2. *For every $H \geq 3$, Ψ_H contains at least one rule $\tilde{\Gamma}^H$ for which $\lim_{\delta \to 0} \hat{\theta}(\tilde{\Gamma}^H | \hat{\theta}^{FB}, \delta) > \hat{\theta}^{FB}$. Therefore, for every $H \geq 3$, Ψ_H contains at least one rule that is optimal for some $\delta \in (0, 1)$.*

A natural question to ask next is whether optimal rules exist for all δ. However, Γ^H is a subset of Ω, its entries thereby being rational numbers. This implies that even for large N, the first best outcome may not exactly be restored by a deterministic rule, although some optimal probabilistic counterpart exists. Therefore, in order to be able to answer this question, it is necessary to extend the set of rules to probabilistic schemes. As will be shown, to derive a general result, it suffices to consider rules which are probabilistic only to a small degree.

Extended set of rules: probabilistic schemes. We strive to generate continuity of $\hat{\theta}^{FB}(\Gamma^H | \hat{\theta}^{FB}, \delta)$ both within and across coarseness categories H. In particular, define $\mathcal{G}$ as the set of probabilistic rules with the following properties: every score q^n which is not placed on a cut, that is, $q^n \neq \gamma_h$ for any h, is deterministically awarded a grade, as before. All other scores may be assigned grades stochastically. In particular, denote r_h a probability vector, with its elements summing up to one. In particular, an element r_h^g denotes the probability with which score $q^n = \gamma_h$ is assigned grade g. As before, we require that every grade g is awarded with strictly positive probability. A rule in $\mathcal{G}$ is then characterized by the set $G^H = \{\gamma_1, \gamma_2, ..., \gamma_h, ...\gamma_H, r_1, ..., r_H\}$, with $\gamma_H = q^N$, as before. If r_h is such that $r_h^h = 1$ for all h, the rule equals the respective deterministic rule. As turns out, it suffices to consider rules which have that (a) at most one r_h contains elements different from 1 and 0, and (b) at most two elements of r_h are different from 0, and these are neighbouring elements. In words, these rules are probabilistic only to a small degree: at most one score is assigned to at most two grades with positive probability. The following can be shown:

Proposition 2.3. *For any δ for which $\hat{\theta}(\Gamma^N|\delta) \geq \hat{\theta}^{FB}$, an optimal rule in $\mathcal{G}$ exists.*

The proposition states that optimal rules exist for each δ which induces too many students to choose $\mathcal{L}_e$ when information is fully revealed. The idea of the proof is as follows: the rule which does not disclose any information to the market verifies that $\hat{\theta}(\Gamma^1|\phi, \delta) = 0$ for any tuple (ϕ, δ). That is, whatever the belief and however strong the bias, all students choose lottery $\mathcal{L}_0$. Probabilistic rules generate continuity between student reactions to a no-disclosure policy and a fully revealing rule. This is done as follows: Consider the following class of deterministic rules, from which probabilistic rules are derived: for any coarseness category H, $\tilde{\Gamma}^H$ is such that $\gamma_h = q^h$ up to $H-1$. Also, $\gamma_H = q^N$ (by definition). That is, $\tilde{\Gamma}^2 = \{q^1, q^N\}$, $\tilde{\Gamma}^3 = \{q^1, q^2, q^N\}$, $\tilde{\Gamma}^4 = \{q^1, q^2, q^3, q^N\}$ and so forth. Then, continuity between $\tilde{\Gamma}^H$ and $\tilde{\Gamma}^{H+1}$ is generated by implementing a rule in which score q^{H-1} is awarded grade C_{H-1} with probability $r \in (0, 1]$ and C_H with probability $1-r$.

Consider for instance $\tilde{\Gamma}^2$ and $\tilde{\Gamma}^3$. q^2 is then awarded either grade C_2 (probability r) or C_3 (probability $1-r$). If it is assigned C_2 for sure, the behaviour induced is the same as under $\tilde{\Gamma}^3$. On the other hand, when $r \to 0$, it limits $\hat{\theta}(\tilde{\Gamma}^2|\phi, \delta)$ for some given ϕ and δ. Respective probabilistic rules generate continuity of the behaviour function for all combinations of H and $H+1$. As a result, whenever $\hat{\theta}(\Gamma^N|\delta) \geq \hat{\theta}^{FB}$, some rule of this kind is optimal.

For a given salience function $\sigma(\cdot)$ and given probability vectors p_e, p_0, an example is depicted in figure 2.2 for $N = 5$.[16] In particular, the picture shows the probability r, denoted by $r^*(\delta, H)$, which, for the described class of rules then yields optimality for given δ and H. For instance, for the approximate interval $(0.2, 0.54]$, a rule with three

[16]In particular, salience function and probability vectors are the same as used in the numerical example further below. The probability distributions are linear with $\beta_0 = 0$ and $\beta_e = 1$

distinct grades, where q^1 is awarded grade C_1, q^2 is awarded grade C_2 with probability r and C_3 with probability $1-r$ while all other scores are assigned C_3, is optimal.

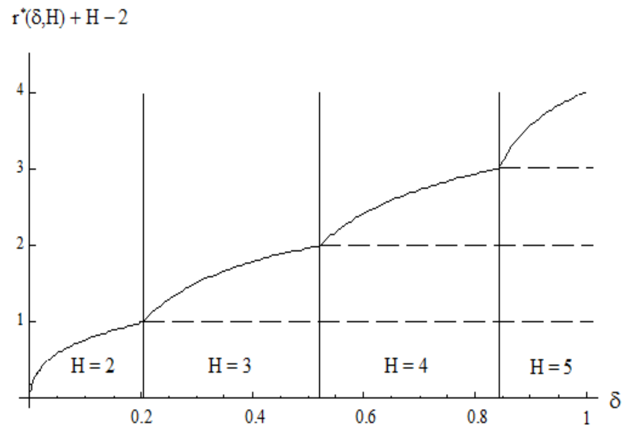

Figure 2.2: An example for optimal rules for $N = 5$

Also, note that probabilistic rules in $\mathcal{G}$ can be constructed such that deterministic rules within a coarseness category are linked. To see how, consider the following example: let there be two rules of the same coarseness category H, and assume they are identical with respect to $H-1$ elements, and differ in one single element where the difference between these two elements is one score, i.e. $1/N$. For instance, let $H = 2$ and consider $\Gamma^{2'} = \{q^n, q^N\}$ and $\Gamma^{2''} = \{q^{n+1}, q^N\}$ Now consider the following probabilistic rule: the score q^{n+1} is assigned grade C_1 with probability r and C_2 with probability $1-r$, while all other scores are graded according to the initial deterministic rules, that is, q^1 to q^n are assigned C_1 and q^{n+2} to q^N are assigned C_2. Denote this new continuous rule $G^2(r)$. Then, student behaviour can be written as a function hereof, i.e. $\hat{\theta}(G^2(r)|\phi,\delta)$ with $\hat{\theta}(G^2(0)|\phi,\delta) = \hat{\theta}(\Gamma^{2'}|\phi,\delta)$ and $\hat{\theta}(G^2(1)|\phi,\delta) = \hat{\theta}(\Gamma^{2''}|\phi,\delta)$ for each tuple (ϕ,δ).

So far, we have found that the focus on salient payoffs can account for the occurrence of coarse schemes and we have derived conditions under which optimal rules exist. However, nothing has been said about how the bias measure relates to optimal standard setting, that is, how strict should the passing standard be for instance when Pass/Fail schemes turn out to be optimal? Also, how does the optimal number of cuts H behave in δ?

In the following, we compute solutions for a salience function proposed by BGS, found to well predict choice behaviour in experimental settings. It is given by $\sigma_{s_{hg}} = (|v_h - v_g|)/(|v_h| + |v_g|)$. For the probability mass functions, both linear and binomial distributions are considered.

When scores are distributed according to binomial distributions $B(N-1,\alpha_0)$ or $B(N-1,\alpha_e)$ respectively, with $\alpha_e, \alpha_0 \in (0,1)$, p_0^n and p_e^n describe the probability that the number of successes is $n-1$. Thereby, an exam is understood to be separable into N subproblems with the probability of one such problem being given by some absolute term which again is determined by the lottery choice.

By contrast, when we speak of linear distributions, we mean that probability mass functions can be represented by a linear function, i.e.

$$p_0^n(\beta_0) = \frac{1}{N}(1+\beta_0\frac{2n-(N+1)}{(N-1)})$$
$$p_e^n(\beta_e) = \frac{1}{N}(1+\beta_e\frac{2n-(N+1)}{(N-1)}),$$

with $\beta_e, \beta_0 \in [-1,1]$. The monotone likelihood ratio property is then fulfilled if and only if $\alpha_e > \alpha_0$ or $\beta_e > \beta_0$ respectively. Since, as has been demonstrated above, simple probabilistic rules can be constructed to link rules within each coarseness category H, results are computed for deterministic rules only.

Results. Figure 2.3 displays $\hat{\theta}(\Gamma^N|\delta)$ for three linear and three binomial distributions and $N = 50$, showing that a fully revealing grading system is optimal only for $\delta = 1$. Therefore, from Proposition 2.3, optimal rules exist for all $\delta \in (0,1]$. Moreover, the share of students choosing lottery $\mathcal{L}_e$ is strictly decreasing in δ. $\delta^0(\cdot)$ denotes a value of δ for which a rule is optimal, that is, $\mathring{\theta}(\Gamma^H|\delta^0) = \hat{\theta}^{FB}$. For Pass/Fail schemes, i.e. $H = 2$, a rule $\Gamma^2 \subset \Psi_2$ is fully described by some $\gamma_1 \in \Omega$. For such schemes and the same distributions, $\delta^0(\Gamma^2)$ is depicted in Figure 2.4. For all these distributions, $\hat{\theta}(\Gamma^2|\hat{\theta}^{FB}, \delta)$ is strictly decreasing in δ, $\delta^0(\Gamma^2)$ thereby being unique. All distributions have in common that Pass/Fail schemes are optimal for δ very close to zero, but never for high δ. In particular, with linear distributions, extreme biases require rules to be extreme, that is, the passing standard should either be very high or very low. Albeit intuitive, this is not an obvious result: with lenient or strict standards, incentives to invest are low, but the decision weight ω_{12} is large. The first effect however turns out to outweigh the second.

On the other hand, when scores are distributed according to binomial distributions, most Pass/Fail schemes induce too few students to employ effort even for extreme biases. Moreover, $\delta^0(\Gamma^2)$ is larger for rules which let fail the expected score for lottery $\mathcal{L}_0$, $\alpha_0 N$, but let pass the expected score for lottery $\mathcal{L}_e$, $\alpha_e N$.

Next, consider rules with two cuts. Figure 2.5 shows graphs of $\delta^0(\Gamma^3)$ for a fixed second cut γ_2. Uniqueness of δ^0 is given, as before. The pictures reveal that for all distributions, these rules outrule both a fully revealing system as well as simple Pass/Fail schemes for intermediate values of δ. The contour plots for δ^0 are presented in the appendix. The range very much depends on the distribution: while for binomial distributions with

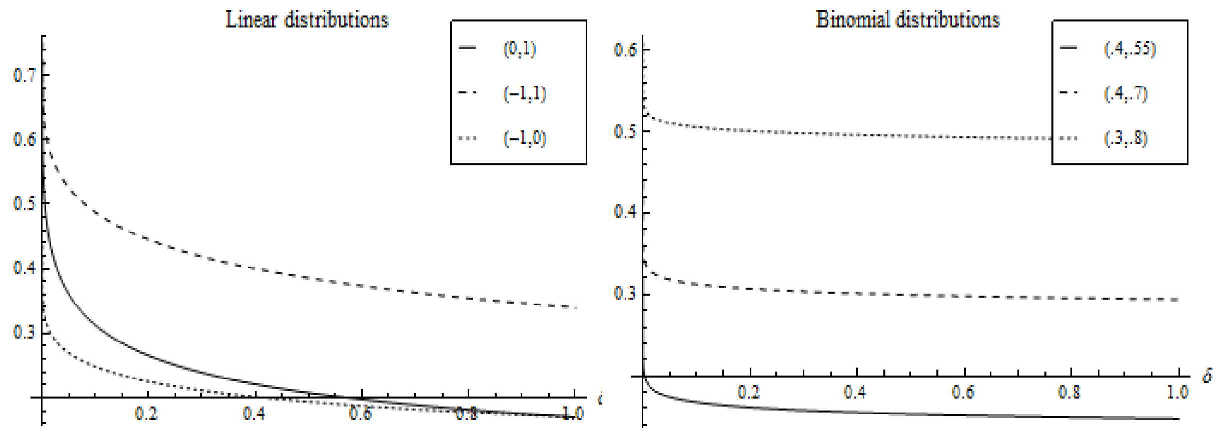

Figure 2.3: Full disclosure: $\hat{\theta}(\Gamma^N|\delta)$ for $N = 50$ and different distributions

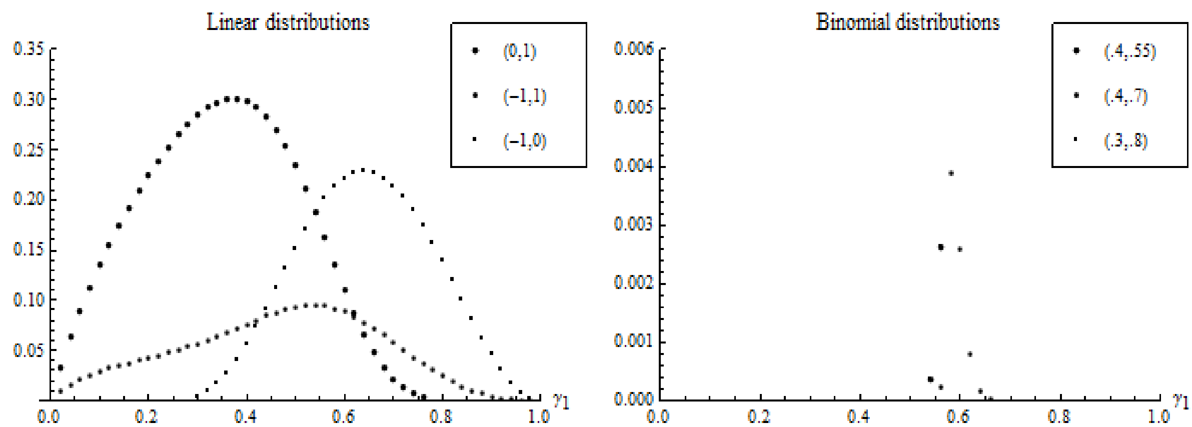

Figure 2.4: Pass/Fail schemes: $\delta^0(\gamma_1)$ for $N = 50$ and different distributions

$\alpha_0 = 0.3$ and $\alpha_e = 0.8$, A, B, C rules are optimal for almost all δ, they are roughly for δ between 0.002 and 0.085 for $(\alpha_0, \alpha_e) = (0.4, 0.55)$. Also, while δ^0 maximizes where each grade contains about the same number of scores when distributions are linear, it does where the number of scores is relatively low in the B grade when distributions are binomial.[17] Rules where the most common grade is an A (γ_1 and γ_2 low) are chosen for rather low values of δ, a result which holds for all distributions.

[17]This is basically the strategy to induce high effort levels given the constraint that there be no more than two cuts. The whole calculation is however biased by the weighting function which exhibits polynomials. This becomes also obvious in the contour plots which show that for some binomial distributions, δ^0 exhibits two local maxima.

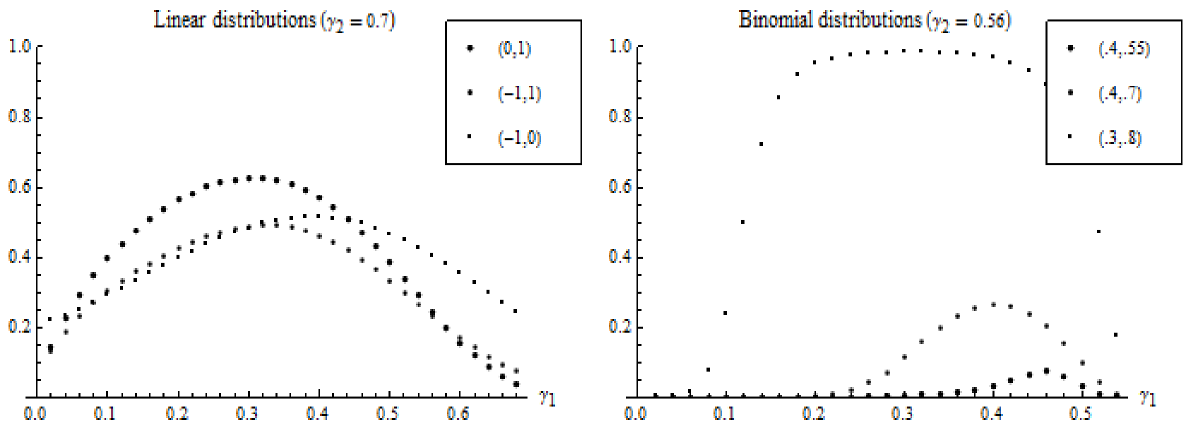

Figure 2.5: A,B,C schemes: $\delta^0(\gamma_1, \gamma_2)$ for given γ_2, $N = 50$ and different distributions

2.6 Discussion

2.6.1 Differentiated students

So far, we have assumed that students of the same institutions are equal with respect to (a) the probabilities distributions and (b) their inherent competitiveness (their δ). In the following, we shortly discuss deviations from these assumptions.

By contrast, consider students of different ability levels. In particular, let there be M student groups $t_1, t_2, ..., t_M$, whose probability distributions have different supports. That is, students of group t_m score between $\underline{q}_m$ and $\overline{q}_m$, $\underline{q}_m < \overline{q}_m$ and $\underline{q}_m, \overline{q}_m \in \Omega$.

Denote ϕ_m the belief employers hold about the critical cost type in groups m, $\hat{\theta}_m$, and denote $\phi^t = (\phi_1, \phi_2, ..., \phi_M)$. The market value of a certificate v_h is given by

$$v_h(\Gamma|\phi^t, \delta) = \frac{\sum_{m=1}^{M} \sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} q^i(\phi_m p_e^{im} + (1-\phi_m)p_0^{im})}{\sum_{m=1}^{M} \sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} (\phi_m p_e^{jm} + (1-\phi_m)p_0^{jm})},$$

where p_e^m and p_0^m are probability vectors for group m with *zero* elements outside the support interval. If students are disparate in the sense that their supports do not overlap, our results from the previous analysis apply accordingly: every group can be analysed separately, and cuts are chosen such that optimality is restored. There is no benefit from further distorting information. As an example, when all groups are optimally incentivized by employing a Pass/Fail scheme, the now optimal rule is one with M cuts.

Similarly, if supports overlap and the optimal cuts are placed outside the overlapping area, the optimal solution can easily be restored. If not so, the decision maker has to trade off the effect of merging the different cuts against that of moving cuts outside the support intervals of students of different abilities. Optimality is then typically not restored. We leave the in-depth analysis of these effects for further research.

A related question is, in how far does the behaviour of differently biased students affect the behaviour their classmates? Assume the school has chosen a coarse rule $\tilde{\Gamma}$ which in equilibrium implemented just the right effort levels given the average δ, denoted by $\tilde{\delta}$. More specifically, $\hat{\theta}(\tilde{\Gamma}|\hat{\theta}^{FB}, \tilde{\delta}) = \hat{\theta}^{FB}$. Consider a single (atomistic) student whose salience bias is lower, namely $\hat{\delta}$, therefore $\hat{\theta}(\tilde{\Gamma}|\hat{\theta}^{FB}, \hat{\delta}) < \hat{\theta}^{FB}$.

Consider by contrast an institution where the average δ is $\hat{\delta}$ but which has employed the same, then non-optimal, rule $\tilde{\Gamma}$. It can be verified that market values for each grade are lower, the reason being that employers foresee the reduced willingness to work. As a result, our analysis predicts that for a given grading scheme, less focused students employ more effort in the presence of biased classmates.

2.6.2 Concluding comments

By presenting a model in which a decision maker overweights salient payoffs, BGS account for a range of identified basic violations of expected utility theory. This study applies their approach to a simple educational model. In particular, an institution seeking to maximize its students' expected utility designs a grading system that does so when students are drawn to salient payoffs associated with grades. Our results suggest that this focus induces too much effort in students, from the institutions' perspective, and the more so the stronger the bias. The common intuition that coarsening reduces incentives applies.

By assuming that this 'focus on grades' is present first and foremost in student bodies of very selective colleges, the model provides an explanation for the occurrence of information suppression in these institutions. Further, the numerical analysis suggests that within one coarseness category, choosing standards such that most students receive $A's$ is a best response to large distortions.

Our results also account for different conceptions of desired states: for instance, consider schools that wish to maximize tuition fees. If students are aware of their bias, and institutions compete for students, employing the grading rule that maximizes overall student welfare is an equilibrium outcome.

2.7 Appendix

2.7.1 Proofs

Proof of Lemma 2.1. It is to show that $W^{FD}(\mathring{\theta}(\Gamma)) = W^{\Gamma}(\mathring{\theta}(\Gamma))$. Denote $E[q^n|\mathcal{L}_e] = V_e^{FD}$ and $E[q^n|\mathcal{L}_0] = V_0^{FD}$. Then $W^{FD}(\hat{\theta}) = W^{\Gamma}(\hat{\theta})$ is given if and only if

$$\hat{\theta}V_e^{FD} + (1-\hat{\theta})V_0^{FD} = \hat{\theta}V_e^{\Gamma}(\hat{\theta}) + (1-\hat{\theta})V_0^{\Gamma}(\hat{\theta}). \tag{2.6}$$

Using V_e^{Γ} and V_0^{Γ} as given in equation (2.1) and v_h as in equation (2.3) transforms equation (2.6) to

$$\sum_{i=1}^{N} q^i(\hat{\theta}p_e^i + (1-\hat{\theta})p_0^i) = \sum_{h=1}^{H} \frac{\sum_{i=\gamma_{h-1}N+1}^{\gamma N} q^i(\phi p_e^i + (1-\phi)p_0^i)}{\sum_{i=\gamma_{h-1}N+1}^{\gamma N} \phi p_e^i + (1-\phi)p_0^i} \sum_{i=\gamma_{h-1}N+1}^{\gamma N} \hat{\theta}p_e^i + (1-\hat{\theta})p_0^i.$$

Since $\mathring{\theta}(\Gamma) = \phi$, the condition becomes

$$\sum_{i=1}^{N} q^i(\phi p_e^i + (1-\phi)p_0^i) = \sum_{h=1}^{H} \sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} q^i(\phi p_e^i + (1-\phi)p_0^i),$$

which is clearly given. □

Proof of Proposition 2.1. Define

$$\begin{aligned}\Delta^{\Gamma^H} :=& \hat{\theta}^r(\Gamma^H|\phi) - \hat{\theta}^{FB} \\ =& \sum_{h=1}^{H} v_h(\Gamma^H|\phi) \sum_{i=1+\gamma_{h-1}N}^{\gamma_H} (p_e^i - p_0^i) - \sum_{n=1}^{N} q^n(p_e^n - p_0^n)\end{aligned}$$

It is to show that $\Delta^{\Gamma^H} < 0$ for all $H < N$ and $\Gamma^H \subset \Psi_H$. The proof is by induction. First step: consider some rule $\tilde{\Gamma}^{N-1} \subset \Psi_{N-1}$ such that grade C_t pools scores q^t and q^{t+1}.

$$\Delta^{\tilde{\Gamma}^{N-1}} = (v_t - q^t)(p_e^t - p_0^t) - (q^{t+1} - v_t(p_e^{t+1} - p_0^{t+1}))$$

Defining $z^n = (\overline{\phi}p_e^n + (1-\overline{\phi})p_0^n)$ and using v_t as given in equation (2.3) gives

$$\begin{aligned}\Delta^{\tilde{\Gamma}^{N-1}} \frac{z^t + z^{t+1}}{q^{t+1} - q^t} =& (z^{t+1}(p_e^t - p_0^t) - z^t(p_e^{t+1} - p_0^{t+1})) \\ =& p_0^{t+t}p_e^t - p_0^t p_e^{t+1} = p_0^t p_0^{t+1}(\tau_t - \tau_{t+1}) < 0.\end{aligned}$$

Second step: consider some $\Delta^{\tilde{\Gamma}^H} \subset \Psi_H$ and assume $\Delta^{\tilde{\Gamma}^H} < 0$. Now consider some $\Delta^{\tilde{\Gamma}^{H-1}} \subset \Psi_{H-1}$ which pools two neighbouring grades $C_{t'}$ and $C_{t'+1}$ into a new grade C_s, all other grades remain unchanged.

$$\begin{aligned}\Delta^{\tilde{\Gamma}^{H-1}} - \Delta^{\tilde{\Gamma}^H} =& \hat{\theta}^r(\Gamma^{\tilde{H}-1}|\hat{\theta}^{FB}) - \hat{\theta}^r(\Gamma^{\tilde{H}}|\hat{\theta}^{FB}) \\ =& (v_s - v_{t'}) \sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} (p_e^i - p_0^i) - (v_{t'+1} - v_s) \sum_{i=1+\gamma_{t'}N}^{\gamma_{t'+1}N} (p_e^i - p_0^i)\end{aligned}$$

with

$$\begin{aligned}v_s - v_{t'} =& \frac{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j) + \sum_{i=\gamma_{t'}N+1}^{\gamma_{t'}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i \sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i} \\ =& \frac{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i \sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i} \\ v_{t'+1} - v_s =& \frac{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i z^j (q^i - q^j) + \sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i \sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i} \\ =& \frac{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i \sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i}\end{aligned}$$

Therefore, $\Delta^{\tilde{\Gamma}^{H-1}} - \Delta^{\tilde{\Gamma}^H} < 0$ if and only if

$$\begin{aligned}& \frac{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i} \Big(\frac{\sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} (p_e^i - p_0^i)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i} - \frac{\sum_{i=1+\gamma_{t'}N}^{\gamma_{t'+1}N} (p_e^i - p_0^i)}{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i} \Big) < 0 \\ \Leftrightarrow & \Big(\frac{\sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} (p_e^i - p_0^i)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i} - \frac{\sum_{i=1+\gamma_{t'}N}^{\gamma_{t'+1}N} (p_e^i - p_0^i)}{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i} \Big) < 0 \\ \Leftrightarrow & \sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} \sum_{j=1+\gamma_{t'}N}^{\gamma_{t'+1}N} p_0^j p_e^i - p_0^i p_e^j = \sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} \sum_{j=1+\gamma_{t'}N}^{\gamma_{t'+1}N} p_0^i p_0^j (\tau_i - \tau_j) < 0,\end{aligned}$$

which is given due to the monotone likelihood ratio property. □

Proof of Lemma 2.2. The problem can be written as a maximization problem where p_0 is being fixed. It is to show that there the maximium is not strictly positive. For technical convenience, we consider $\tau_n = 0$, i.e. $p_0^n = 0$ for some small n and $\tau_n = \tau_{n+1}$ for some n (in contradicion to Assumption (2.1) where we assumed strictly increasing τ_n).

$$\max_{\tau} \; f(\tau) := -(N-1)\frac{\tau_N - \tau_1}{\tau_N + \tau_1} + \sum_{n=1}^{N} n p_0^n (\tau_n - 1)$$
$$s.t. \; \tau_n \geq \tau_{n-1} \geq 0 \text{ for all } n, \; \sum_{n=1}^{N} p_0^n \tau_n = 1$$

τ_N can be written as a function of τ_{-N}. Then, for every $2 \leq n \leq N-1$,

$$\frac{\partial f(\tau_{-N})}{\partial \tau_n} = p_0^n (n - N + \frac{2(N-1)\tau_1}{(\tau_N + \tau_1)^2 p_0^N})$$

Possible solutions to the program are:
(a) $\tau_1 \to 0$. Define $E_e^n := \sum_{n=1}^{N} n p_e^n = \sum_{n=1}^{N} n p_0^n \tau_n$ and $E_0^n := \sum_{n=1}^{N} n p_0^n$. Then: $\lim_{\tau_1 \to 0} = -(N-1) + (E_e^n - E_0^n) < 0$.
(b) $1 > \tau_1 > 0$. Since $\partial f(\tau_{-N})/\partial \tau_n$ is monotonically increasing in n, a candidate solution has the property that there exists some critical n, denoted by $\hat{n}$ such that $\tau_{\hat{n}}^1 := \tau_1 = \tau_2 = ... = \tau_{\hat{n}} < \tau_{\hat{n}+1} = ... = \tau_N := \tau_{\hat{n}}^N$. Denote $p^{\hat{n}} := \sum_{i=\hat{n}+1}^{N} p_0^i$.
Then $\tau_{\hat{n}}^N = (1 - (1 - p^{\hat{n}})\tau_{\hat{n}}^1)/p^{\hat{n}}$ and

$$f(\tau) = -\frac{(N-1)(1-\tau_{\hat{n}}^1)}{p^{\hat{n}}(\tau_{\hat{n}}^N + \tau_{\hat{n}}^1)} - (1-\tau_{\hat{n}}^1)\sum_{i=1}^{\hat{n}} i p_0^i + (1-\tau_{\hat{n}}^1)\frac{1-p^{\hat{n}}}{p^{\hat{n}}} \sum_{i=\hat{n}+1}^{N} i p_0^i.$$

$f(\tau) < 0$ iff

$$-(N-1) + (\tau_{\hat{n}}^N + \tau_{\hat{n}}^1)(1-p^{\hat{n}}) \sum_{i=\hat{n}+1}^{N} i p_0^i - (\tau_{\hat{n}}^N + \tau_{\hat{n}}^1) p^{\hat{n}} \sum_{i=1}^{\hat{n}} i p_0^i < 0.$$

It holds that $\sum_{i=1}^{\hat{n}} i p_0^i > (1 - p^{\hat{n}})$ and $\sum_{i=\hat{n}+1}^{N} i p_0^i < N p^{\hat{n}}$. Therefore, $f(\tau) < 0$ if

$$1 - (\tau_{\hat{n}}^N + \tau_{\hat{n}}^1)(1-p^{\hat{n}})p^{\hat{n}} > 0.$$
$$\Leftrightarrow 1 > (1-p^{\hat{n}})(1 + \tau_{\hat{n}}^1(2p^{\hat{n}} - 1))$$

which can be checked to hold for all $\tau_{\hat{n}}^1 \in (0,1)$ and $p^{\hat{n}} \in (0,1)$.
(c) $\tau_1 \to 1$ yields $\lim_{\tau_1 \to 1} f(\tau) = 0$. □

Proof of Proposition 2.2. From the analysis in the text, it follows that for any Γ^H, $H \geq 3$ with $\gamma_1 = q^1$ and $\gamma_{H-1} = q^{N-1}$, it holds that $\omega_{s_{1H}} = \omega_{s_{H1}} = (p_0^1 p_e^N + p_0^N p_e^N)^{-1}$ and

$\omega_{s_{hg}} = 0$ for all other states, thereby:

$$\lim_{\delta \to 0} \hat{\theta}(\Gamma^H|\phi, \delta) = \lim_{\delta \to 0} \hat{\theta}(\Gamma^N|\phi, \delta),$$

which is strictly greater than $\hat{\theta}^{FB}$ by Lemma 2.2. From Proposition 2.1, these rules are strictly smaller than $\hat{\theta}^{FB}$ at $\delta = 1$. By continuity of $\hat{\theta}(\Gamma^H|\phi, \delta)$ in δ, there exists some δ for which $\hat{\theta}(\Gamma^H|\phi, \delta) = \hat{\theta}^{FB}$. □

Proof of Proposition 2.3. First, if no information is revealed, i.e. $H = 1$, then $\hat{\theta}(\Gamma^1|\phi, \delta) = 0$ for all tuples (ϕ, δ).

Some probabilistic rule in $\mathcal{G}$ establishes continuity between any $\hat{\theta}(\Gamma^1|\phi, \delta)$ and $\hat{\theta}(\Gamma^2|\phi, \delta)$ for any $\Gamma^2 \subset \Psi_2$. Consider the following rule in Ψ_2: $\tilde{\Gamma}^2 = q^1$ and the probabilistic rule in $\mathcal{G}$, $\tilde{G}^2 = \{q^1, q^N, (r, 1-r), (0, 1)\}$ with $r \in (0, 1]$. That is, score q^1 is awarded grade C_1 with probability r and C_2 with probability $1 - r$. All other scores are awarded grade C_2. The function $\hat{\theta}(\tilde{G}^2|\phi, \delta)$ is clearly continuous in r. When $r = 1$, $\tilde{G}^2 = \tilde{\Gamma}^2$.

Also, when $r \to 0$, $\tilde{G}^2 \to \Gamma^1$, and $\hat{\theta}(\tilde{G}^2|\phi, \delta) \to \hat{\theta}(\Gamma^1|\phi, \delta) = 0$, as can be verified by inspection of equation 2.5. Then, all sum elements of equation 2.5 which contain $g = 1$ or $h = 1$ approach 0, while the value of grade C_2 limits the value of C_2 of Γ^0. Further, the probability that grade C_2 occurs if $\mathcal{L}_0$ is chosen, is given by $\sum_{i=q^2N}^{q^N N} p_0^i + (1-r)p_0^1$ (analog presentation for $\mathcal{L}_e$), while the probability that grade C_1 occurs is rp_0^1, which is clearly zero for $r \to 0$. As a result, the weighting functions as given in 2.4 also limit those of Γ^0, i.e. 1.

The same procedure links all higher coarseness categories up to the full disclosure rule Γ^N. More specifically, the new probabilistic rules are all derived from the deterministic rules $\tilde{\Gamma}^H = \{\gamma_1 = q^1, \gamma_2 = q^2, ...\gamma_{H-1} = q^{H-1}, \gamma_H = q^N\}$. Then, the rules $\tilde{G}^H = \{q^1, ..., q^{H-1}, q^N, (1, 0, ..., 0), (0, 1, 0, ..., 0), ...(0, ..., 1, 0, 0), (0, ..., 0, r, 1-r), (0, ..., 1)\}$, $r \in (0, 1]$ establish the link between $\tilde{\Gamma}^{H-1}$ and $\tilde{\Gamma}^H$ for all H. The last step is the transition from $\tilde{\Gamma}^{N-1}$ to the full disclosure rule Γ^N.

By continuity of $\hat{\theta}(\tilde{G}^H|\phi, \delta)$ in r for any tuple (ϕ, δ) and any rule $\tilde{G}^H$, continuity is established between $\hat{\theta}(\Gamma^0|\hat{\theta}^{FB}, \delta)$ - student behaviour when no information disclosure - and $\hat{\theta}(\Gamma^N|\hat{\theta}^{FB}, \delta)$ - student behaviour for full disclosure. As a result, optimal rules exist for all δ for which $\hat{\theta}(\Gamma^N|\hat{\theta}^{FB}, \delta) \geq \hat{\theta}^{FB}$.

□

2.7.2 Numerical results

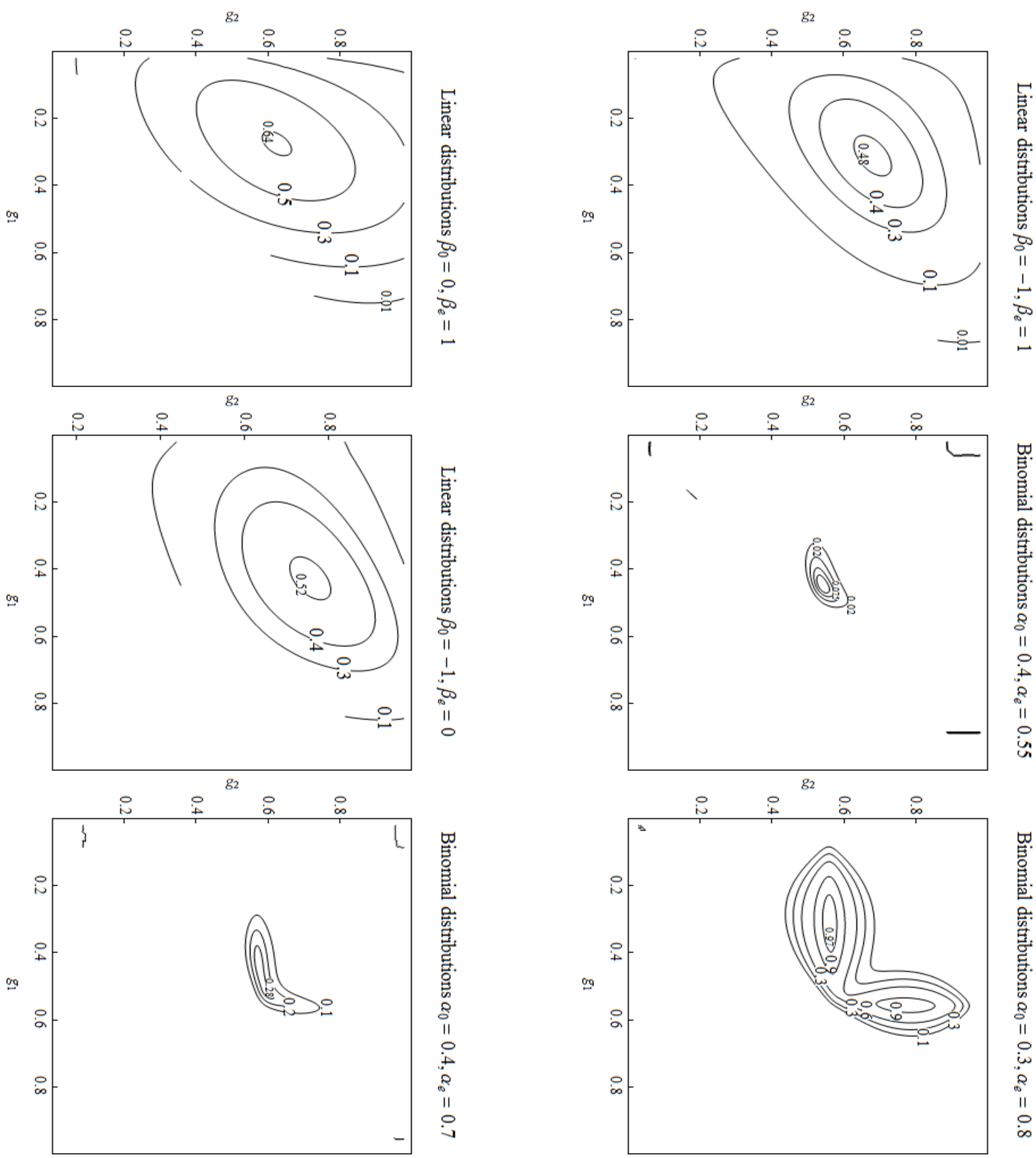

Figure 2.6: A,B,C schemes: $\delta^0(\gamma_1, \gamma_2)$ for $N = 50$ and different (p_0, p_e)

3 Informational opacity and honest certification

Chapter Abstract

This paper studies the interaction of information disclosure and reputational concerns in certification markets. We argue that by revealing less precise information a certifier reduces the threat of capture. Opaque disclosure rules may reduce profits but also constrain feasible bribes. For large discount factors a certifier is unconstrained in the choice of a disclosure rule and full disclosure maximizes profits. For intermediate discount factors, only less precise, such as noisy, disclosure rules are implementable. Our results suggest that contrary to the common view, coarse disclosure may be socially desirable. A ban may provoke market failure especially in industries where certifier reputational rents are low.

3.1 Introduction

In markets that exhibit informational asymmetries, product quality is typically reduced. This in turn may provoke a breakdown of trade. The lack of credible communication between informed and uninformed parties may result in the emergence of certification intermediaries. Certifiers inspect products whose characteristics are private information to agents, and publicly reveal this information. Examples abound: rating agencies, eco-labels, wine certificates or technical inspections. Often however, results are revealed on a coarse scale, although the information at hand would allow for a more precise disclosure.

A stream of literature starting with Lizzeri (1999) has identified profit concerns as the reason for imprecise information disclosure. We provide an alternative explanation for such opacity. We show that partially revealing rules can serve as a safeguard against fraud: certifiers may be tempted to accept bribes for releasing favourable certificates. This behaviour, which we call capture, enables the certifier to extract payments other than the certification fee. If consumers are aware of this threat of capture, then the certifier must find ways to credibly commit to honesty. We show that one way to do so is to employ an opaque disclosure rule. opacity reduces the producer's willingness to pay for bribery, because a more opaque disclosure rule lowers differences in the values of certificates. Hence, opacity can be welfare enhancing since it may prevent market failure. This result is surprising because it contradicts the commonly held view that reducing informational asymmetries is socially desirable per se.

We show our result in a model with moral hazard where, in each period, short-lived producers first have to make an investment choice, which in turn determines the probability distribution of their products' qualities. Thus, the payoffs assigned to each quality outcome determine the incentives to invest. The long-lived certifier has two instruments at his disposal: a flat certification fee and the disclosure rule. Consumers experience the true quality of a product only after consumption. If it does not match the awarded certificate, capture is detected. This makes the certifier face a classical reputation dilemma because he trades off short-run gains from capture against regular future profits.

We characterize feasible disclosure rules in this setting. Our major finding is that for sufficiently low discount factors, honest certification requires partial disclosure of quality information, which in our model implies noisy disclosure. In the short run, a certifier may gain from making a capture offer that is acceptable for at least some producers. The maximum producer willingness to pay for bribes is directly affected by the publicly announced disclosure rule. It is greatest for full disclosure and can be substantially reduced by revealing less precise information. But if consumers detect a bribe and therefore lose trust, a certifier gives up his future profits. Static certifier profits are maximal for full disclosure and any deviation will typically reduce the long-run loss from losing credibility. As will be shown, the first effect exceeds the latter.

We moreover obtain the counterintuitive result that a threat of capture increases social welfare.[1] Whenever information is fully revealed, sharing profits necessarily reduces producer investments as compared to the first-best level, obtained under complete information. We show that whenever capture offers are made before a certifier observes the true quality level, these are such that they are accepted by either all producers or only by low quality producers. If the highest threat of capture stems from offers that are accepted by all producers and the disclosure rule is noisy, credibility can be maintained by making honest certification more attractive to high quality producers. This in turn increases equilibrium investment levels as compared to full information disclosure.

Related literature. A stream of literature seeks to explain why certifiers often choose to only partially reveal quality information. Lizzeri (1999) finds that it is optimal for a monopolistic certifier in a static adverse selection environment to reveal almost no information. In this setting, this result is robust to introducing capture because a no revelation policy simply annihilates producer incentives to bribe. In the presence of moral hazard however, information revelation is necessary to create incentives for the provision of quality. Albano and Lizzeri (2001) study optimal disclosure rules in a static model of both moral hazard and adverse selection. In their setting, a certifier chooses to employ noisy disclosure if his set of actions is restricted to flat fees. According to Farhi et al. (2014), opacity in certification markets is caused by information averse sellers. In Dubey and Geanakoplos (2010), it is shown that coarse grading schemes can help to induce all students to employ effort if they are disparate and care about their status in class. Kartasheva and Yilmaz (2013) explain imprecise ratings in a model with partially informed investors and heterogeneous liquidity needs of issuers. A static adverse selection model where quality is not fully observable by the seller is analysed by Faure-Grimaud et al. (2009). They identify conditions under which the ownership of certification results is left to firms and under which firms reveal their ratings.

The threat of capture in certification markets has been analysed by Strausz (2005). In a pure adverse selection setting with full disclosure, he analyses the effects of a threat of capture on certification prices. He finds that in order to maintain credibility, for low

[1] We analyse a belief system that substantially restricts the set of feasible disclosure rules. For different belief systems and sufficiently low discount factors, other (opaque) rules may be chosen by the certifier. The effect on social welfare is therefore not a general result.

discount factors, a certifier raises fees above the static monopoly price. This result is consistent with our finding in that as less information is disclosed, the certification fee generates a cut-off value that specifies a minimal certified producer quality. A larger fee increases this cut-off but this implies that less information is revealed in equilibrium. Although this effect is also present in Strausz (2005), he however does not explicitly point it out. As in the present paper, credibility is maintained by reducing the maximal willingness to bribe. In Strausz (2005), this is affected by the value of not being certified, which, in turn, is an increasing function of the certification fee.

There is a rich literature on reputation building in markets with informational asymmetries. For example, Shapiro (1983) analyses the forces at work when sellers build reputation. Biglaiser (1993) investigates the role of market intermediaries when sellers are unable to build their own reputation.

Examples of works that treat reputational concerns of rating agencies are Mathis et al. (2009) and Bolton et al. (2012). In contrast to the present paper, these works follow the asymmetric information approach to reputations, where certifiers are assumed to always be committed (i.e. honest) with positive probability.[2] This, however, restricts the set of allowed certification fees and disclosure rules for non-committed certifiers. The reason is that a departure from the equilibrium strategy immediately reveals the certifier type. Instead of assuming that testing by the certifier is imperfect as is done in those works, we show how imperfect testing may endogenously arise in equilibrium.

Levin (2003) extends the standard moral hazard setting to situations where contractual agreements are enforceable only to a certain degree and where reciprocal relations are long-term. The optimal contract derived by Levin has a coarse structure, which parallels our finding of coarse disclosure being optimal.

The remainder of the paper is organized as follows. Section 2 presents the model. Section 3 analyses the static game in the absence of bribery. In section 4, we treat the general case of certification under the threat of capture. Section 5 concludes. All proofs are presented in the appendix.for.

3.2 The model

We consider a dynamic framework in discrete time. In each period $t = 1, 2, \ldots, \infty$, a short-lived monopolistic producer is born. He produces a single unit of quality $q_t \in \{q^l, q^h\}$, where $0 \leq q^l < q^h$. In the following, we refer to a *high type* producer if his product quality is q^h and to a *low type* producer otherwise. Prior to production, a producer chooses

[2]See Mailath and Samuelson (2012, Chapter IV) for a discussion of this approach.

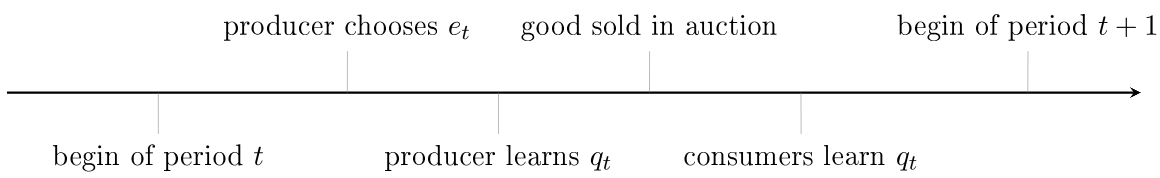

Figure 3.1: Timing in one period without certification

some investment level $e_t \in [0, 1]$. Quality is stochastic and the probability of the produced good being of high quality q^h is given by $Prob(q_t = q^h | e_t) = e_t$. This probability function is independent of t, i.e. quality levels are independent across time. Investment costs are given by the function $k(\cdot)$. We assume $k(\cdot)$ to be increasing and strictly convex. For technical reasons we assume a non-negative third derivative, such that the certifier's profit function is concave. To guarantee interior solutions we additionally assume $k'(0) \leq q^l$ and $k'(1) \geq q^h$.

Consumers' reservation prices equal (expected) qualities. Both investment and quality level are private information to the producer. Consumers observe the product quality only after consumption. All other components of the model are common knowledge. The equilibrium concept we use throughout is the perfect Bayesian equilibrium.

Each producer is short-lived and leaves the market at the end of a period. Goods are sold in a second-price auction.[3] Figure 3.1 summarizes the timing in period t.

To simplify notation, we set $q^l \equiv 0$ and define $v := q^h - q^l$. In the benchmark case with complete information, high quality goods are sold in the second-price auction at price v and low quality goods are sold at price 0. The producer then chooses e to maximize his expected profits $ev - k(e)$. The first-best investment level e^* is thus given by $k'(e^*) = v$. It lies in the interval $[0, 1]$, because $k'(1) \geq v$ by assumption. In particular, $e^* > 0$.

When the market exhibits asymmetric information, a producer cannot persuade consumers that he offers a high quality good. As a result, the market price cannot be made contingent on a good's quality. It is standard to show that the Perfect Bayesian market outcome involves a market breakdown. Consumers form a belief q_t^e about the offered quality reflecting their willingness to pay. In equilibrium, this belief has to be consistent with the actual expected quality, $E(q_t|e_t)$. Given any belief, a producer's optimal investment choice will be $e_t = 0$ as he maximizes $q_t^e - k(e_t)$. But since $E(q_t|0) = 0$, in the unique equilibrium, producers choose $e_t = 0$ in every period. Thus, quality is zero in each period. The result is a market failure: high quality is never offered in equilibrium.

[3] The second price auction results in a standard monopoly price that equals consumers' valuations. It circumvents signaling issues, e.g. letting the informed party take a publicly observed action that might be interpreted as a signal.

We summarize this finding in the following lemma.

Lemma 3.1. *Without certification, producers choose $e_t = 0$ in each period. In equilibrium, quality is given by $q_t = 0$ and the price is 0 in each period.*

This inefficiency calls for the emergence of alternative market institutions to facilitate supply of high quality. The focus of this paper lies on certification as one such institution. Assume that an infinitely long-lived certifier enters the market. She screens quality and offers to disclose the result of some potentially imperfect test, prior to it being sold. More precisely, at the beginning of the game, in period $t = 0$, the certifier announces a fee $f \geq 0$ and a disclosure rule $D = (\mathcal{C}, \alpha^l, \alpha^h)$.[4] [5] The fee has to be paid by any producer who wishes to have his product tested. The disclosure rule consists of a set $\mathcal{C} = \{C^1, \ldots, C^m\}$ of potential certificates and probability vectors α^l and α^h, where the k-th entry of vector α^i indicates the probability that a product of quality q^i is awarded certificate C^k if tested. We do not assume that those probabilities add up to one, i.e. we allow for $\sum_{k=1}^{m} \alpha_k^i < 1$. Hence, a product may remain uncertified with the conforming probability and will be sold as such. We assume that consumers cannot observe whether a product was tested, unless it is offered with a certificate.[6] Possible disclosure rules encompass for example *full disclosure*, where $\mathcal{C} = \{C^1, C^2\}$ and $\alpha^2 = (0, 1)$ as well as $\alpha^1 = (1, 0)$, or *no disclosure*, where $\mathcal{C} = \{C\}$ and $\alpha^i = (1)$.[7] For a given certificate C^k, consumers form a belief $\widetilde{q}^{C^k}$ about the true quality of a product. The belief for uncertified products is denoted $\widetilde{q}^{\emptyset}$. For notational convenience we henceforth add $\emptyset$ to the set of certificates $\mathcal{C}$, which refers to uncertified products. Hence, $\mathcal{C} = \{C^1, \ldots, C^m, \emptyset\}$.

An interpretation of the disclosure rule, which we shall use throughout the paper, is the following: the certifier can create any test that leads to a grading scheme with grades from the set $\mathcal{C}$ and results in the respective grades with conforming probabilities. This may be done by a computer program or a statistical test. In particular, after the test result is obtained, both certifier consumers share the same beliefs on product quality.

[4] Assuming a single fee f, that does not depend on the certificate, is without loss in the setting with only two quality levels. The best a certifier could do is, following the revelation principle, to offer a menu of 'contracts' for the two potential producer types. Eventually, there is one payment referring to the high type and one referring to the low type. It can be easily shown that the optimal contract corresponds to the full disclosure rule, where high types pay f and low types pay 0 and true quality is revealed.

[5] The fee f creates a distortion as will become clear later on. The certifier could implement the first-best outcome, but only when moving first, i.e. when demanding an upfront payment *before* producers choose their investment. This timing however seems unreasonable in many certification markets.

[6] Hence products which "failed" the test are sold under the same label as products that didn't even take the test. This assumption is not crucial, since the certifier can replicate any outcome of a game where consumers are able to observe whether a product applied for certification.

[7] Note that certificates do not carry an intrinsic value. In the case that quality is fully revealed, whether C^1 or C^2 is the valuable certificate depends on the choice of α.

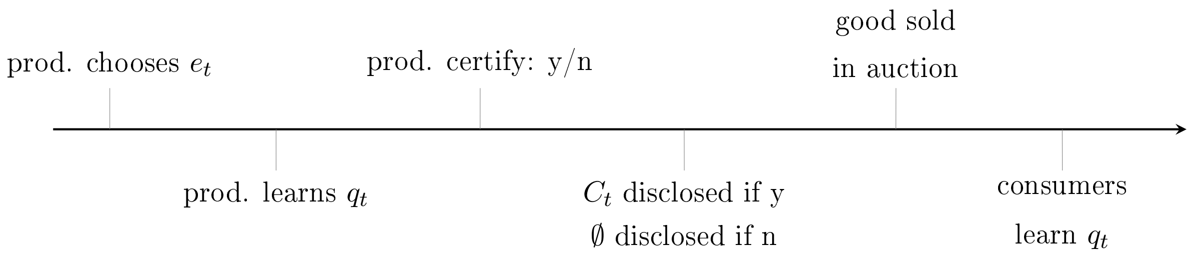

Figure 3.2: Timing of a period t with certification

Finally we assume that the certifier's inspection costs are zero[8] and that she discounts future profits at rate $\delta \in (0, 1)$. Figure 3.2 illustrates the timing of the game with certification.

3.3 Optimal honest certification

In this section, we analyse certifier equilibrium strategies when the certifier is honest. By the stationary structure of the model, we can restrict our analysis to the certifier decision (D, f) plus a single period of production. Let $\pi^D(f)$ denote the equilibrium profit of the certifier, when adopting disclosure rule D with certification fee f.

We first study a full disclosure rule in some detail. It will turn out that this disclosure rule maximizes certifier profits. Consider the case where quality is fully revealed such that $\alpha^h = (1, 0)$ and $\alpha^l = (0, 1)$. Any product that is awarded C^1 is sold at a price v, whereas C^2 is worth nothing. The only plausible equilibrium is one where only high types apply for certification.[9] A producer chooses his investment according to

$$e = arg \max_{\widetilde{e}} \ \widetilde{e} \cdot (v - f) - k(\widetilde{e}). \tag{3.1}$$

This implies $k'(e) = v - f$ in equilibrium and certifier expected equilibrium profits can be expressed as

$$\widehat{\pi}^{FD}(e) = e \cdot (v - k'(e)). \tag{3.2}$$

Denote e^{FD} the equilibrium effort level under a full disclosure rule and f^{FD} the corresponding profit maximizing fee. The following lemma proves that these values exist and are unique.

[8]This assumption simplifies the analysis without substantially affecting the results, which continue to hold as presented here for small but strictly positive inspection costs. High inspection costs do not invalidate most of our results, but create cumbersome case distinctions.

[9]Trivially, low quality producers do not demand certification when $f > 0$ since their revenues are zero at most.

Lemma 3.2. *Under full disclosure, there exists a unique fee f^{FD} that maximizes certifier profits. The uniquely defined equilibrium investment level e^{FD} is implicitly given by*

$$k''(e^{FD}) \cdot e^{FD} = v - k'(e^{FD}). \tag{3.3}$$

The fee is $f^{FD} = v - k'(e^{FD})$ and the (subgame-) equilibrium profit is $\pi^{FD} = e^{FD} \cdot f^{FD}$.

We continue with an analysis of general disclosure rules. The entire set of disclosure rules is complex and difficult to handle analytically. A closer look at equation (3.2), which allows us to express the certifier profit as a function of the implemented investment level e, points to the advantages of using an indirect approach. We can express the attained profit of any certifier policy (D, f) solely in terms of the induced investment level e. This allows for a straightforward comparison of attained profits and leads us to the following proposition.

Proposition 3.1. *For any disclosure rule $D = \left(\mathcal{C}, \alpha^1, \alpha^0\right)$ and any fee $f \geq 0$, it holds that $\pi^D(f) \leq \pi^{FD}$ in equilibrium.*

Proposition 3.1 states that the certifier will always find it optimal to employ a full disclosure rule. The reason is that investment incentives depend on the difference between payoffs from selling high and low quality products. Given full disclosure, the certification fee is sufficient to fully control this difference.

We complete this section by pointing to the fact that full disclosure is not the unique optimal (i.e. profit maximizing) disclosure rule. First of all, the same outcome can be implemented by adding redundant certificates to various rules – either additional certificates for high types, which then all have the same value in equilibrium, or by adding certificates for low types that will not be issued in equilibrium.

Because certification is assumed to be costless for the certifier, other rules also maximize certifier per-period profits: issue two different certificates C^1 and C^2. Low quality products are only eligible for certificate C^2, hence $\alpha^l = (0, 1)$. High quality products receive certificate C^1 with probability $\alpha \in (0, 1)$ and C^2 otherwise, therefore $\alpha^h = (\alpha, 1-\alpha)$. With rules of this structure, it is possible to sustain equilibria in which all producer types apply for certification.[10] The optimal certifier profit π^{FD} can be obtained by choosing f and α accordingly.[11] This class of disclosure rules plays a crucial rule in the remainder of this paper. We henceforth refer to these as *partial disclosure* rules.

[10] For this, we have to set the off-equilibrium belief $\widetilde{q}^{\emptyset} = 0$ and all other beliefs underly Bayesian updating.

[11] We formally show this in the proof of Proposition 3.6.

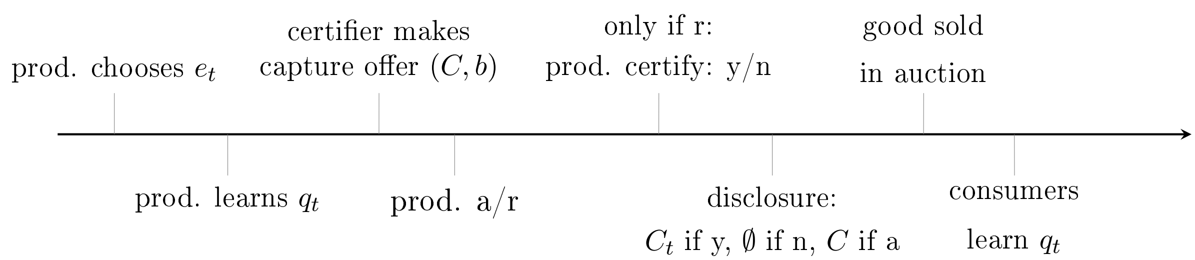

Figure 3.3: Timing of a period t with certification and capture

3.4 The capture problem

So far we have assumed that the certifier sticks to the previously announced disclosure rule. This implies that she conducts the lottery honestly and grants the respective certificate. However, there is pressure from producers who wish to be awarded better certificates. For instance, if disclosure is meant to be noisy, a certifier might be willing to guarantee a producer a high value certificate in exchange for a bribery payment. In this section we address this issue by formally introducing the possibility of capture.

We follow Strausz (2005) in modelling the threat of capture, using the framework of enforceable capture as initiated by Tirole (1986). Hence we assume that the certifier and the producer can write an enforceable side-contract with transfers. Consumers are fully aware of the possibility of these side-contracts, but cannot observe them. Specifically, we model capture as follows: after a producer has learned his type q_t, but before deciding whether to apply for certification, the certifier, without having observed q_t, may make an offer (C, b) to the producer.

The offer consists of a certificate C, issued in case of acceptance, and a financial transfer b to be paid by the producer. The certifier thus offers to "sell" the sure certificate C at the price b, circumventing the customary certification procedure. Hence, (C, b) are the terms at which she is willing to become captured. A producer however can reject this offer and insist on being honestly certified by paying the fee f. This assumption is motivated following Kofman and Lawarrée (1993) who argue that the certifier cannot forge certification without the help of the producer. Figure 3.3 displays the timing in a representative period t, allowing for the possibility of capture.

Note that the choice of disclosure rule puts some limits on the set of *feasible* capture offers. For a general disclosure rule $D = \{\mathcal{C}, \alpha\}$ only offers of the form (C, b) with $C \in \mathcal{C}$

are feasible.[12]

Within the framework presented here, capture may subvert honest certification for two reasons.[13] First, producers with low quality products are willing to side-contract with the certifier in order to obtain better certification. Second, high types may want to avoid uncertainty if disclosure is noisy.

In this section, we are interested in the existence and characterization of equilibria where the certifier resists the temptation of making a capture offer. Throughout, we will work with different specifications of trigger beliefs. This becomes necessary as the ability of consumers to detect capture varies across disclosure rules. We assume consumers to be able to perfectly observe quality after consumption. Therefore, if D is full disclosure or if certain certificates are awarded exclusively to high types, capture detection is also perfect.

In particular, when speaking of consumer trigger beliefs we mean the following: they stop trusting the certifier immediately as soon as a false testimony about a product's quality is detected. Knowing this, producers are not willing to pay for certification anymore. Consequently, the certifier will lose future demand and makes zero profits henceforth. This prevents the certifier from becoming captured in the first place. We shall make this more precise in the following subsections.

3.4.1 Capture under full disclosure

Because, by Proposition 3.1, a certifier wishes to employ full disclosure whenever possible, we start by investigating capture under a full disclosure rule. We assume that consumers trust certificates as long as no deviation has been detected. A certifier who anticipates this behaviour may be prevented from succumbing to the temptation of becoming captured by the fact that losing credibility will leave her without demand in future periods.

Denote $h_t = (C_t, q_t)$ the certification outcome in period t, where C_t is the issued certificate in period t and q_t is the true quality observed after consumption. If certification in period t did not take place, then $C_t = \emptyset$. Now let $H_t = (h_1, \ldots, h_{t-1})$ summarize the history of certification at the beginning of period t. Finally, we denote $\widetilde{q}_t(C_t, H_t)$ a consumer's belief in period t when faced with a product carrying certificate C_t and when having observed history H_t. The following assumption on consumer beliefs formalizes the

[12] This will be made more precise when formally introducing consumer beliefs. Granting a certificate which is not contained in D is certainly perceived as cheating by consumers. Consequently consumers believe to be faced with a worthless product and they will lose trust in the certifier's credibility.

[13] When certification is costly for the certifier, there is a third reason: saving certification cost! As already mentioned in Footnote 8 our analysis can be extended to $c > 0$, but this involves some troubling case-by-case distinctions.

described behaviour.[14]

Assumption 3.1. *Consumer beliefs* $\widetilde{q}_t(C_t, H_t)$ *satisfy* $\widetilde{q}_t(C_t, H_t) = \widetilde{q}^{C_t}$ *whenever* $\{\tau < t | q^{C_\tau} \neq q_\tau \vee C_\tau \notin \mathcal{C} \cup \{\emptyset\}\} = \emptyset$. *Moreover* $\widetilde{q}_t(C_t, H_t) = 0$ *whenever* $\{\tau < t | \widetilde{q}^{C_\tau} \neq q_\tau \ \vee \ C_\tau \notin \mathcal{C} \cup \{\emptyset\}\} \neq \emptyset$ *and* $\widetilde{q}_t(C_t, H_t) = 0$ *whenever* $C_t \notin \mathcal{C}$.

The assumption states that consumers trust the certifier if capture was not observed in the past. They however lose trust forever, once cheating is detected. Losing trust implies that consumers, whatever the certifier's claim may be, believe in being faced with a low quality product.

With full disclosure, there are (at most) two types of bribing offers that can be made: (C^1, b) and (C^2, b). Obviously, an offer (C^2, b) is turned down by all types of producers, as it is worth nothing. Hence, in the following we focus on offers (C^1, b) and talk of a bribe b rather than (C^1, b). An offer b is accepted by high producer types whenever $b < f$. Low quality producers accept any bribe $b < v$ because acceptance will yield positive profits as compared to zero profits for a rejection. In equilibrium, the certifier assigns probability $e(f)$ to the event that a producer is of high type, where $e(f)$ is the producer's optimal investment under full disclosure, derived from (3.1). We are interested in equilibria where capture does not occur. In all these equilibria, a producer chooses his optimal investment level while being aware of the fact that he will not receive an acceptable capture offer. The acceptance probability $p(b|f)$ of a bribing offer b given the a certification fee f is given by

$$p(b|f) = \begin{cases} 1, & \text{if } b < f, \\ 1 - e(f), & \text{if } f \leq b < v, \\ 0, & \text{if } b \geq v. \end{cases} \tag{3.4}$$

We denote by $\Pi^D(f) = \sum_{t=1}^{\infty} \delta^{t-1} \pi^D(f) = \pi^D(f)/(1-\delta)$ the certifier's expected profit from honest certification under disclosure rule D and fee f. The certifier's expected profit from offering bribe b is denoted by $\widehat{\Pi}^D(b|f)$ and depends on whether the consumer detected capture as follows: whenever $b < f$, all producer types will accept the bribe, but only for low quality producers this is detected. Hence, $\widehat{\Pi}^{FD}(b|f) = b + e(f)\delta\Pi^{FD}(f)$. For $f \leq b < v$, only low quality producers accept the bribe and $\widehat{\Pi}^{FD}(b|f) = (1 - e(f))b + e(f)(f + \delta\Pi^{FD}(f))$. Whenever $b \geq v$, all producers reject the bribe and the certifier obtains $\widehat{\Pi}^{FD}(b|f) = \Pi^{FD}(f)$.

If $\widehat{\Pi}^{FD}(b|f)$ exceeds $\Pi^{FD}(f)$, the certifier is actually better off becoming captured

[14] Note that consumers do not lose trust in the certifier when a product is awarded certificate C^2, although this should not happen in equilibrium. It is not necessary to include this case into consumer beliefs, because any such event can only follow a non-profitable deviation by the certifier.

with the associated probability $p(b|f)$. We say that certification at a fee f is *capture proof* if and only if

$$\Pi^{FD}(f) \geq \widehat{\Pi}^{FD}(b|f) \tag{3.5}$$

for all b. Note that $\widehat{\Pi}^{FD}(b|f)$ is increasing in b, both on $[0, f)$ and $[f, v)$ and it is constant for $b \geq v$. Furthermore $\widehat{\Pi}^{FD}(\cdot|f)$ is continuous at $b = f$.[15] Therefore, certifier profits from bribery are largest when b approaches v. Evaluating this yields the following proposition:

Proposition 3.2. *Under a full disclosure rule, an equilibrium satisfying Assumption 3.1 is capture proof. It exists if and only if*

$$\delta \geq \delta^{FD}(f) \equiv \frac{v}{v + \pi^{FD}(f)}. \tag{3.6}$$

The proposition highlights the crucial role the discount factor plays for the existence of honest, i.e. capture proof, equilibria: the critical discount factor determines the relative weights of the short run gain – the bribe b – and the long run loss of capture – foregone future profits from certification. To see this, note that all bribes $b < v$ are accepted with some positive probability and therefore, the largest possible short-run gain equals v. In the long run, a certifier risks her per-period profits $\pi^{FD}(f)$. As the certification fee only enters via the per-period profit, $\delta^{FD}(f)$ depends on f only through $\pi^{FD}(f)$, which is concave in f. Therefore $\delta^{FD}(f)$ must be convex in f and minimized at the profit maximizing fee f^{FD}. The following corollary summarizes.

Corollary 3.1. *For any discount factor $\delta \geq \delta^{FD}$ there exists an interval of fees $[f_l(\delta), f_h(\delta)]$, which sustains capture-proof certification under full disclosure, where*

$$\delta^{FD} \equiv \frac{v}{v + \pi^{FD}}. \tag{3.7}$$

In the right part of Figure 3.4 the set of feasible (δ, f)-combinations for full disclosure is depicted.

An immediate consequence from this is that the static monopoly fee f^{FD} can sustain honest certification for all discount factors $\delta \geq \delta^{FD}$. Alternatively one might ask the question, what level of producer investment can be implemented via capture-proof certification with a full disclosure rule? The analysis follows the same arguments as above, only that certifier profits in the inequality of Proposition 3.2 are expressed in terms of e.

[15]To see this compare the left and right limit: $\lim_{b \nearrow f} \widehat{\Pi}^{FD}(b|f) = f + e(f)\delta\Pi^{FD}(f) = (1 - e(f))f + e(f)\big(f + \delta\Pi^{FD}(f)\big) = \lim_{b \searrow f} \widehat{\Pi}^{FD}(b|f)$.

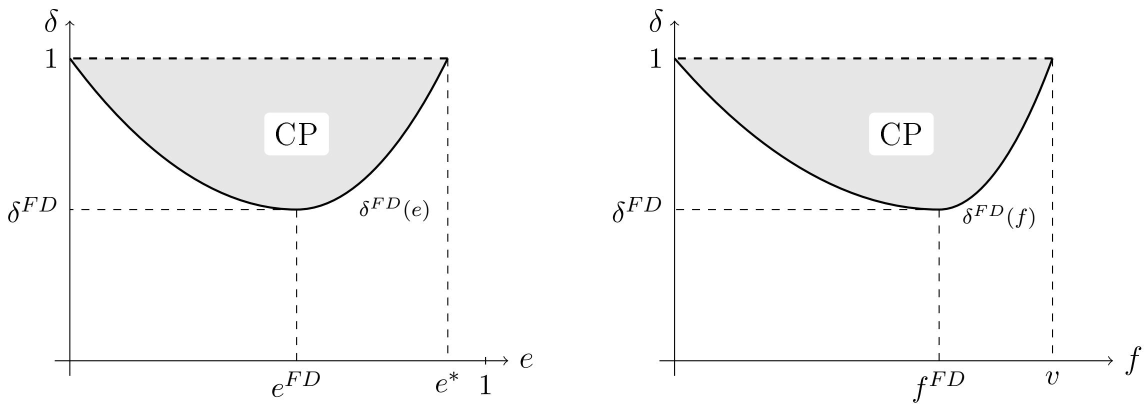

Figure 3.4: Capture proof combinations of (e, δ) resp. (f, δ) under full disclosure.

Proposition 3.3. *For any $\delta \geq \delta^{FD}$ there exists an interval of investment levels $[e_l^{FD}(\delta), e_h^{FD}(\delta)]$ that can be implemented in a capture-proof equilibrium. A particular investment level $e \in [0, e^*]$ can be implemented in a capture-proof equilibrium with full disclosure if and only if*

$$\delta \geq \delta^{FD}(e) \equiv \frac{v}{v + e \cdot (v - k'(e))} \tag{3.8}$$

.

The set of feasible (e, δ)-combinations is depicted in the left part of Figure 3.4. Note that the first-best investment level e^* can only be (virtually) implemented for $\delta = 1$. Whenever $\delta < 1$, fees must be strictly positive in order to induce the certifier to remain honest. But then, the producer does not obtain the entire return on his investment. Hence, it must be that $e < e^*$.

3.4.2 Capture under partial disclosure

We next argue that alternative noisy disclosure rules can improve certifier credibility in the sense that they increase the range of discount factors that allow for capture-proof equilibria.

To gain an intuition for this consider condition (3.6). This condition summarizes the trade-off between short-run gains and long-run losses. A larger profit $\pi^D(f)$ reduces the critical discount factor and full disclosure guarantees maximal per-period profits. On the other hand, $\delta^{FD}(f)$ is decreasing in v, which represents the the maximal bribe still accepted by low-type producers and therefore the largest possible short-run gain from capture. Using noisy disclosure the certifier can affect the maximal short-run gain in various dimensions. First of all, lowering the value of the best certificate or increasing the value of the worst certificate (resp. the value of uncertified products) decreases the gap between particular certification outcomes. This effect can be used to reduce the maximal bribe which producers are willing to pay. Second, with noisy disclosure the certifier can sustain an outcome where both producer types demand certification. Upon colluding with a producer type the certifier foregoes the regular certification fee, which reduces the effective gain from becoming captured.

Before analysing noisy disclosure rules, we have to reconsider the detection possibilities by consumers. An implication of noisy rules is that consumers may hold probabilistic beliefs about a product's quality. In order to simplify matters and because it suffices to make our point clear, we focus on partial disclosure rules as introduced in section 3.3. Other noisy disclosure rules are discussed in section 3.5 and in the appendix. Under partial disclosure, there are again two certificates C^1 and C^2, where certificate C^1 is awarded exclusively to high quality products and C^2 is awarded to high quality sellers with probability $1-\alpha$ and to low quality sellers with probability 1. With an appropriately chosen fee f, all producer types demand certification, hence there are no uncertified products in equilibrium. The corresponding off-equilibrium belief is $q^{\emptyset}=0$. The fact that C^1 is awarded exclusively to high quality products makes effective trigger punishment possible. In particular, it then suffices that the certifier is punished only if probability zero events (a low quality product was awarded certificate C^1) are observed.

The fact that capture detection is not possible if bribes are being paid in exchange for the low value certificate C^2, which is assigned to both high and low types, turns out not to be crucial. This relies on the fact that in the equilibria under consideration all producer types demand certification, hence receiving certificate C^2 is the worst possible outcome. Certificate C^2 can therefore not be part of a profitable bribing offer, as we will argue later.

To specify consumer beliefs, let $h_t=(C_t,q_t)$ denote the certification outcome in period t and, as before, $H_t=(h_1,\dots,h_{t-1})$ describes the history of certification before period t. Consumer beliefs are specified as follows

Assumption 3.2. *Consumer beliefs* $\widetilde{q}_t(C_t,H_t)$ *satisfy* $\widetilde{q}_t(C_t,H_t)=\widetilde{q}^{C_t}$ *whenever* $\{\tau<t|Prob(C=C_\tau|q=q_\tau)=0\ \vee\ C_\tau\notin\mathcal{C}\cup\{\emptyset\}\}=\emptyset$. *Moreover* $\widetilde{q}_t(C_t,H_t)=0$ *when either* $C_t\notin\mathcal{C}$ *or* $\{\tau<t|Prob(C=C_t|q=q_t)=0\ \vee\ C_\tau\notin\mathcal{C}\cup\{\emptyset\}\}\neq\emptyset$.

Note that in contrast to Assumption 3.1, consumers trust the certifier unless probability zero events occurred in the past. Because the crucial bribe entails certificate C^1, which is exclusively awarded to high quality producers, this essentially says that consumers stop trusting the certifier whenever they find a low quality product carrying certificate C^1.

Bribing offers can now be of two kinds: (C^1,b) and (C^2,b). Offer (C^2,b) is never beneficial. It would only be accepted for $b<f$ since any producer receives at least the

certificate C^2 when applying for (honest) certification and the certifier gets f from any producer who is honestly tested. Thus, we can focus on bribing offers of the form (C^1, b), which we will simply refer to as b. Recall that certificate C^1 can only be awarded to high quality products. Hence, $q^{C^1} = v$. To simplify notation, denote V_2 the value of a C^2-certified product, i.e. $V_2 = q^{C^2}$. Furthermore, recall that α is the probability with which a high type is awarded C^1.

A bribe b is accepted by low types whenever $V_2 - f < v - b$. High quality producers accept b if $\alpha v + (1-\alpha)V_2 - f < v - b$. Denote $e(\alpha)$ the equilibrium investment.[16] Then bribery acceptance probabilities are

$$p(b|\alpha, f) = \begin{cases} 1 & \text{if } b < f + (1-\alpha)(v - V_2), \\ 1 - e(\alpha) & \text{if } f + (1-\alpha)(v - V_2) \leq b < f + (v - V_2), \\ 0 & \text{if } b \geq f + (v - V_2). \end{cases}$$

Let $\Pi^{PD}(\alpha, f)$ denote the expected profit from employing a partial disclosure rule and honestly disclosing information in each period. The corresponding expected certifier profits from bribing offer b are

$$\widehat{\Pi}(b|\alpha, f) = \begin{cases} b + e(\alpha)\delta\Pi^{PD}(\alpha, f) & \text{if } b < f + (1-\alpha)(v - V_2), \\ (1 - e(\alpha))b + e(\alpha)\left(f + \delta\Pi^{PD}(\alpha, f)\right) & \text{if } f + (1-\alpha)(v - V_2) \leq b < f + (v - V_2), \\ \Pi^{PD}(\alpha, f) & \text{if } b \geq f + (v - V_2). \end{cases}$$

Note that whenever high types accept the bribery offer, this is not perceived as cheating because the certificate then matches the observed quality level. Again, $\widehat{\Pi}(b|\alpha, f)$ is increasing in the respective subintervals. But the function now exhibits a downward-jump at $b = f + (1-\alpha)(v - V_2)$. The reason is that high types are willing to accept bribes strictly larger than the certification fee f to avoid the lottery between the good and the bad certificate. Therefore, at least locally, the certifier is better off bribing all producers instead of only the low types as it was the case with full disclosure. Furthermore, the maximal bribe that is accepted by at least some types is now $f + v - V_2$, which is weakly lower than under full disclosure, where the maximal bribe is v.[17] The analysis of condition (3.5) yields the following proposition.

[16] The investment decision does not depend on the fee because in equilibrium, all types apply for certification and therefore pay f anyway. The expected producer profit is $e(\alpha V_1 + (1-\alpha)V_2) + (1 - e)V_2 - f - k(e)$ and consequently the optimal investment level depends on α but not on f.

[17] In order to have all producer types demand certification it has to hold that $f \leq V_2$. Consequently $f + v - V_2 \leq v$.

Proposition 3.4. *With partial disclosure, an equilibrium satisfying Assumption 3.2 is capture-proof. It exists if and only if*

$$\delta \geq \delta^{PD}(\alpha, f) \equiv \max\left\{\delta^{l}(\alpha, f), \delta^{l,h}(\alpha, f)\right\}, \tag{3.9}$$

where $\delta^{l}(\alpha, f) = \frac{v-V_2}{v-V_2+f}$ *and* $\delta^{l,h}(\alpha, f) = \frac{(1-\alpha)(v-V_2)}{(1-\alpha)(v-V_2)+(1-e(\alpha))f}$.

The result gives a lower bound on the discount factor δ to guarantee existence of a capture-proof equilibrium with partial disclosure. The critical discount factor $\delta^{PD}(\alpha, f)$ depends on the parameters in the way how they affect short-run gain and long-run loss from capture and on which producer types accept the bribing offer that yields largest deviation profits. The term $\delta^{l}(\alpha, f)$ refers to the case where the largest threat stems from bribes accepted only by low types. The numerator $v - V_2$ is the effective bribe, defined as the bribery payment minus foregone payments. In the denominator we find again the effective bribe and the per-period profit f, reflecting the long-run loss from capture. The term $\delta^{l,h}(\alpha, f)$ refers to the case where the largest threat stems from bribes accepted by all types. Here the effective bribe is $(1-\alpha)(v-V_2)$. Since the long-run profit is only at stake if quality is low, long-run profits are lost with probability $(1-e(\alpha))$. Although the classical trade-off between short-run gain and long-run loss, that we already identified for full disclosure, prevails, the derivation of the maximal short-run gain is more involved for partial disclosure.

From Proposition 3.4 we identify a third notable difference between capture under full and noisy disclosure. Short-run gains from capture can be reduced due to the different equilibrium structure: all producers certify in equilibrium which implies that the certifier always loses fee payments if he is captured. Therefore, a larger fee f not only increases the long-run losses but at the same time reduces the short-run gains from capture.

It is now straightforward to see that $\delta^{PD}(\alpha, f)$ is decreasing in the certification fee f. This implies that for any partial disclosure rule (i.e. any α) the threat of capture is lowest when f is maximal. To ensure that all producer apply for certification, f is not allowed to exceed V_2. It is therefore optimal to set $f = V_2$, which leaves low quality producers with an expected profit of zero. The following corollary summarizes.

Corollary 3.2. *With partial disclosure a capture-proof equilibrium satisfying Assumption 3.2 exists if and only if*

$$\delta \geq \delta^{PD}(\alpha) \equiv \max\left\{\delta^{l}(\alpha), \delta^{l,h}(\alpha)\right\}, \tag{3.10}$$

where $\delta^{l}(\alpha) = \frac{v-e(\alpha)(v-k'(e(\alpha)))}{v}$ *and* $\delta^{l,h}(\alpha) = \frac{1}{1+e(\alpha)}$.

Corollary 3.2 allows us to reduce the problem of finding the critical discount factor for partial disclosure to the one-dimensional problem of finding the optimal level of α, the probability that high quality is revealed. In fact, $\delta^{PD}(\alpha)$ depends on α only through the equilibrium value for producer investment $e(\alpha)$. The set of investment levels that can be implemented by partial disclosure is $(0, e^*)$, the same set as for full disclosure. Defining $\delta^{PD} \equiv \min_\alpha \delta^{PD}(\alpha)$ allows us to formulate the analogue of Proposition 3.3 for partial disclosure.

Proposition 3.5. *For any $\delta \geq \delta^{PD}$ exists an interval of investment levels $[e_l^{PD}(\delta), e_h^{PD}(\delta)]$ that can be implemented in a capture-proof equilibrium. A particular investment level $e \in [0, e^*]$ can be implemented in a capture-proof equilibrium with noisy disclosure if and only if*

$$\delta \geq \delta^{PD}(e) = \max\left\{\delta^{PD,l}(e), \delta^{PD,l,h}(e)\right\} \tag{3.11}$$

where $\delta^{PD,l}(e) = \frac{v - e(v - k'(e))}{v}$ and $\delta^{PD,l,h}(e) = \frac{1}{1+e}$.

Proposition 3.5 makes implementation of capture-proof equilibrium under full and partial disclosure directly comparable. Before investigating this in the next section we want to highlight some properties of the function $\delta^{PD}(e)$. Writing $e(v - k'(e)) = \pi^{PD}(e) = f$ the term $\delta^{PD,l}(e)$ can be expressed as $(v - f)/(v - f + \pi^{PD}(e))$. This resembles the trade-off between short-run gain and long-run loss, already identified above. Only the maximal short-run gain with partial disclosure is the maximal bribe minus foregone regular payments. The same trade-off leads to $\delta^{PD,l,h}(e)$, which is however independent of the producer's cost function $k(e)$. The maximal bribe that is accepted from both producer types in particular must be accepted from high quality producers. For them, the difference between the sure certificate C^1 and the lottery faced when certifying honestly matters. This difference is closely related to a producer's investment incentives. In fact one can show that the maximal bribe equals $v - k'(e)$. Now both short-run gain and long-run loss depend in a similar way on the investment incentives[18] and consequently the fraction $\delta^{PD,l,h}(e)$ does not depend on the producer's cost function anymore.

Which of the two terms, $\delta^{PD,l}(e)$ and $\delta^{PD,l,h}(e)$, is now larger? $\delta^{PD,l,h}(e)$ is decreasing in e, starting at 1 for $e = 0$ towards $1/2$ for $e = 1$. On the other hand $\delta^{PD,l}(e)$ is convex in e with a unique minimum at $e = e^{FD}$. Furthermore $\delta^{PD,l}(0) = \delta^{PD,l}(1) = 1$. Therefore, δ^{PD} is either $\delta^{PD,l}(e^{FD})$, that is the minimum of $\delta^{PD,l}$, or it is the intersection of both fractions lying to the right of $e = e^{FD}$. Figure 3.5 illustrates the two cases, the latter in its left part.

[18] As discussed, the short-run gain equals $v - k'(e)$. The long-run loss is the per-period profit, which was already shown to be $e(v - k'(e))$.

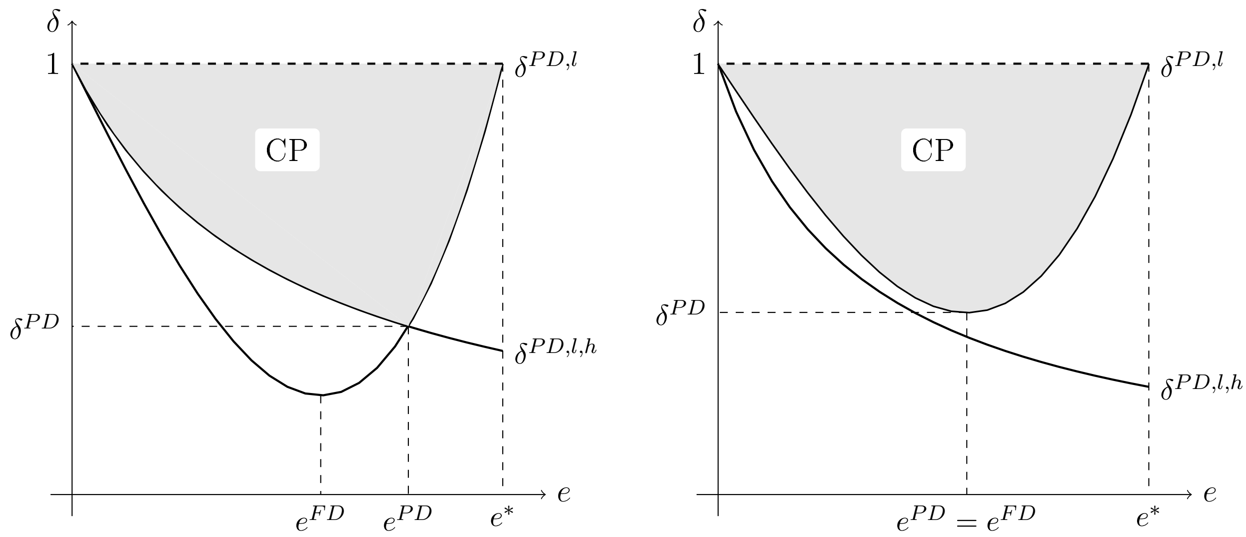

Figure 3.5: Capture-proof (e, δ)-combinations for low (left) and high (right) marginal costs k' at $e = e^{FD}$.

3.4.3 Sub-optimality of full disclosure

In the previous sections, we have identified the conditions under which capture-proof equilibria exist for full disclosure and a special class of noisy disclosure rules. These conditions are expressed in terms of the critical discount factors δ^{FD} and δ^{PD}. It is the aim of this study to show that opaque disclosure rules can be used by the certifier to improve her credibility. Comparing the critical discount factors δ^{FD} and δ^{PD} is short-hand for comparing the entire sets of (e, δ)-combinations, for which a capture-proof equilibrium exists with the respective disclosure rule. We are going to prove in this section that the two sets are different and, more importantly, that the respective set for full disclosure is contained in the respective set for partial disclosure. Consequently there exists an intermediate range of discount factors for which a capture-proof equilibrium with full disclosure does not exist. Yet, it is still possible to sustain capture-proof equilibria with partial disclosure.

As discussed earlier, the key trade-off for implementing a capture-proof equilibrium is that of short-run gain versus long-run loss. Either disclosure rule leads to a per-period profit of $\pi(e) = e(v - k'(e))$ when implementing effort level e, the potential long-run loss is therefore the same. However, with partial disclosure, the short-run gain from being bribed only by low quality producers is $v - f$, as compared to v for full disclosure. Therefore, partial disclosure accounts for lower discount factors. This presumes that the largest threat of capture indeed stems from low quality producers. This is generally true for full disclosure, but ceases to hold for partial disclosure. When the maximum threat stems from a bribe accepted by all producer types, the long-run loss is reduced. Obviously, this is perceived as fraud only when the product quality was low. As a result, per-period profits are lost with probability $1-e$ only. On the other hand, the bribe need to be lower in order to be accepted by high quality producers, which again reduces the short-run gain. The following proposition proves that the latter effect outweighs the former.

Proposition 3.6. *It holds that $\delta^{PD} < \delta^{FD}$. For any $\delta \in [\delta^{PD}, \delta^{FD}]$, a capture-proof equilibrium can only be sustained applying a noisy disclosure rule. Furthermore, for any $\delta \geq \delta^{FD}$, we have that $[e_l^{FD}(\delta), e_h^{FD}(\delta)] \subsetneq [e_l^{PD}(\delta), e_h^{PD}(\delta)]$.*

Proposition 3.6 shows our main result, namely that opacity can be used as a tool to improve certifier credibility. For any level of producer investment e, the range of discount factors that allow for capture-proof implementation of e is strictly larger for partial disclosure as compared to full disclosure. Similarly, for any discount factor δ, the set of investment levels that are implementable in a capture-proof equilibrium with partial disclosure is strictly larger than the corresponding set for full disclosure. Therefore, the superiority of partial disclosure takes place along two dimensions. Figure 3.6 displays these differences. The dark-grey area corresponds to (e, δ)-combinations that can be implemented as a capture-proof equilibrium under full disclosure. The light-grey area shows the *additional* (e, δ)-pairs that allow for implementation in capture-proof equilibrium under partial disclosure.

In Section 3.3, we show that a certifier would always want to implement e^{FD} as this maximizes her per-period profits. With full disclosure, this is only possible when $\delta \geq \delta^{FD}$. Partial disclosure allows for capture-proof equilibria also for lower discount factors. It is remarkable that, at least for a range of discount factors, this can be achieved without waiving any profits. To see this, denote $\tilde{\delta}(\pi^{FD})$ the smallest discount factor, such that a capture-proof equilibrium is sustained and achieves per-period profits of π^{FD}. The following corollary is an immediate consequence of Proposition 3.6.

Corollary 3.3. *It holds that*

$$\tilde{\delta}(\pi^{FD}) = \max\left\{\frac{v - \pi^{FD}}{v}, \frac{1}{1+e^{FD}}\right\} < \delta^{FD}.$$

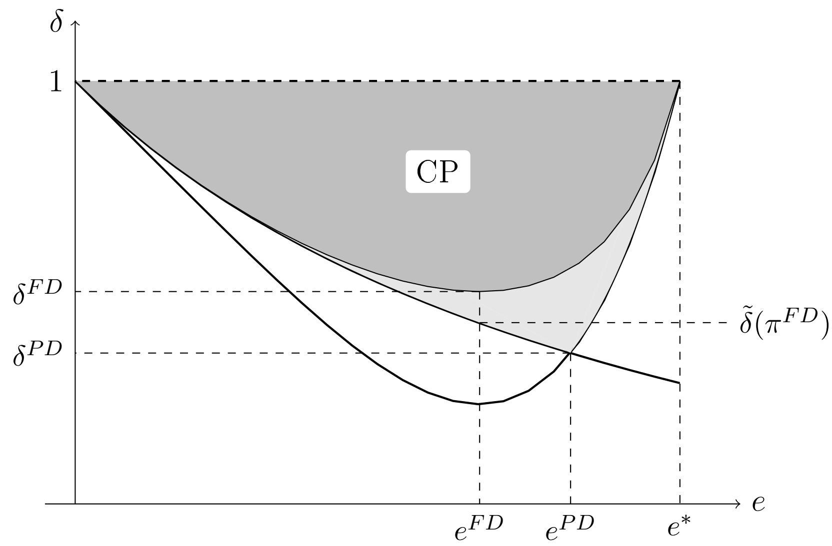

Figure 3.6: Dark-grey: capture-proof certification with full disclosure. Light-grey: (additional) capture-proof certification with noisy disclosure.

3.4.4 Welfare properties of partial disclosure

In this subsection, we study welfare properties of capture-proof equilibria with partial disclosure. When $\tilde{\delta}(\pi^{FD}) = \delta^{PD}$ we also have $\delta^{PD} = (v - \pi^{FD})/v$. In this case, the largest threat of capture stems from low quality producers, i.e. the largest deviation profit for the certifier is achieved for $b = v$. Then the certifier can still achieve the maximal per-period profits π^{FD} in a capture-proof equilibrium for any $\delta \geq \delta^{PD}$, which implies implementing $e = e^{FD}$.

This is however not true when $\tilde{\delta}(\pi^{FD}) > \delta^{PD}$. As can be seen from Figure 3.6, for discount factors below $\tilde{\delta}(\pi^{FD})$ the profit maximizing level of investment e^{FD} is no longer capture-proof implementable. Instead only larger values of producer investment can be implemented when $\delta \in [\delta^{PD}, \tilde{\delta}(\pi^{FD}))$. To provide an intuition for this, note the following: Bribing offers b that are accepted by all producer types pose the largest threat. Now, implementing a larger e leads to a reduction in V_2, as otherwise profits would increase beyond π^{FD}. To incentivize producers to make larger investments, the certifier therefore has to increase α. As now shown, for high quality producers the difference in expected profits between the lottery of the certification process and the sure certificate v is reduced.[19] This in turn lowers the maximum bribe they are willing to pay for capture and therefore reduces the short-run gain for the certifier from any such offer. From a

[19] Honest certification yields an expected payoff $\alpha v + (1-\alpha)V_2$. This value is reduced when α increases and V_2 decreases at the same time.

welfare perspective this increase in investment is beneficial. Social welfare is given by $e \cdot v - k(e)$ in each period. The first-best investment level e^* was shown to be strictly larger than e^{FD} and welfare is strictly increasing on $[0, e^*]$. Implementing certification with partial disclosure for discount factors $\delta \in [\delta^{PD}, \tilde{\delta}(\pi^{FD})]$ therefore increases social welfare as compared to doing so for larger levels of the discount factor. Put differently, a severe threat of capture increases welfare. We summarize this in the following proposition.

Proposition 3.7. *Assume $\tilde{\delta}(\pi^{FD}) > \delta^{PD}$. For intermediate levels of the discount factor, i.e. $\delta \in [\delta^{PD}, \tilde{\delta}(\pi^{FD}))$, only investment levels that are strictly larger than e^{FD} can be capture-proof implemented with partial disclosure. This leads to increased social welfare.*

3.5 Discussion

We have analysed the effects of reputational concerns on optimal disclosure rules from the point of view of a monopolistic certifier. Our main finding is that if capture is an issue, a certifier benefits from resorting to coarser certification in order to reduce the threat of capture. In particular, for medium discount factors, sustaining honest certification is impossible if information is fully disclosed whereas it is still possible if information disclosure is noisy.

Implications of our analysis are manifold. First of all we provide a novel explanation for the occurrence of imperfect testing. In many papers on e.g. rating agencies (examples include Mathis et al. (2009) and Bolton et al. (2012)) imperfect testing is exogenously given, whereas here it arises in equilibrium. An empirical implication is that for low discount factors we expect disclosure to be coarser. This is consistent with the casual observation that certification in markets with low volume, such as wine, technical inspections or eco-labels often involves only a few different certificates. On the other hand, the high volume rating market exhibits a rather wide variety of different but still coarse certificates per rating agency.

Our findings also have important policy implications. Politics tend to push certifiers to precisely reveal information. Our results suggest that doing so may have unforeseen consequences for the functioning of those markets. In particular, reputation building and the resistance to capture is impeded. Similarly, regarding the current financial crisis, forcing rating agencies to issue more precise information might even exacerbate capture problems.

We have demonstrated our results in a highly stylized model, but the intuition behind our results is general. In particular, they carry over to more than only two quality specifications. The appendix provides examples for honest equilibria for the case of more than two quality specifications.

Our restriction to a particular class of noisy disclosure rules is without loss of generality. First, offering various coarse certificates generates incentives for the certifier to always offer the best among the noisy certificates in a bribing offer. This will be accepted (at least by low quality producers) in order to avoid a lottery that includes the worst certificates. As deviations of this kind remain undetected they will occur with certainty, and this destroys the equilibrium. Second, disclosure rules that do not allow for unambiguous detection of deviations call for a different type of trigger beliefs. Consumers lose trust in the certifier whenever they first detect low quality sold with the best certificate. This leads to punishments even if collusion did not take place. The harsher punishments makes it impossible to sustain capture proof equilibria for low discount factors. Proposition 3.8 in the appendix makes this statement precise.

Finally, we have used a specific extensive form to model capture. More sophisticated forms to study imply non-uniform bribing offers, e.g. menus, to elicit the producers' private information. A possible extension is the analysis of posterior bribing, i.e. after the certifier learned q. Also, producers might be equipped with tools to to signal their private information. The exact extensive form may well affect parts of the analysis albeit not the main finding.

3.6 Appendix

3.6.1 Proofs

Proof of Lemma 3.1. Follows immediately from the arguments given in the text. □

Proof of Lemma 3.2. Following the arguments given in the text the certifier maximizes (3.2). Recall that we assume $k'''(\cdot) \geq 0$, which ensures that this profit function is concave in e, thus the first-order condition is sufficient for an optimum. This first-order condition is $0 = v - k'(e) - ek''(e)$. Define $\Psi(e) = v - k'(e) - ek''(e)$. We have $\Psi(0) = v > 0$ and $\Psi(1) = v - k'(1) - k''(1) \leq 0$ by our assumptions on $k(\cdot)$. Furthermore Ψ is strictly decreasing due to strict concavity of $k(\cdot)$. Hence there exists a unique e^{FD} such that $\Psi(e^{FD}) = 0$, which consequently is the unique maximizer of the certifier profit. The formulas for e^{FD} and f^{FD} follow easily from the formulas above. □

Proof of Proposition 3.1. First of all a disclosure rule can potentially lead to four different subgames: (1) no producer demands certification, (2) only low quality producers demand certification, (3) only high quality producers demand certification, and (4) all producers demand certification. Note that we do not explicitly consider mixed strategies by producers. The reason is that any outcome where some producers randomize their certification decision can be replicated by a disclosure rule that adds the respective probabilities for not certifying to the probabilities of remaining uncertified though paying for certification. To see this, assume type i chooses to certify with probability $\gamma \in (0,1)$. Now multiply every α^i by γ and increase the probability of remaining uncertified appropriately. After changing the fee from f to γf, it is easy to see that this adjusted disclosure with the reduced fee leads to the same investment incentives and also to the same equilibrium prices for (un-)certified products and the certifiers profit is unchanged.
Case (1) trivially leads to zero profits and the claim is proven.
Case (2) leads to consumers paying zero in equilibrium for certified products.[20] To make low quality producers "pay" for certification we consequently must have $f = 0$ which leads to zero profits and proves our claim also in this case.
Case (3) can be analysed as follows: If only high types certify, rational behaviour by consumers dictates that a certified product is sold at a price v. Uncertified products however can be of either high or low quality and have some price $q^{\emptyset} \in [0, v)$.

[20] A disclosure leading to this particular subgame is given by $\mathcal{C} = \{C\}, \alpha^l = 1$ and $\alpha^h = 0$.

A producer's investment decision is given by the solution of

$$\max_e \ e\Big(\sum_k \alpha_k^1 v + (1 - \sum_k \alpha_k^1)q^\emptyset - f\Big) + (1-e)q^\emptyset - k(e),$$

which yields the following first-order condition for producer investment:

$$\Big(\sum_k \alpha_k^1(v - q^\emptyset) - f\Big) = k'(e).$$

Rewriting this constraint in terms of induced investment yields $f = v - k'(e) - (1 - \sum_k \alpha_k^1)(v - q^\emptyset) - q^\emptyset$. Now we have for the certifier profit

$$\pi^D(f) = e(f, D) \cdot f = e \cdot \big(v - k'(e) - (1 - \sum_k \alpha_k^1)(v - q^\emptyset) - q^\emptyset\big) \leq e \cdot (v - k'(e)) \leq \pi^{FD}.$$

This proves the claim for case (3).

Finally consider case (4): When both producer types demand certification, the resulting certifier profit in the subgame is $\pi^D(f) = f$. The price at which a product holding certificate C^i can be sold is

$$q^{C^i} = v \cdot \frac{e\alpha_i^h}{e\alpha_i^h + (1-e)\alpha_i^l}.$$

Uncertified products are sold at price $q^\emptyset = v \cdot \frac{e\left(1 - \sum_i \alpha_i^h\right)}{e\left(1 - \sum_i \alpha_i^h\right) + (1-e)\left(1 - \sum_i \alpha_i^l\right)}$. A producer's investment decision follows from maximizing his expected payoff from certification, given by

$$e \cdot \left(\sum_i \alpha_i^h q^{C^i} + \left(1 - \sum_i \alpha_i^h\right) q^\emptyset\right) + (1-e) \cdot \left(\sum_i \alpha_i^l q^{C^i} + \left(1 - \sum_i \alpha_i^l\right) q^\emptyset\right) - f - k(e).$$

The resulting investment constraint is

$$k'(e) = \sum_i (\alpha_i^h - \alpha_i^l)(q^{C^i} - q^\emptyset). \tag{3.12}$$

On the other hand, from the formula given for q^{C^i} we have $e\alpha_i^h q^{C^i} + (1-e)\alpha_i^l q^{C^i} = ev\alpha_i^h$. Similarly $e(1 - \sum_i \alpha_i^h)q^\emptyset + (1-e)(1 - \sum_i \alpha_i^l)q^\emptyset = ev(1 - \sum_i \alpha_i^h)$. Summing those expressions yields

$$\sum_i \left(e\alpha_i^h q^{C^i} + (1-e)\alpha_i^l q^{C^i} \right) + e(1 - \sum_i \alpha_i^h) q^{\emptyset} + (1-e)(1 - \sum_i \alpha_i^l) q^{\emptyset} = ev. \tag{3.13}$$

Rewriting the left hand side of equation (3.13) yields

$$e \sum_i (\alpha_i^h - \alpha_i^l)(q^{C^i} - q^{\emptyset}) + \sum_i \alpha_i^l q^{C^i} + \left(1 - \sum_i \alpha_i^l \right) q^{\emptyset} = ev. \tag{3.14}$$

Finally, to make all producer types demand certification we must have in particular

$$f \leq \sum_i \alpha_i^l q^{C^i} + \left(1 - \sum_i \alpha_i^l \right) q^{\emptyset} \tag{3.15}$$

i.e. low quality producers ecpected payoff from certification must be non-negative.[21] From this we can derive an upper bound on certifier profits:

$$\begin{aligned} \pi^D(f) = f &\stackrel{(3.15)}{\leq} \sum_i \alpha_i^l q^{C^i} + \left(1 - \sum_i \alpha_i^l \right) q^{\emptyset} \\ &\stackrel{(3.14)}{=} ev - e \sum_i (\alpha_i^h - \alpha_i^l)(q^{C^i} - q^{\emptyset}) \\ &\stackrel{(3.12)}{=} ev - ek'(e) = e(v - k'(e)). \end{aligned}$$

But $e\,(v - k'(e))$ is the profit from implementing effort level e optimally with a full disclosure rule, therefore we have proven $\pi^D(f) \leq \pi^{FD}$. □

Proof of Proposition 3.2. In any equilibrium in which Assumption 3.1 holds capture may not take place, since otherwise the beliefs of consumers are not consistent with the behaviour of the certifier. Hence, condition (3.5) must be satisfied for all b. As mentioned in the text, certifier profits from deviating $\widehat{\Pi}^{FD}(b|f)$ are largest for b approaching v. Taking this limit yields

$$\begin{aligned} \lim_{b \nearrow v} \widehat{\Pi}^{FD}(b|f) &= (1 - e(f))v + e(f) \cdot \left(f + \delta \Pi^{FD}(f) \right) \\ &= (1 - e(f))v + \pi^{FD}(f) + \frac{\delta}{1-\delta} e(f) \pi^{FD}(f) \\ &= (1 - e(f))v - \frac{\delta}{1-\delta}(1 - e(f))\pi^{FD}(f) + \Pi^{FD}(f). \end{aligned}$$

[21] More conditions are required in subgame where all producer types demand certification, but the one presented her is the only required for our proof.

Condition (3.5) is thus equivalent to

$$(1-e(f))v \le \frac{\delta}{1-\delta}(1-e(f))\pi^{FD}(f).$$

Rearranging this expression yields that condition (3.5) is satisfied if and only if

$$\delta \ge \delta^{FD}(f) \equiv \frac{v}{v+\pi^{FD}(f)}.$$

□

Proof of Proposition 3.3. We first argue how condition (3.6) can be translated towards (3.8). Recall $\pi^{FD}(f) = e(f)\cdot f$ and optimal investment by producers requires $k'(e) = v - f$. Replacing f by $v - k'(e)$ yields (3.8). All other statements are straightforward reformulations of Proposition 3.2 and Corollary 3.1. □

Proof of Proposition 3.4. In any equilibrium in which Assumption 3.2 holds capture may not take place, since otherwise the beliefs of consumers are not consistent with the behaviour of the certifier. Hence, condition (3.5) must be satisfied for all b. We compute the respective critical discount factors. Taking the limit of $\widehat{\Pi}^D(b|f)$ as b approaches $f+(1-\alpha)(v-V_2)$ we get

$$\begin{aligned}
\lim_{b \nearrow f+(1-\alpha)(v-V_2)} \widehat{\Pi}^D(b|f) &= f+(1-\alpha)(v-V_2)+e(\alpha)\delta\Pi^{PD}(f) \\
&= f+(1-\alpha)(v-V_2)+e(\alpha)\frac{\delta}{1-\delta}f \\
&= (1-\alpha)(v-V_2)-\frac{\delta}{1-\delta}(1-e(\alpha))f+\Pi^{PD}(f).
\end{aligned}$$

Consequently this limit lies below $\Pi^{PD}(f)$ if and only if

$$(1-\alpha)(v-V_2) \le \frac{\delta}{1-\delta}(1-e(\alpha))f,$$

respectively whenever

$$\delta \ge \delta^{l,h}(\alpha,f) = \frac{(1-\alpha)(v-V_2)}{(1-\alpha)(v-V_2)+(1-e(\alpha))f}.$$

Similarly the limit of $\widehat{\Pi}^D(b|f)$ as b approaches $f+(v-V_2)$ can be rewritten as follows

$$\begin{aligned}\lim_{b\nearrow f+(v-V_2)} \widehat{\Pi}^D(b|f) &= (1-e(\alpha))\cdot(f+(v-V_2))+e(\alpha)\big(f+\delta\Pi^{PD}(f)\big)\\ &= (1-e(\alpha))(v-V_2)-\frac{\delta}{1-\delta}(1-e(\alpha))f+\Pi^{PD}(f).\end{aligned}$$

Therefore $\lim_{b\nearrow f+(v-V_2)} \widehat{\Pi}^D(b|f)\leq \Pi^{PD}(f)$ if and only if

$$\delta \geq \delta^l(\alpha,f)=\frac{v-V_2}{f+v-V_2}.$$

Because capture-proofness requires $\widehat{\Pi}^D(b|f)\leq\Pi^{PD}(f)$ for all b, (3.9) follows. □

Proof of Corollary 3.2. As discussed in the text, the certifier may set $f = V_2$ to minimize the threat of capture. We consider $\delta^l(\alpha,f)$ first. Making use of $f=V_2$ allows us to simplify it to $(v-V_2)/v$. From the proof of Proposition 3.1 we get $V_2=e(v-k'(e))$ and therefore

$$\delta^l(\alpha)=\frac{v-V_2}{v}=\frac{v-e(\alpha)(v-k'(e(\alpha)))}{v}.$$

Now consider $\delta^{l,h}(\alpha,f)$. With $f=V_2$ we may rewrite

$$\delta^{l,h}(\alpha,f)=\frac{(1-\alpha)(v-V_2)}{(1-\alpha)(v-V_2)+(1-e(\alpha))V_2}$$

By Bayesian updating we have $V_2=v\cdot\big((1-\alpha)e(\alpha)\big)/\big(1-\alpha e(\alpha)\big)$ in equilibrium, which implies $v-V_2=v\cdot\big(1-e(\alpha)\big)(\big(1-\alpha e(\alpha)\big)$. Replacing V_2 and $v-V_2$ accordingly yields

$$\frac{(1-\alpha)(v-V_2)}{(1-\alpha)(v-V_2)+(1-e(\alpha))V_2}=\frac{1}{1+e(\alpha)}.$$

□

Proof of Proposition 3.6. Recall, that with full disclosure the critical discount factor is $\delta^{FD}(e)=\frac{v}{v+\pi^{FD}(e)}=\frac{v}{v+e(v-k'(e))}$ and this term is minimized for the profit maximizing effort e, yielding $\min_e\delta^{FD}(e)=\frac{v}{v+\pi^{FD}}$. For all $e\in(0,e^*)$ we have $\frac{v-e(v-k'(e))}{v}<\delta^{FD}(e)$. To see this:

$$\frac{v-e(v-k'(e))}{v}<\delta^{FD}(e)=\frac{v}{v+e(v-k'(e))}\quad\Leftrightarrow\quad(e(v-k'(e)))^2>0.$$

Also

$$\frac{1}{1+e}<\delta^{FD}(e)=\frac{v}{v+e(v-k'(e))}\quad\Leftrightarrow\quad ek'(e)>0$$

Therefore also $\max\{\frac{1}{1+e}, \frac{v-e(v-k'(e))}{v}\} < \delta^{FD}(e)$ for all $e \in (0, e^*)$ and hence we can define

$$\delta^{PD} := \min_e \max\left\{\frac{1}{1+e}, \frac{v - e(v - k'(e))}{v}\right\}$$

and it follows that $\delta^{FD} > \delta^{PD}$. Since both $\delta^{PD,l}(e) < \delta^{FD}(e)$ and $\delta^{PD,l,h}(e) < \delta^{FD}(e)$ the last statement follows immediately. $\square$

Proof of Proposition 3.7. When $\delta^{PD} < \tilde{\delta}(\pi^{FD})$ we must have $\tilde{\delta}(\pi^{FD}) = \delta^{PD}(e^{FD}) = \frac{1}{1+e^{FD}}$. Since $1/1+e$ decreases in e we have $\delta^{PD}(e) > \tilde{\delta}(\pi^{FD})$ for any $e < e^{FD}$. Consequently we must have $\delta^{PD}(e) < \tilde{\delta}(\pi^{FD})$ on some interval $[e^{FD}, \hat{e}]$. This proves our result. $\square$

3.6.2 Extensions

Examples for more than two levels of quality

Let quality levels be $\{0, 0.5, 1\}$ and $P(q = 0.5|e) = P(q = 1|e) = e/2$. Consequently $P(q = 0|e) = 1 - e$. The cost of effort is $k(e) = e^2/2$. If we restrict the analysis to deterministic disclosure rules, it is straightforward to show that full disclosure with a fee $f = 3/8$ maximizes certifier profits. With this fee both quality levels 0.5 and 1 get certified in equilibrium. Using the same line of argument as in the main text, this disclosure rule can be sustained as a capture-proof equilibrium whenever $\delta \geq \frac{16}{19}$.

A cut-off disclosure rule that certifies any product whose quality exceeds 0, but does not distinguish any further, achieves the same static profit as the mentioned full disclosure rule. However, the largest possible bribe is then not equal to 1 since no certificate which yields a price of one is available. Instead, the best certificate yields 3/4, the value of a certified product. Consequently, a capture-proof equilibrium with this disclosure rule exists whenever $\delta \geq \frac{16}{20}$. While profits remain the same, the largest acceptable bribe is lowered.

Alternative disclosure rules for the two-quality case

Proposition 3.8. *For any $\delta < \delta^{FD}$ and any disclosure rule which is such that the highest certificate's value is different from v, no capture-proof equilibrium exists.*

Proof. We restrict the proof to the following simple disclosure rule[22]: there are two

[22] For any other rule, the argument is the same for selling the best certificate in a capture offer to the low quality producer. However, there are even more feasible bribing offers, which make it even harder to resist the threat of capture.

certificates, C_1 and C_2, where high quality always receives C_1 and low quality receives C_1 with probability $\alpha \in (0,1)$. Denote V the value of C_1, certificate C_2 is always worth zero (in equilibrium). The first-order condition for producer investment reads as

$$k'(e) = (1-\alpha)V$$

and from Bayes' rule we have

$$V = v\frac{e}{e+\alpha(1-e)}.$$

Thus, to implement a particular e, the certifier has to set[23]

$$\alpha = \frac{e(v-k'(e))}{e(v-k'(e))+k'(e)}$$

The fee must be such that low quality producers are willing to get their product certified, i.e. $f \leq \alpha V$.

When a purchased product with certificate C_1 turns out to be of low quality, consumers cannot be sure whether this was due to bad luck or to a captured certifier. Appropriate trigger beliefs have to be such that the certifier is punished whenever low quality is sold with certificate C_1. This can well happen without any deviation by the certifier. The probability of entering punishment, absent any deviation, is $p = (1-e)\alpha$ and expected profits from honest play are given by

$$\Pi^h(\alpha, f) = f + (1-p)\delta f + (1-p)^2\delta^2 f + \ldots = \frac{f}{1-(1-p)\delta}.$$

The maximal bribe is given by $b \approx (1-\alpha)V + f$, where only low quality producers accept it. The profit from making such an offer is

$$\Pi(b|f,\alpha) = (1-e)b + e(f + \delta\Pi^h(\alpha, f))$$

We have $\Pi(b|f,\alpha) \leq \Pi^h(\alpha, f)$ for $b \to (1-\alpha)V + f$ whenever

$$\delta \geq \frac{(1-e)b-(1-e)f}{(1-e)(1-p)b-epf} = \frac{b-f}{\big(1-(1-e)\alpha\big)b-e\alpha f}$$

This is both increasing in b and in f, such that the largest threat is exercised for $f = \alpha V$

[23]Note that $\lim_{e\to 0}\alpha$ equals 1 whenever $k''(0)=0$ and otherwise equals $\frac{v}{v+k''(0)\in(0,1)}$, that is in the latter case not all α are implementable.

and $b = V$, which results in the condition

$$\delta \geq \frac{1}{1 + e\alpha}.$$

We have $\frac{1}{1+e\alpha} \geq \frac{v}{v+e(v-k'(e))}$ if and only if

$$v - k'(e) \geq v\alpha \quad \Leftrightarrow \quad 1 \geq e.$$

Hence, for all e to be implemented, this is only possible with a noisy rule without sure high quality certificate, when this is also possible using a full disclosure rule. $\square$

4 Non-persuasiveness mitigates competition in the market for life insurances

Chapter Abstract

In this article, we analyse a duopolistic insurance market with bargainers and price takers. Firms compete in two stages. We set up a model in which firms, but not applicants themselves, have the option to screen an individual's risk of incurring a loss. In equilibrium, insurers are asymmetrically informed on individual applicants. When an informed insurer is able to signal high risk, she is said to be *persuasive*. If not, competition is invalidated on high risk bargaining applicants, resulting in strictly positive expected profits for the uninformed firm. It is shown that all equilibria are non-persuasive, list prices may be asymmetric and both firms expect to earn strictly positive profits.

4.1 Introduction

In insurance markets that provide economic coverage for a biometric risk (e.g. life, health or disability insurances), contract offers are based on the outcome of a screening, which is conducted by an insurance firm. More specifically, if a person wishes to buy insurance, he is required to provide inter alia detailed information on his medical record. Based on the screening result, the person is either accepted and offered a pre-specified contract, or rejected.

The medical record is not the only basis for the decision: consumers are furthermore required to provide information on their application history. In particular, insurers ask whether a consumer has applied for this kind of insurance before and if so, what the outcome was. The information provided is twofold: first, if the answer is 'yes', the respective insurer will be aware of being faced with a person in search for the best offer, a so called bargainer. Second, information about previously received offers may help the firm to engage in competition without having to incur costs for screening.[1]

Another common method to collect the same kind of information is by means of an information exchange system. In these systems acceptance/rejection decisions are made public to competitors. Examples for these systems are the German Hinweis- and Informationssystem (H.I.S.) and the U.S. Medical Information Bureau (MIB).

In the present study, we analyse this specific market with two competing insurers. In particular, we are interested in the role of bargainers versus price takers (persons that apply only once) in determining industry profits. Surprisingly, we find that the presence of bargainers helps both firms to positive expected profits. By contrast, as the share of price takers converges to one, competition works effectively and firms do not make positive profits.

These results also help to explain the occurrence of information sharing systems: with bargainers being present, both firms earn positive profits. By contrast, if information about acceptance/rejection decisions is not shared, both firms opt to screen applicants, producing a perfectly competitive market environment. This outcome is counterintuitive only on first sight – because an insurer shares information also about the bargaining behaviour of applicants, the competitor understands he will not earn positive profits, and consequently forgo costly screening.[2]

It is not obvious why rejections happen at all. With informed consumers, higher risk

[1]Note that there are no incentives for misreporting, as far as the application history and the medical record are concerned. If misreporting is detected by the time a claim is submitted, or earlier, an insurer may rescind the contract without having to pay back any paid premiums.

[2]This result is immediate and its analysis considered in the appendix.

types would not remain uninsured, since they would be willing to pay more for their insurance. On the other hand, firms dispose of advanced technological means to estimate the statistical risk of a loss. Although a consumer is aware of his medical condition, he usually does not have the necessary experience to evaluate his own risk. A relatively young stream of literature considers the possibility that insurers are better informed than consumers. In particular, Villeneuve (2005) studies a model in which applicants can infer their default probabilities only from the contract offers they receive. This signalling structure provides an explanation for the actual occurrence of rejections: since an applicant's willingness to pay for insurance directly depends on the beliefs he holds about his risk type, an informed firm may not be able to persuade consumers that their risk is high by setting high prices. In that case, the only way for a firm to convincingly signal high risk may be to reject the applicant.

In this study, we analyse a model in which firms, but not applicants themselves, have the option to screen an individual's risk of incurring a loss. More precisely, the two-stage competition game proceeds as follows. Firms compete in list prices. Persons who wish to purchase insurance apply to the firm that offers the best list price. There, the person's medical profile is screened if the firm spends a small amount (this decision is not observable). Depending on the screening outcome (or the belief about the personal risk of the applicant), an applicant is considered eligible for insurance at the list price, or not, in which case he is rejected. Price takers consider only this offer, while bargainers induce, in a next stage, competition between both insurers. Before entering the competition for bargainers, the competitor chooses whether to screen the applicant, or not, and observes the applicant's previously received offer.

In equilibrium, the firm to which applicants apply first (the winner of the list price competition), screens his risk profile, while her competitor does not. This implies that the acceptance/rejection decision transmits information to both the competitor and the applicant.

One effect, which we shall call *persuasiveness* of the informed firm, is central to our analysis: consider the firm that has won the list price competition and assume she faces a low risk applicant. By accepting this applicant, she reveals the true risk type. This earns him the list price minus potential compensations in case of damage, but only if the applicant is a price taker. If he is a bargainer, the true risk is revealed and competition works effectively, leaving the insurer without positive profits.

He could also reject the applicant. If he is a price taker, the person remains uninsured and the firm earns nothing. But in case he is a bargainer, both the competitor and the applicant believe that the person's risk is high. In the second round, the competitor offers the fair price for the high risk type. That implies that the informed firm can benefit, and earns strictly positive profits.

We say that the insurer is *persuasive* whenever she weakly prefers to accept the low risk applicant to rejecting him. If not, then in equilibrium, a high risk bargaining applicant is rejected twice by the informed firm[3] – for the benefit of her competitor. The insurer would like to undercut her competitor's offer, but she can't: doing so induces the applicant to believe to have a low default risk. Note that the term persuasiveness refers only to bargaining applicants. High risk price takers by contrast are always rejected, in the sense of Villeneuve (2005).

Our first main finding is that signalling is never persuasive in equilibrium. The reason is that, in order to give persuasion, list prices would need to be relatively high. In particular, the lowest list price would have to exceed some critical threshold. When persuasion is given, the standard Bertrand logic applies and competition works. Once competition drives price offers into an interval in which persuasion is not given, being the ignorant firm becomes relatively more attractive however.

This impacts list prices: on account of the incentives of not winning the price competition and thereby incurring the ignorant firm's position list price competition remains imperfect. Under list price competition, the 'natural' symmetric equilibrium in which both firms earn the same profits may not exist, either as a result of the non-persuasion requirement or of the maximum willingness to pay for a low risk individual. In this case – our second important finding – different list prices are offered in equilibrium: one firm will offer a high price in the persuasion interval. Thereby, the winner of the list price competition, although earning less, has no incentives to switch roles. The observation that list prices differ between firms is in line with the real market situation.

Another finding is that overall industry profits increase in the share of bargainers in the market. If this share is low, on the contrary, list prices converge to the competitive solution: since the uninformed firm does not, in the absence of bargainers, expect to earn positive profits, winning the list price competition becomes relatively attractive again. A high share of bargainers creates the opposite effect: the future ignorant firm expects high profits.

Related literature A small stream of literature analyses the effects of consumer bargaining behaviour on prices and firm profits. In Gill and Thanassoulis (2013), firms first compete for price takers in list prices and only in a second step for bargainers. They find that a higher share of bargainers dampens competition both in the price listing stage and in the

[3]More precisely, the informed firm's offer is never accepted.

bargaining stage. Their result is in line with our findings, although the effect is somewhat different: there, list prices serve as an outside option for second stage prices and thereby reduce incentives to undercut the rival. In Gill and Thanassoulis (2009), the second stage effect is also present, while the unique market list price is determined through Cournot competition. In the present study, by contrast, list prices rise due to the inability of an informed firm to convincingly signal high risk. Desai and Purohit (2004) analyse a model in which the bargaining strategy is an endogenous decision variable of firms. Raskovich (2007) demonstrates that list prices can jump from the fully competitive outcome to the monopoly price if the share of bargainers exceeds some critical threshold. The logic is that in this juncture firms find it more worthwhile to negotiate prices with bargainers individually.

In our paper, meaningful signalling is possible since the beliefs restrict the maximum willingness to pay for insurance. Beliefs may be such that an insurer and a high risk applicant cannot agree on a price, forcing firms to reject these individuals. As a result, equilibria with fixed and commonly known sender preferences need not necessarily be pooling but can be separating. Put differently, the only tool to signal high risk may be to reject an applicant.

To the best of our knowledge, we are the first to point out how customer bargaining behaviour affects such persuasive signalling. We do not allow for firm tools to generate persuasiveness, like reputation building (e.g. Sobel 1985, Gentzkow and Shapiro 2006), advertising (e.g. Milgrom and Roberts 1986) or multiple senders (e.g. Battaglini 2002, Villeneuve 2005). As to what the latter is concerned, with endogenous information acquisition firms will be asymmetrically informed in equilibrium.

A large stream of literature analyses quality signalling through prices (e.g. Bagwell and Riordan 1991, Ellingsen 1997, Adriani and Deidda 2011, Janssen and Roy 2010). There, high quality sellers manage to signal their type by distorting prices upwards and thereby reducing sold quantities. Adriani and Deidda (2011)) show that this might not be possible, i.e. high quality sellers drop out of the market if competition is strong, while Janssen and Roy (2010) demonstrate that equilibria exist even in a setting without additional product differentiation. This literature differs from the present study mainly in that we consider a market with inelastic demand. There, in order for separating equilibria to arise, additional conditions are required, such as buyers differing in their willingness to pay (Ellingsen 1997).

Our model adds to the literature of asymmetrically informed firms. As far as we know, a structure where one firm has private information over the customer and its competitor has only been considered in the credit lending literature, namely by Inderst and Mueller

(2006) and Inderst (2008). Asymmetrically informed firms with informed customers have been studied by de Garidel-Thoron (2005) in an insurance setting and by Sharpe (1990) and Hauswald and Marquez (2006) in the bank lending market. Asymmetrically informed firms are also considered in the literature on information sharing in competitive markets (see for instance Vives 1984, Gal-Or 1985, Spulber 1995, Kühn and Vives 1995 and Vives 2008). These works trade off the ex-ante effects of decreasing uncertainty against those of a changing competitive environment. There, the symmetrically imperfectly informed firms agree ex-ante on a certain way of sharing information and are subsequently bound to report honestly.

This study is organized as follows. Section 4.2 introduces the model. In Section 4.3, a modified version of the game is considered in order to grasp the role of persuasiveness. The game is then analysed in Section 4.4. Section 4.5 concludes. The effect of a prohibition of information sharing systems is discussed in the appendix. Then, both firms opt to screen applicants, resulting in fully revealing and perfectly competitive offers. All proofs are relegated to the appendix.

4.2 The model

Consumers with wealth level W incur the risk of running a loss d, which is normalized to 1. Their utility is given by a von Neumann-Morgenstern-function $u(\cdot)$ where $u'(\cdot) > 0, u''(\cdot) < 0$. A loss occurs with some probability $\theta \in \{L, H\}$ where $0 < L < H < 1$. We say θ is a consumer's risk type. Let q_θ with $\sum_\theta q_\theta = 1$ be the commonly held prior on the distribution of risk types in the population. At the beginning of the game, θ is unobservable for all players. A share λ of consumers does not bargain. In the following, we refer to them as price takers. Correspondingly, bargaining consumers are referred to as bargainers. Furthermore, we consider two competing risk-neutral insurance companies $i \in \{A, B\}$.[4] The firms can perfectly observe an individual's risk type (although not his bargaining behaviour) after having invested some arbitrarily small but strictly positive amount γ. We assume that this decision is not observable by other players. If a firm does not acquire information, prior beliefs are preserved unless the firm receives some meaningful signal from an informed firm. Consumers remain uninformed.

Offers. If a person applies for insurance, insurers make offers. In particular, offers may be considered acceptable or non-acceptable by consumers, contingent on their beliefs about

[4]In this setting we analyse, the results are supported only in a duopolistic structure. We discuss this in the last section.

their own risk type. If informed firms are not able to signal high risk because they have no tool to do so, these persons must be rejected, as shown in Villeneuve (2005).

Let Ω be the set of possible price offers. Formally, a rejection is identical to a price that is never acceptable to any consumer, whatever his beliefs. To make rejections more explicit, we denote by Δ the set of all offers in Ω that are not acceptable to a consumer who believes that he is an H-type. A consumer will accept an offer if his expected utility from it is at least as high as his expected utility from not buying insurance (his reservation utility).

Beliefs in every period can be expressed as the expected risk which is a mapping of the set of vectors of offers in t, P^t into the interval $[L, H]$. It is denoted as $\hat{\theta}^t = \hat{\theta}(P^t) \in [L, H]$. Ex-ante, it equals prior risk expectations, i.e. $\hat{\theta}^0 = q_L L + q_H H$. Then, a person's expected utility $U(\cdot)$ from accepting an offer p is given by

$$U(p) = u(W - p),$$

while his expected utility from not buying any insurance is

$$U^0(\hat{\theta}^t) = (1 - \hat{\theta}^t)u(W) + \hat{\theta}^t u(W - 1).$$

A consumer judges an offer p acceptable if and only if $U(p) \geq U^0(\hat{\theta}^t)$. In the following, the maximum willingness to pay for given beliefs will play an important role. In particular, three prices will be of relevance: $\tilde{p}_L$ and $\tilde{p}_H$ are the prices that correspond to the certainty equivalents for L and H-types, i.e. $U(\tilde{p}_\theta) = U^0(\theta)$. A consumer with the prior risk expectation θ^0 is indifferent between insurance at price $\tilde{p}_u$ and no insurance. To circumvent existence problems (see below), we assume the following.

Assumption 4.1. $\tilde{p}_u < H$.

From the definition of Δ follows that all prices in Δ are strictly greater than $\tilde{p}_H$. A firm's posterior profit is $\pi_i(p) = p - \theta$ if her offer p is selected by a consumer and zero otherwise. The lowest price a firm asks if she is informed about θ is the fair price, i.e. θ.

The game proceeds as follows:

t=0 Firms set list prices p_A and p_B. Consumers apply to the firm that offers the lowest list price if this price does not exceed $\tilde{p}_L$. If it does, the game ends. Applicants randomize with probability 1/2 if $p_A = p_B \leq \tilde{p}_L$.

t=1 Price taking stage: the firm where the consumer applies, denoted firm i, makes an investment choice $\gamma_i \in \{0, \gamma\}$ and contingently observes θ. Contingent on this observation, the firm publicly and individually accepts the applicants and offers the list price p_i or rejects him.

t=2 Bargaining stage: bargainers ask both firms to make better offers: firms choose whether to acquire information on consumers that have not been assigned to them. Firms simultaneously offer $p_{i,2}, p_{j,2}$ for $i, j \in \{A, B\}$ and $j \neq i$. Bargainers choose among all offers, i.e. $p_i, p_{i,2}, p_{j,2}$ and no insurance.

Therefore, firms compete for price taking applicants. More specifically, competition takes place on low risk consumers only, while high risks are being rejected. By modelling list price competition accordingly, we follow Inderst (2008) and Inderst and Mueller (2006). In the following section, it is demonstrated in a modified version of this game why firms have no tool to signal high risk. The basic idea is that if an applicant was willing to believe that he has a high risk when offered a high price, firms would always want to induce this belief to make him pay higher prices. By Assumption 4.1, it must also hold that $\tilde{p}_L < H$, and all equilibria must be such that high risk applicants end up rejected.

It is assumed that all applicants first apply to the firm that offers the best list price. That is, we abstract from the possibility that bargaining consumers may wish to employ a search strategy that differs from that of price taking consumers.

In this game, the decision whether to acquire information about risk types is made on an individual basis (every applicant is an 'individual market'). That is to say, a firm may have an informational advantage on some consumers while not on others vis-à-vis her competitor. Also note that the insurer who does not win the list price competition in $t = 0$ has observed the other firm's acceptance/rejection decision. This game specification is motivated by the fact that insurers indeed share information about offers via an information sharing system.[5] As a result, in equilibria in which the insurer who has not won the price competition in $t = 0$ chooses to remain uninformed while her competitor does not, the price offer p_i serves both uninformed players – this insurer and the applicant – to update beliefs, see further below. A further note refers to consumer preferences: for technical convenience, we assume that whenever a consumer is indifferent between offers in $t = 2$, he chooses the offer from the firm to which he has applied in the first place.

Equilibrium selection. Consumers hold beliefs $\hat{\theta}^0$ about their own risk type and the firms' informational state. Updating takes place after the price taking stage, i.e. after the first

[5] The effects of a prohibition are being discussed in the appendix.

offer has been made, and after the bargaining stage. Since bargainers do not have to make a decision after the price taking stage, we are not concerned with their beliefs after the price taking stage. We assume that whenever consumers observe off-path behaviour, they make inferences only about their risk type. Put differently, beliefs on investment behaviour remain fixed and are never updated. We require that off-path beliefs satisfy the Intuitive Criterion by Cho and Kreps (1987).

The equilibrium concept is the Perfect Bayesian Nash Equilibrium, meaning that beliefs have to be consistent in equilibrium. We restrict our analysis to pure strategy equilibria and full insurance contracts.[6]

Comments on game structure. Bertrand games may not have equilibria if utility functions exhibit discontinuities (see e.g. Dasgupta and Maskin 1986). This problem arises here because consumers update their beliefs based on the obtained contract offers which may render offers in some intervals unacceptable.[7] Equilibria in the bargaining game are affected whenever firms are asymmetrically informed and the market situation is such that a consumer who believes to be an L type is not willing to pay at least H while, if priors are preserved, the willingness to pay is at least H. Since we are more concerned with the equilibria in which an informed firm reveals information in the price taking stage, we circumvent this problem by considering only situations where the willingness to pay of a consumer who holds prior beliefs is less than H (Assumption 4.1).

4.3 Preliminary analysis: persuasiveness

In this section, in order to grasp the role of persuasiveness here, we analyse the subgame starting in $t = 1$. This is to say, we abstract from list price competition for a moment and assume that in $t = 0$, applicants are assigned randomly (with probability 1/2) to one of the two firms. The firm to which a consumer has been assigned makes him an offer, and bargainers then proceed to the second stage of the game. More precisely, the game we are considering here is the following:

t=0 With probability 1/2, nature assigns consumers to firm i for $i \in \{A, B\}$.

[6] This may require further assumptions on the market situation. We discuss this topic in the last section.

[7] In particular, consider the following situation: let firm A be informed about θ, while B only knows that $\theta \in \{L, H\}$. If the consumer, whose belief is L, is not willing to accept at least the fair price for H, i.e. H, there is an interval of prices that would never be accepted by the consumer if the offer is made by A.

t=1 Price taking stage: the firm to which the applicant has been assigned, denoted firm i, makes an investment choice $\gamma_i \in \{0, \gamma\}$ and contingently observes θ. Contingent on this observation, the firm publicly announces an individual price offer p_i^m. Price takers accept the offer or choose no insurance.

t=2 Bargaining stage: bargainers ask both firms to make better offers: firms choose whether to acquire information on consumers that have not been assigned to them. Firms simultaneously offer $p_{i,2}^m, p_{j,2}^m$ for $i, j \in \{A, B\}$ and $j \neq i$ (and where i denotes the firm to which an applicant has been assigned in $t = 0$). Bargainers choose among all offers, i.e. $p_i^m, p_{i,2}^m, p_{j,2}^m$ and no insurance.

Note also that in this game formulation, we allow period-one firm i to freely choose offers from the contract space Ω on each individual applicant in $t = 1$, instead of being restricted to the formerly set list price. Here, we demonstrate that no distinct price offers are made in $t = 1$ to applicants of different risk types. Put differently, the restriction to a single price in the general formulation of the game is without loss of generality.

This section is organized as follows: it is first shown that only certain informational states may arise in equilibrium. In a next step, we analyse properties of such potential equilibria before investigating which of these equilibria exist.

In particular, consider some equilibrium in which each firm chooses to gather information on each single applicant, that is, no matter whether being his period-one firm or not.

Let $p_{i,t}^{m,\theta}$ for $\theta \in \{L, H\}$ be denoted the equilibrium offers made by an informed firm i in t. From Villeneuve (2005), Proposition 6, we know that whenever both firms are informed, in any equilibrium it holds that $(p_{A,2}^{m,L}, p_{B,2}^{m,L}) \neq (p_{A,2}^{m,H}, p_{B,2}^{m,H})$ and that $p_{A,2}^{m,L} = p_{B,2}^{m,L} = L$.[8] In words, any equilibrium in the bargaining game is separating and the competitive outcome arises on the L-type. But because profits are zero on the L-type, the second firm may as well not acquire information and always offers the equilibrium contract for the H-type in order to save costs. As a result, these equilibria never exist.

On the other hand, consider an equilibrium in which no firm gathers information on an applicant's risk type, that is, whether being his period-one firm or not. But in this case, firms will always benefit from clandestinely acquiring information. We can therefore focus on equilibria in which either the period-one firm acquires information while her competitor does not, or the inverse. Consider an equilibrium in which j is informed while the period-one insurer i is not. By Assumption 4.1, it holds that if i is not informed,

[8] The result also applies to the general game.

$p_i^{m,u} = \tilde{p}_u$ and $p_i^{m,u} \in \Delta$ are equilibrium candidates, where $p_i^{m,u}$ denotes the price offer made by an uninformed firm i in $t = 1$. Further, because of the firm's ignorance, any deviation does not shape consumer beliefs. Therefore, a deviation is always profitable. Proposition 4.1 proves this intuition.

Proposition 4.1. *Any equilibrium verifies that the period-one firm i acquires information while her competitor j does not.*

Therefore, we can focus on this class of equilibria.

By Assumption 4.1, the consistency condition requires, first, that price taking consumers must rule out the possibility that they are a high risk type whenever accepted and second, that high risk applicants are indeed rejected in the price taking stage. The rough intuition for the first point is that i never wins on H-type bargainers in $t = 2$, since her competitor's offer is not restricted by beliefs. As a result, it must be that either j's offer is accepted in equilibrium or that i's offer is accepted but the outcome is competitive. Put differently, i never wins on H-types in the bargaining stage (i.e. she earns no positive profits on these applicants), and she is therefore not willing to accept losses on the same type in the price taking stage. Further, i's equilibrium profits in $t = 2$ are at least as high on H-types as they are on L-types – on which they are zero in separating equilibria (different prices in $t = 1$) by the Bertrand logic – resulting in the lack of signalling opportunities for the L-type. Following the logic of Villeneuve (2005), price taking H-types must therefore end up rejected. As a result, equilibria are either separating or pooling with both types being rejected in the first place.

At the beginning of the bargaining stage, the uninformed firm shares the price taker's (consistent) belief. Consequently, in separating equilibria, the uninformed firm's information sets on the equilibrium path are singletons: $I^A = \{L\}$ is the information set when a person has been accepted and was offered the equilibrium price for an L-type in $t = 1$. The corresponding information set for a rejection is $I^\Delta = \{H\}$. In pooling equilibria, the rejection decision does not reveal any information, i.e. I^Δ then contains both types. We denote by $p_{j,2}^{m,I^A}$ and $p_{j,2}^{m,I^\Delta}$ equilibrium offers made by the uninformed firm.

The following proposition characterizes the set of possible equilibria given that period-one firm i is informed while her competitor remains ignorant.

Proposition 4.2. *Any equilibrium where firm i has acquired information, verifies that either*

(A) (POOLING) in the price taking stage $p_{i,1}^{m,L} = p_{i,1}^{m,H} \in \Delta$, and in the bargaining

stage, $p_{i,2}^{m,L} \leq p_{j,2}^{m,I^A}$ *and*

$$p_{i,2}^{m,L} \geq L + \frac{\lambda}{1-\lambda}(\tilde{p}_L - L). \text{ Or,}$$

(B) *(SEPARATING) in the price taking stage,* $p_{i,1}^{m,L} = \tilde{p}_L$, $p_{i,1}^{m,H} \in \Delta$, *and in the bargaining stage,* $p_{i,2}^{m,L} = L$ *and* $p_{j,2}^{m,I^A} = L$.
Furthermore, there exists a $\mathring{\lambda} \in (1/2, 1)$ *such that*

(p) $p_{i,2}^{m,H} = p_{j,2}^{m,I^\Delta} = H$ *iff* $\lambda \geq \mathring{\lambda}$, *or*

(np) $p_{j,2}^{m,I^\Delta} \in (H, \tilde{p}_H]$ *and* $p_{i,2}^{m,H} = p_{j,2}^{m,I^\Delta} + \varepsilon$ *or* $p_{i,2}^{m,H} > p_{j,2}^{m,H}$ *if* $p_{j,2}^{m,H} = \tilde{p}_H$,

where ε *is arbitrarily small.* $\mathring{\lambda}$ *is given by*

$$\mathring{\lambda} = \frac{H-L}{H-L+\tilde{p}_L-L}.$$

Part (A) of Proposition 4.2 states that for a pooling equilibrium to arise, period-one firm i must earn strictly positive profits on the L-type in the bargaining stage. The reason is that otherwise, a deviation to an offer which is considered acceptable by some applicant whatever his beliefs is clearly a profitable deviation.

Part (B) of Proposition 4.2 characterizes candidates for separating equilibria, i.e. equilibria where only the H-type is rejected in the first place. Low risk applicants are offered the monopoly price $\tilde{p}_L$ in the price taking stage. It cannot be lower since any deviation from a lower price offer will be profitable and is not restricted by beliefs. Bargaining L-type applicants receive the actuarially fair price. Competition works here because information is revealed already in $t = 1$, which means that both firms are informed in $t = 2$ and price offers are never restricted by beliefs on these types.

By contrast, consider $\theta = H$ and an equilibrium in which the bargaining H-type applicant is offered the actuarially fair price by both firms. This can be an equilibrium only if a deviation on low risk consumers is not profitable in the first round, that is to say, if the firm will not earn higher profits on bargainers by rejecting all applicants in $t = 1$ in order to make them believe that they are high risks. This in turn depends on the share of price takers in the market. A higher share implies that opportunity costs from rejecting low risk applicants are high.

As remarked in the introduction, the informed insurer is said to be persuasive if the period-one firm i would not want to deviate and reject the L-type applicant in $t = 1$ given that the competitive equilibrium arises on H-type bargainers. This implies that if period-one firm i is persuasive, a bargainer that has been rejected in the first period accepts the informed firm's offer in $t = 2$. Here, persuasiveness is given whenever $\lambda \geq \mathring{\lambda}$ (Part (B), (p) of Proposition 4.2).

The competitive outcome is unique whenever persuasiveness is given. Consider instead a consumer who is rejected in the first stage but offered, by the same firm, a high price in the bargaining stage. If the firm is persuasive, this person must believe that he is a high risk. This is inherent in the definition of persuasiveness and the logic of the Intuitive Criterion: since by employing this reject/accept strategy, the informed party would, in case of a deviation to rejecting the L-type, earn less than equilibrium profits on the same person, the consumer must not put positive probability on being a low risk (the deviation is not admissible on the L-type).

On the other hand, if the firm is not persuasive, all equilibria must be such that the informed firm earns nothing on high risk bargainers. This is because the firm would otherwise deviate and reject the L-type in the first place. Proposition 4.2, Part (B), (np) (for 'non-persuasiveness') states that if the informed firm is not persuasive, in all equilibria, the uninformed firm earns strictly positive profits while her informed competitor earns nothing. All these equilibria are such that the consumer, in $t = 2$, upon observation of an offer that he would accept from the informed firm, puts strictly positive probability on $\theta = L$. These beliefs are always allowed by the definition of non-persuasiveness: any deviation $p'_{i,2}$ that would be profitable on the H-type, i.e. $p^{m,I^{\Delta}}_{j,2} \geq p'_{i,2} \geq H$, is also profitable on the L-type.

Proposition 4.2 characterizes equilibria for a given informational state. To see whether these indeed exist, it is necessary to consider incentives to deviate. Consider a pooling equilibrium as presented in Part (A) of the proposition. Clearly, firm j will benefit from clandestinely collect information because in order for this equilibrium to exist, the winning offer implies strictly positive profits on the L-type in the second stage. Note that we have assumed costs for information acquisition to be arbitrarily small.

On the other hand, equilibria as presented in Part (B) of the proposition do not induce firms to deviate from the informational state. However, in order to guarantee existence, λ has to be sufficiently large.

Assumption 4.2. $\lambda \geq \frac{1}{2}$.

Separating equilibria can exist only if this assumption is met. The reason is that otherwise, firm i could still profitably deviate to rejecting the L-type applicant in $t = 1$ and offering him $\tilde{p}_L$ in $t = 2$. Proposition 4.3 summarizes:

Proposition 4.3. *In any equilibrium, the firm to which the consumer has been initially assigned acquires information while the second firm does not. Any equilibrium is fully characterized by Proposition 4.2, Part (B). Such an equilibrium always exists.*

Therefore, although the firms' expected profits are the same at the beginning of the game, (initial assignment in $t = 0$ to each firm takes place with probability $1/2$), they are not once applicants have been assigned. This provides incentives to compete for initial assignment, a topic we will treat in the following section.

4.4 List price competition

In this section, we analyse the game as specified in section 4.2. In the previous section, we have given an intuition for the notion of persuasiveness and we have shown that it may or may not be given if applicants are randomly assigned, that is, if competition in list prices does not take place.

Further, we have found that in all equilibria of the modified game, the period-one firm i is informed while her competitor is not. In order to analyse the game with list price competition, we now assume that the firm which wins the list price competition in $t = 0$ (to which the applicant applies first) acquires information. Afterwards, we check whether this indeed is an equilibrium. Investigating equilibria accordingly essentially simplifies the analysis.

Whether or not the informed firm is persuasive in equilibrium now depends on the lowest list price which is offered in equilibrium. It turns out that with list price competition, persuasiveness is never given in equilibrium and that, although list prices are sometimes asymmetric, all firms expect to earn strictly positive profits.

The informed firm can only be persuasive if she makes strictly positive profits on price taking L-type consumers. Otherwise, rejecting low risks in the price taking stage and thereby hoping to convince consumers and the other firm that the type is H would always be profitable.

Lemma 4.1. *There exists some $\hat{p}(\lambda) > L$ such that persuasiveness is given if and only if $p_i \geq \hat{p}(\lambda)$. It is given by*

$$\hat{p}(\lambda) = \frac{1-\lambda}{\lambda}(H - L) + L$$

This is to say, if the lowest price does not exceed some threshold $\hat{p}(\cdot)$, persuasiveness is not given. In that case, as before, the competitor makes strictly positive profits, $(p^H - H)$

on H bargaining type consumers. Here, p^H denotes the equilibrium price of the modified game, $p_{j,2}^{m,I^{\Delta}}$ if the informed firm is not persuasive, see Proposition 4.2, Part (B), (np).

Now consider some equilibrium price $p_i \geq \hat{p}(\lambda)$, such the firm is persuasive. Because the uninformed firm earns nothing on bargainers, both firms strive to win the list price competition. Hence, competition works effectively until $\hat{p}_L(\lambda)$ is reached. Now consider a possible equilibrium with $p_A = p_B = \hat{p}_L(\cdot)$. Firms have an incentive to deviate to a lower price. However, once one firm does so, the competitor may not wish to keep up. The reason is that as soon as the lowest offer is smaller than $\hat{p}(\cdot)$ the insurer which does not win the list price competition expects to earn strictly positive profits. As a result, persuasive equilibria cease to exist under list price competition, independently of the share of price taking consumers in the market. Therefore, all equilibria must exhibit non-persuasiveness and $p_i < \hat{p}(\lambda)$ in equilibrium.

A 'natural', symmetric equilibrium price would be one where both firms earn the same in expectation. Define the corresponding list price as $\mathring{p}(\cdot)$ with

$$\mathring{p}(\lambda) = \frac{1-\lambda}{\lambda}\frac{q_H}{q_L}(p^H - H) + L.$$

This list price depends on the distribution of risk types in the population and on the share of price takers. More high risks drive the price up, because positive profits are less likely to be made, the same holds true for the share of bargainers.

It turns out that the list price is bound from above by $\mathring{p}(\lambda)$. The reason is that a deviation is always profitable for firm j if her competitor's profits exceeded her owns. It should also not be lower, since otherwise, both firms will find it preferable to embrace the position of the uninformed firm. However, two restrictions apply: first, L-type consumers may not be willing to pay the corresponding price; second, as demonstrated above, the equilibrium list price is necessarily lower than the persuasion threshold $\hat{p}(\lambda)$.

These equilibria exist if an informed firm will not find it profitable to reject L-types in the first place in order offer them $\tilde{p}_L$ in the bargaining stage. To secure this, the following assumption is made.

Assumption 4.3. *Let* $q_H(\tilde{p}_H - H) \geq q_L(\tilde{p}_L - L)$.

This is to say, expected profits of a monopolistic insurer in a market that exhibits symmetric information should be higher on high risk applicants than on low risks. Let $p_{i,2}^{\theta}$ be denoted the informed firm's equilibrium offer to an applicant of risk type θ in the bargaining stage, and $p_{j,2}^{I^A}$ and $p_{j,2}^{I^{\Delta}}$ her competitor's equilibrium offer to applicants that she believes to be L or H-types, respectively. The following proposition specifies candidate equilibria.

Proposition 4.4. *Assume that firm i is informed. Then an equilibrium has the following properties:* $p^L_{i,2} = p^{IA}_{j,2} = L$, $p^{I\Delta}_{j,2} \in (q_L/q_H(\tilde{p}_L - L) + H, \tilde{p}_H]$ *and* $p^H_{i,2} = p^{I\Delta}_{j,2} + \varepsilon$ *or* $p^H_{i,2} > p^H_{j,2}$ *if* $p^H_{j,2} = \tilde{p}_H$*, with* ε *being arbitrarily small.*
In the list price competition,

$$p_i = \min\{\tilde{p}_L; \hat{p}(\lambda) - \varepsilon; \mathring{p}(\lambda)\}, \text{ and } p_j \begin{cases} \geq \hat{p}(\lambda) \text{ if } p_i = \tilde{p}_L \\ = \hat{p}(\lambda) \text{ if } p_i = \hat{p}(\lambda) - \varepsilon \\ = \mathring{p}(\lambda) \text{ if } p_i = \mathring{p}(\lambda). \end{cases}$$

Therefore, as argued above, the lowest list price is the minimum of (a) the maximum willingness to pay for insurance for a low risk, $\tilde{p}_L$, (b) the persuasion threshold, $\hat{p}(\lambda)$, and (c) the price at which both firms' expected profits are equal, $\mathring{p}(\lambda)$.

It is most notable that asymmetric equilibrium list prices arise if the lowest list price is $\tilde{p}_L$ or $\hat{p}(\cdot)$. In the case where $\tilde{p}_L$ is the equilibrium price, one firm will offer some list price that equals or exceeds the persuasion threshold $\hat{p}(\lambda)$, ensuring thereby that the other firm has no incentives to seek to switch roles by offering a higher price and the ignorant firm earns strictly more than her competitor. Also, j has no interest in undercutting the best offer since in expectation, these firm's profits are higher than her competitor's, since $p_i < \mathring{p}(\cdot)$. Asymmetry is also given in (b), that is, when the persuasion threshold is just met. This has been seen above: while one firm has an incentive to undercut the persuasion threshold, the competitor has not if her expected profits are higher. Again, in this case, this is given because $p_i < \mathring{p}(\cdot)$.

As a result, the firm that does not win the list price competition expects to earn at least as much as the winning firm. Figure 4.1 shows in an example how the list price develops in the share of price takers for two different values of q^H. $q^H = 1/3$ is chosen such that Assumption 4.3 is met with equality. It then holds that $\tilde{p}_L = \mathring{p}(0.5)$. For larger shares of high risk applicants, $\tilde{p}_L$ is the equilibrium list price for small values of λ. As the share of price takers increases, however, competition works fine and the list price limits the competitive outcome.

So far, it was assumed that i acquires information. Therefore, it remains to show that the offers as given in Proposition 4.4 indeed form an equilibrium.

Proposition 4.5. *An equilibrium where firm i acquires information and price offers are such as given in Proposition 4.4 exists.*

This is a straightforward result: assume i chooses to remain uninformed. Then she can either accept or reject all applicants. Accepting H-types is not profitable because firms play the competitive outcome on bargaining L-types. Therefore, doing so would imply negative expected profits. On the other hand, rejecting all applicants is also not

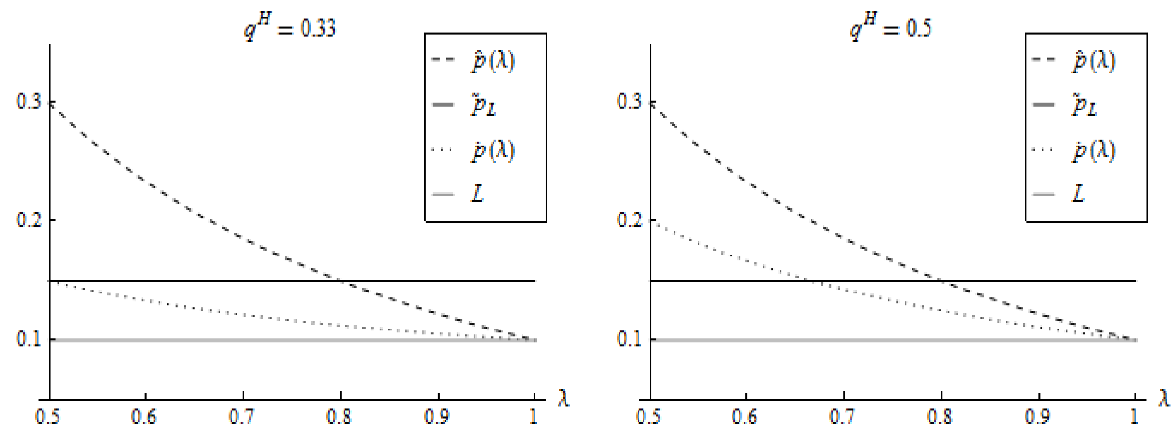

Figure 4.1: $p_i(\lambda)$ for $L = 0.1, \tilde{p}_L = 0.15, H = 0.3, p^H = 0.4$ and different q^H

profitable, since the firm's offer on H-types is never accepted.

Industry profits. Now consider overall industry profits. They are given by

$$\Pi(\lambda) = \lambda q_L(p_i(\lambda) - L) + (1 - \lambda)q_H(p^H - H).$$

Deriving with respect to λ gives

$$\frac{d\Pi(\lambda)}{d\lambda} = q_L(p_i(\lambda) - L) - q_H(p^H - H) + \lambda q_L \frac{dp_i}{d\lambda}. \tag{4.1}$$

The last term of equation (4.1) is non-positive. In particular, as shown in figure 4.1, it is zero for small λ and negative for large λ. Some transformations then reveal that industry profits are decreasing for all $\lambda \geq 1/2$. This is shown in the proof of the following proposition:

Proposition 4.6. $\frac{d\Pi(\lambda)}{d\lambda} \leq 0$ *for all* $\lambda \geq 1/2$.

This means that the more bargainers in the market, the higher are industry profits. This result is independent of the distribution of risk types in the population. The reason for it is that the expected profits of the insurer that has not won the list price competition weakly exceed that of her competitor. This is the case even if the share of high risk applicants is low. This then results in almost competitive list prices, yielding both firms lower profits.

4.5 Discussion

We have analysed the role of bargainers in the market for life insurances. This market exhibits special features: on the one hand, applicants are required to provide information on their bargaining history. On the other hand, since firms dispose of advanced technological means to estimate the statistical risk of applicants they may be better informed than the applicants themselves. This produces a signalling environment in which prices possibly reveal information about risk types.

We have shown that an informed firm may not have the tool to signal high risk to no applicant of that type (neither price taker nor bargainer), in which case we say she is not persuasive. In this case, her offer in the bargaining stage is never accepted, resulting in strictly positive profits for her competitor.

In a modified setting where applicants randomly choose to which firm to apply first, an informed firm can convincingly signal high risk in the bargaining stage if the share of bargaining consumers does not exceed some critical threshold. List price competition however undermines persuasiveness because competition works effectively as long as it is given. It stops to do so only when the lowest list price is such that persuasiveness is not given. As a result, the informed firm is never persuasive and the uninformed firm always earns strictly positive profits. This affects the first competition stage: list prices are not competitive. As a result, and perhaps surprisingly, both firms – also the informed insurer – benefit from this ability to signal high risk only through a non-acceptable offer.

Moreover, the overall industry profit increases in the share of bargainers as their absence fosters competition on the list price level. This is because the firm that does not offer the best list contract will earn strictly positive profits on bargaining applicants exclusively.

We have mentioned in the introduction that insurers usually exchange information about customer data through the means of an information sharing system. Alternatively, or additionally, they require consumers to reveal their application history. Our model helps to understand the use of such procedures: they help firms to gather information on the bargaining behaviour of applicants. In particular, a firm that is not aware of the actual stage of the game may find it useful to always acquire information that leaves firms symmetrically informed in equilibrium. Implications hereof are discussed in the appendix. In particular, symmetric information produces the fully competitive outcome.

In the model studied here, the value of non-persuasiveness is owed to the presence of only two firms in the market. If more firms compete on the bargaining stage, the effect vanishes due to the Bertrand structure. It can be restored, however, if competition is imperfect. In particular, it is visible if the informed firm cannot find any price at which her expected equilibrium profit on a low risk type is higher than the deviation profit on the same type. Otherwise, rejections are not expressed twice, although the persuasion issue may still affect price setting behaviour. We leave this issue for further research.

In order to rationalize rejections in the way we have done, it is necessary to consider a discrete risk type distribution. Alternatively, our model can be interpreted as a continuous type space with imperfect screening technology: consumer willingness to pay will change, but the logic remains the same.

Finally, firms may also offer partial insurance contracts while we have only taken into account full insurance. We believe that this is not so much of a concern here, as applicability seems limited in life or private occupational disability insurances. Furthermore, it is of no relevance to our analysis, as long as risk types are sufficiently distinct.

4.6 Appendix

4.6.1 Proofs

Proof of Proposition 4.1. .
First, there is no equilibrium in which both firms are informed. Assume by contrast that there is a such equilibrium. From Villeneuve (2005), Proposition 6, in equilibrium with two informed firms, it must hold that $p_{i,2}^{m,L} = p_{j,2}^{m,L} = L$ and profits are zero on a bargaining L type. A firm that is not informed can always secure at least zero profits by deviating to $p_{j,2}^{m,H}$. This will be profitable since information acquisition is costly.

Second, there is no equilibrium in which no firm is informed. Consider such an equilibrium. Then, by the standard Bertrand logic, $p_{i,2}^{m} = p_{j,2}^{m} = E[\theta]$ for all applicants in equilibrium. In that case, information acquisition by firm j and her deviating to $p_{j,2}^{L} < E[\theta]$ on the L-type is a profitable deviation.

Third, there is no equilibrium in which i is not informed while j is. Consider a such equilibrium. In such an equilibrium, i offers, in the price taking stage some price p_i^m to all types. The price is acceptable if and only if $p_i^m \leq \tilde{p}_u$. Consider by contrast $p_i^m > \tilde{p}_u$. In that case, no consumer accepts. To acquire information is then always profitable because i can make an offer to a L-type that is considered acceptable, yields strictly positive profits, and does not influence consumer beliefs. By contrast, consider $p_i^m \leq \tilde{p}_u$. All types accept this offer, but profits are strictly negative on H-type consumers. Therefore, acquiring information and rejecting the H-type is a profitable deviation. □

Proof of Proposition 4.2. .
(A) First, consider a pooling equilibrium, i.e. $p_{i,1}^{m,L} = p_{i,1}^{m,H}$. There are two possibilities:

(i) Consider an equilibrium $p_{i,1}^{m,L} = p_{i,1}^{m,H} < H$. But firm i never strictly gains on bargaining H-type consumers: j is not informed. Therefore, j's offer is not restricted by beliefs and i cannot make positive profits on the bargaining H-type. A deviation to the first period offer $p_i^{'} \geq H$ is then always profitable, whatever the hereby induced beliefs.

(ii) $p_{i,1}^{m,L} = p_{i,1}^{m,H} \geq H$ can only be an equilibrium if $p_{i,1}^{m,L}, p_{i,1}^{m,H} \in \Delta$ since $\tilde{p}_u < H$. For this case, consider some equilibrium price $p_{i,2}^{m,L}$ in the bargaining game. Note that (1) as in i), j's offer is not restricted by beliefs and i cannot make positive profits on the bargaining H-type (see also argument below), and (2), if the belief following a deviation in $t = 1$ is L, the uninformed firm j offers L, according to standard Bertrand logic.

According to the logic of the Intuitive Criterion, every deviation to a price greater than H is admissible on H (which means that uninformed parties are allowed to then believe H). If the belief is H, the deviation is profitable. Therefore, the equilibrium is eliminated if the deviation is not admissible on the L-type. This is the case whenever the deviation is not profitable on the L-type given that the belief is H, that is if $(1-\lambda)(p_{i,2}^{m,L} - L) > \lambda(p'_{i,1} - L) + (1-\lambda)(H-L)$ for some deviation offer $p'_{i,1}$. This is never the case, implying that any deviation is admissible on the L-type.

As a result, deviation incentives must be checked. For some $t = 1$ deviation price greater in than H, the offer will not be accepted by some applicant who believes L. Therefore, a deviation in $t = 1$ to some price $p'_{i,1}$ greater than H is never profitable. The equilibrium is not destroyed for deviations to offers of at least H.

Further, a deviation to some price smaller than H is also admissible on the L-type, while not on the H-type. Therefore, the belief on such a deviation must be L. It can be profitable if and only if it is accepted by the L-type. The deviation in $t = 1$ is profitable on the L-type if and only if for some deviation price $p'_{i,1}$:

$$\begin{aligned} & \lambda(p'_{i,1} - L) > (1-\lambda)(p_{i,2}^{m,L} - L) \\ \Leftrightarrow \quad & p_{i,2}^{m,L} < \frac{\lambda}{1-\lambda}(p'_{i,1} - L) + L \end{aligned}$$

The equilibrium is not destroyed if the inequality is not fulfilled for any deviation offer lower or equal to $\tilde{p}_L$. Therefore if (part (A) of the Proposition):

$$p_{i,2}^{m,L} \geq \frac{\lambda}{1-\lambda}(\tilde{p}_L - L) + L.$$

(B) *Separating equilibria.* All other equilibrium candidates must be separating. In that case, for given updated beliefs of j, it must be that $p_{i,2}^{m,L} = p_{j,2}^{m,I^A} = L$. This follows immediately from the standard Bertrand logic.

The informed party makes no positive profits on a bargaining H-type either. Consider by contrast an equilibrium in which a price $\tilde{p}_H \geq p_{i,2}^{m,H} > H$ is accepted by the bargainer with strictly positive probability. Then, it must hold that $p_{j,2}^{m,I^\Delta} \geq p_{i,2}^{m,H}$. In any case, a deviation for j to a lower price is always profitable since j's offer does not influence consumer beliefs.

Consider first possible outcomes in the price taking stage: we show that $p_{i,1}^{m,H} \in \Delta$. By contrast, consider $H \leq p_{i,1}^{m,H} \leq \tilde{p}_H$. This can be an equilibrium only if the offer is considered acceptable in equilibrium (i.e. the consumer believes H) and if firm i does not want to deviate to this offer upon observation of an L-type. But in any equilibrium with $p_{i,1}^{m,L} \neq p_{i,1}^{m,H}$, the informed firm cannot make positive profits on the bargaining L-type. A deviation to a higher price is therefore always profitable.

Also, the offer $p_{i,1}^{m,H} < H$ is not an equilibrium since firm i makes no positive profits on bargaining H-type consumers. A deviation to a higher price is therefore profitable whatever the beliefs of the other players.

Then, we show that $p_{i,1}^{m,L} = \tilde{p}_L$. Assume not and consider $p_{i,1}^{m,L} > \tilde{p}_L$. A deviation in $t = 1$ to $p'_{i,1} = \tilde{p}_L$ is clearly always profitable no matter how this deviation affects beliefs. Similarly, if $p_{i,1}^{m,L} < \tilde{p}_L$, a deviation to $p'_{i,1} = \tilde{p}_L$ is profitable on a L-type, no matter how beliefs are affected.

In the following, we show that either in the bargaining stage, (a) $p_{i,2}^{m,H} = H$ and $p_{j,2}^{m,I^{\Delta}*} = H$ (EQa) or (b) $p_{i,2}^{m,H} > H$ and $p_{j,2}^{m,I^{\Delta}} = \min\{p_{i,2}^{m,H} - \epsilon; \tilde{p}_H\}$ with ε being arbitrarily small (EQb).

We also show that EQa is the unique separating equilibrium when $\lambda \geq \mathring{\lambda}$ while an equilibrium of the form EQb arises if $\lambda < \mathring{\lambda}$ and no other separating equilibrium exists in this case. $\mathring{\lambda}$ is given by

$$\mathring{\lambda} = \frac{H - L}{H - L + \tilde{p}_L - L} > \frac{1}{2}$$

(EQa) EQa is an equilibrium if $\lambda \geq \mathring{\lambda}$: i does not want to deviate on the L-type by definition of $\mathring{\lambda}$.

if $\lambda \geq \mathring{\lambda}$, EQa is the only equilibrium: consider some different candidate and $\lambda \geq \mathring{\lambda}$. First, it is not possible that i offers the best price and makes strictly positive profits on the H-type, since firm j's offer is not restricted by beliefs. Second, consider the case where j wins, i.e. j's H-type offer is strictly smaller than i's. This offer is part of an equilibrium if there exists a profitable deviation that is admissible on the L-type. It is admissible on the L-type whenever $(1 - \lambda)(p_{j,2}^{m,I^{\Delta}} - L) > \lambda(\tilde{p}_L - L)$, where $p_{j,2}^{m,I^{\Delta}} > H$. But i can then always find a profitable deviation to some price offer which is admissible on the L-type and sufficiently close to H.

(EQb) EQb is an equilibrium if $\lambda < \mathring{\lambda}$: if $\lambda < \mathring{\lambda}$, a deviation to "reject L-type, offer winning price" is admissible on the L-type. Therefore, belief L is allowed on such a deviation. Further, a deviation to "reject L-type, offer $\tilde{p}_L$ or lower" is not profitable whenever $\lambda \geq 1/2$.

if $\lambda < \mathring{\lambda}$, EQb is the only equilibrium: obviously, i cannot make a the winning offer in $t = 2$ on the H-type which is strictly greater than H. Also, i offers

H cannot be an equilibrium since i would then deviate to "reject L-type, offer H'.

□

Proof of Proposition 4.3. .

Pooling equilibrium, Part (A) of Proposition 4.2. Firm j wants deviate because the actual profit on an L-type is zero while her deviation profit from acquiring information is $p_{i,2}^{m,L} - \varepsilon - L$ on the same type while the profit on the H-type remains the same.

Separating equilibrium, Part (B) of Proposition 4.2. The ex-ante profit for firm i from acquiring information is $\pi_i^{m,e} = \lambda(\tilde{p}_L - L) > 0$ whereas the deviation profit is $\pi_i^{d,e} = 0$ if the firm rejects the consumer in the first place, and $\pi_i^{d,e} = \lambda(q_L(\tilde{p}_L - L) - q_H(H - \tilde{p}_L)) < \pi_i^{m,e}$ if she offers $\tilde{p}_L$. Any other offer is either not considered acceptable and yields no profits, or yields even lower profits than the offer $\tilde{p}_L$.

Firm j has no interest in acquiring information because the equilibrium is separating. □

Proof of Lemma 4.1. .

From Proposition 4.2, Part (B), if actions are separating on the price taking stage, $p_{i,2}^{m,L} = p_{j,2}^{m,IA} = L$. The same holds for the bargaining stage in the general game. In analogy to the proof of Proposition 4.2, it then follows that the H-type is rejected in the price taking stage. The persuasion logic does not change, therefore, persuasion is given if and only if

$$\lambda(p_i - L) \geq (1 - \lambda)(H - L)$$

$\hat{p}(\lambda)$ is the price at which the equation is strict, i.e. $\hat{p}(\lambda) = \frac{1-\lambda}{\lambda}(H - L) + L$. □

Proof of Proposition 4.4. .

$p_{i,2}^{L}, p_{j,2}^{IA}$ are in analogy to the modified game.

Now consider p_i and p_j. We first show that there is no equilibrium in which $p_i > \min\{\tilde{p}_L, \hat{p}(\lambda) - \varepsilon, \mathring{p}(\lambda)\}$.

(a) Assume $p_j \geq p_i > \tilde{p}_L$. Then, the game ends and all firms earn nothing. A deviation in $t = 0$ to $p_i' = \tilde{p}_L$ is profitable.

(b) Assume $p_j \geq p_i \geq \hat{p}(\cdot)$. Persuasiveness is not given and j earns nothing if $p_j > p_i$ and earns $\frac{1}{2}0 + \frac{1}{2}q_L(p_j - L)$ if $p_j = p_i$. A deviation to an arbitrarily smaller price p_j yields $q_L(p_j - L)$ which is greater.

(c) Assume $p_j \geq p_i > \mathring{p}(\cdot)$. In that case, i expects to earns more than j. Therefore, a deviation by j to a lower list price $p_j' < p_i$ is strictly profitable.

Second, we show that there is no equilibrium in which $p_i < \min\{\tilde{p}_L, \hat{p}(\lambda), \mathring{p}(\lambda)\}$.

(a) Assume $p_j \geq p_i < \tilde{p}_L$ and $\tilde{p}_L \leq \min\{\hat{p}(\cdot), \mathring{p}(\cdot)\}$. Then, a deviation to a higher price is always profitable: if $p_j > p_i$, a deviation by i to $p_i' = \tilde{p}_L$ is profitable. If $p_j = p_i$, a deviation to some higher price is profitable because persuasiveness is not given and the uninformed firm earns more than the informed party.

(b) Assume $p_j \geq p_i < \hat{p}(\cdot) - \varepsilon$ and $\hat{p}(\cdot) - \varepsilon \leq \min\{\tilde{p}_L, \mathring{p}(\cdot)\}$. Then, persuasiveness is given and the uninformed firm earns more than the informed firm. Therefore, a deviation to a higher price $p_i = \hat{p}(\cdot)$ is always profitable.

(c) $p_j \geq p_i < \mathring{p}(\cdot)$ and $\mathring{p}(\cdot) \leq \min\{\tilde{p}_L, \hat{p}(\cdot)\}$. Then, the informed firm earns less than the uninformed party and persuasiveness is given. A deviation by i to $p_i' = \mathring{p}(\cdot)$ is always profitable.

Next, we investigate deviation incentives for $p_i = \min\{\tilde{p}_L, \hat{p}(\lambda) - \varepsilon, \mathring{p}(\lambda)\}$.

(a) $\min\{\tilde{p}_L, \hat{p}(\cdot) - \varepsilon, \mathring{p}(\cdot)\} = \tilde{p}_L$.
Consider some $p_j^* \geq \hat{p}(\cdot)$. j does not want to deviate since $\pi_j^e > \pi_i^e$. i's best response is $p_i(p_j^*) = \tilde{p}_L$. Therefore, $p_i^* = \tilde{p}_L$ and any price $p_j^* \geq \hat{p}(\cdot)$ is an equilibrium.
By contrast, consider some equilibrium price $p_j^* < \hat{p}(\cdot)$. But i's best response is $p_i(p_j^*) > p_j^*$.

(b) $\min\{\tilde{p}_L, \hat{p}(\cdot) - \varepsilon, \mathring{p}(\cdot)\} = \hat{p}(\cdot) - \varepsilon$.
Consider $p_j = \hat{p}(\cdot)$. j does not want to deviate since her expected profit is higher than i's. i's best response is $p_i(p_j) = \hat{p}(\cdot) - \varepsilon$. Therefore, $p_i = \hat{p}(\cdot) - \varepsilon$, $p_j = \hat{p}(\cdot)$ is an equilibrium.
By contrast, consider $p_j = \hat{p}(\cdot) - \epsilon$. i's best response is $p_i(p_j) > p_j$.
Further, consider $p_j > \hat{p}(\cdot)$. This cannot be an equilibrium. i's best response to p_j is $p_i(p_j) = \min\{\tilde{p}_L; p_j - \varepsilon\}$.

(c) $\min\{\tilde{p}_L, \hat{p}(\cdot) - \varepsilon, \mathring{p}(\cdot)\} = \mathring{p}(\cdot)$.
Consider $p_j = \mathring{p}(\lambda)$. j does not want to deviate because her profits are higher than i's. i's best response is $p_i(p_j) = p_j$. Therefore, $p_i = p_j = \mathring{p}_L$ is an equilibrium.
By contrast, consider $p_j > \mathring{p}(\lambda)$. i's best response is $p_i(p_j) = \min\{p_j - \varepsilon; \tilde{p}_L\}$.

Last, for given p_i, we have to check deviation incentives in $t = 1$. Persuasiveness is never given, which implies that the deviation profits i can attain by rejecting the applicant in the first place are at most $(1-\lambda)(\tilde{p}_L - L)$. Then, it must hold that

$$\lambda(p_i - L \geq (1-\lambda)(\tilde{p}_L - L) \tag{4.1}$$

(a) $p_i = \tilde{p}_L$. $\lambda(p_i - L) \geq (1-\lambda)(\tilde{p}_L - L)$ since $\lambda \geq 1/2$.

(b) $p_i = \hat{p}(\lambda)$. Inserting this into equation (4.1) gives $H \geq \tilde{p}_L$ which is given.

(c) $p_i = \mathring{p}(\lambda)$. Inserting this gives $q_H(p^H - H) \geq q_L(\tilde{p}_L - L)$. This is given if and only if $q_L/q_H(\tilde{p}_L - L) + H$.

□

Proof of Proposition 4.5. .
In the text. Further, rejecting all types and offering them $\tilde{p}_L$ in $t = 2$ is not a profitable deviation since acquiring information and rejecting L-types and offering them $\tilde{p}_L$ in $t = 2$ is not a profitable deviation: not being informed would imply losses on the H-type bargaining applicants. □

Proof of Proposition 4.6. .

- $p_i = \tilde{p}_L$: $\partial\Pi/\partial\lambda = q_L(\tilde{p}_L - L) - q_H(p^H - H)$. If $p_i = \tilde{p}_L$, it must hold that $\tilde{p}_L \leq \mathring{p}(\lambda)$. Inserting $\mathring{p}(\lambda)$ for $\tilde{p}_L$ gives that

$$\begin{aligned}\frac{\partial\Pi}{\partial\lambda} &\leq q_L(\frac{1-\lambda}{\lambda}\frac{q_H}{q_L}(p^H - H) + L - L) - q_H(p^H - H) \\ &= \frac{1-2\lambda}{\lambda} q_H(p^H - H) \leq 0.\end{aligned}$$

- $p_i = \hat{p}(\lambda)$: inserting $\hat{p}(\cdot)$ and $\partial p_i/\partial\lambda = \partial\hat{p}(\lambda)/\partial\lambda = -1/\lambda^2(H - L)$ into equation (4.1) in the main text gives

$$\frac{\partial\Pi}{\partial\lambda} = -q_L(H - L) - q_H(p^H - H) < 0.$$

- $p_i = \mathring{p}(\lambda)$. Similarly, inserting $\mathring{p}(\cdot)$ and $\partial\mathring{p}(\lambda)/\partial\lambda = -1/\lambda^2(p^H - H)$ gives

$$\frac{\partial\Pi}{\partial\lambda} = -2q_H(p^H - H) < 0.$$

□

4.6.2 Information sharing systems

When information is shared, one would naturally assume asymmetries of information to disappear. This depends however on the type of information that is being shared. In insurance markets, only little information is transferred on the risk type of a prospective customer,[9] leading us to conclude that these systems serve to inform competitors about recent applications, thereby preventing them from screening clients. This in turn implies that firms are symmetrically informed in equilibrium with customers being aware hereof. To see this, consider the following modified game setting (modifications in bold text):

(t=0) With probability 1/2, nature assigns consumers to firm i for $i \in \{A, B\}$. **This is private information to applicants**.

t=1 Price taking stage: **Both firms** make an investment choice $\gamma_A, \gamma_B \in \{0, \gamma\}$ and contingently observe θ. The firms publicly announce price offers p_A^S, p_B^S. Price takers only observe the offer from the firm to which they have been assigned in $t = 0$, and accept the offer or choose no insurance.

t=2 Bargaining stage: Bargainers ask both firms to make better offers: firms choose whether to acquire information on consumers that have not been assigned to them. Firms simultaneously offer $p_{A,2}^S$, $p_{B,2}^S$. Bargainers choose among all offers, i.e. p_i^S, $p_{i,2}^S$, p_j^S, $p_{j,2}^S$ and no insurance.

This game specification is a translation from the following idea: applicants choose one firm where they apply in the first place. Bargainers will then approach the competing insurer to ask for a new offer. This second firm, since not aware of being faced with a bargainer, will acquire information and make an offer for the price taker. It can easily be shown that with informational costs being sufficiently small, acquiring information is the only equilibrium outcome here. As before, firms then compete for bargainers in the last stage.

A similar effect is generated by assuming commitment power: although information is fully revealed in equilibrium, the lack of commitment power towards competitors is a driver of the results. In particular, consider the case where the competitor is truthfully informed about the true risk type. Then, an informed firm is always persuasive and insurers are symmetrically informed in equilibrium (even if the second firm did not invest in information). As a result, information sharing without commitment power serves to increase asymmetries of information instead of reducing them.

[9] In particular, only information of pass/fail decisions are being shared

This section sketches implications of a market in which firms are symmetrically informed. As mentioned above already, Villeneuve (2005) has demonstrated that two firms may, in a subset of the set of existing equilibria that have survived elimination according to the Intuitive Criterion, make positive profits on bargaining consumers if they are symmetrically informed. These equilibria however appear unintuitive.

By contrast, Bagwell and Ramey (1991) propose a two-step equilibrium refinement that turns out to uniquely select the efficient separating equilibrium. They propose the following: Whenever the uninformed party (parties) observe(s) an off-path behaviour, he (they) then count the number of deviating (informed) firms that are necessary to reach this (these) particular disequilibrium offer(s). The belief will then be according to the lower number of deviations. If this lowest number is equal for more than one equilibrium, the Intuitive Criterion applies,[10] the criterion therefore being consistent with previous analysis.

A Perfect Bayesian Equilibrium then requires the following:

(1) Beliefs are consistent with Bayes' rule: whenever both firms are informed, and equilibrium pricing is fully separating, then the consumer must, if he observes some price vector that is only played for given θ, believe $\hat{\theta}^t(\cdot) = \theta$. If equilibrium pricing is pooling for some θ', θ'', then the consumer believes $\hat{\theta}^t(\cdot) = \frac{\Sigma_{\theta',\theta''} q_\theta \theta}{\Sigma_{\theta',\theta''} q_\theta}$.

(2) Consider a disequilibrium pricing tuple $(P_i', P_j') \neq (P_i(\theta), P_j(\theta))$ (the equilibrium price vectors of both firms), and denote $N(\theta) \in \{1, 2\}$ the number of informed deviating firms required to generate (P_i', P_j'). Then, $\hat{\theta}^2(P_i', P_j') = \theta'$ if $N(\theta') < N(\theta'')$.

(3) Consider a disequilibrium pricing strategy (P_i', P_j') for which $N(\theta') = N(\theta'')$. Then the Intuitive Criterion applies.

From this equilibrium selection rule, it follows that bargaining consumers never end up uninsured. By contrast, consider the case where the H-type is rejected by both firms and denote $p_{i,t}^{S,\theta}$ an equilibrium candidate offer made by firm i to an applicant of type θ in period t. Then, $(p_{A,1}^{S,H}, p_{A,2}^{S,H}, p_{B,1}^{S,H}, p_{B,2}^{S,H}) \in \Delta$. A firm would like to deviate to an offer that is profitable on H and that is considered acceptable by the consumer given his beliefs. Because rejections are played only on $\theta = H$, the consumer belief $\hat{\theta}^2(P_i \in \Delta, P_j \in \Delta)$ must be H by the Bayes-consistency requirement. Therefore, if firm i deviates to a price $\tilde{p}_H \geq p_{i,2}' > H$, $N(H) = 1$, whereas $N(L) = 2$ and the off-bath belief must be H too. But then, $(p_{A,1}^{S,H}, p_{A,2}^{S,H}, p_{B,1}^{S,H}, p_{B,2}^{S,H}) \in \Delta$ is not an equilibrium. The following are unique

[10] Bagwell and Ramey (1991) define a very similar concept which they call the ϵ-Intuitive Criterion

equilibrium allocations, independent of the firm to which the applicant has been assigned in the first place (superscripts are being dropped again).

Proposition 4.7. . *The following contracts are offered in equilibrium:* $p^{S,H}_{i,1} = p^{S,H}_{j,1} \in \Delta$, $p^{S,L}_{i,1} = p^{S,L}_{j,1} = \tilde{p}_L$ *for* $i \in \{A, B\}$ *and* $p^{S,H}_{i,2} = p^{S,H}_{j,2} = H$, $p^{S,L}_{i,2} = p^{S,L}_{j,2} = L$.

Proof. The unique equilibrium in $t = 2$ is the perfect competition outcome. Suppose not, then there exists a θ such that $p^{S,\theta}_{A,2}, p^{S,\theta}_{B,2} > \theta$ for some θ. Let A be the period-one firm. Both firms wish to deviate to a lower price. If this price is not played for any other θ, denoted θ', in equilibrium, then $N(\theta) = 1$ and $N(\theta') = 2$. Therefore, beliefs do not change and the deviation is profitable. On the other hand, there is no pooling equilibrium in $t = 2$: suppose the type has not been revealed in the first stage, then the firm to which the applicant has not been assigned in the first place has always an incentive to deviate to $\tilde{p}_L < H$ (it was assumed that with symmetric offers, an applicant prefers the initial firm) if the equilibrium price is at least H. On the other hand, prices lower than H cannot be offered in equilibrium, since firms would always want to deviate on the H-type. Further, if the type has been revealed in the first place, offers cannot be pooling since $\tilde{p}_u < H$. The logic for the equilibrium in $t = 1$ follows the proof of Proposition 4.2. $\square$

Bibliography

Adriani, F. and L. G. Deidda (2011). Competition and the signaling role of prices. *International Journal of Industrial Organization 29*(4), 412–425.

Akerlof, G. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 488–500.

Albano, G. and A. Lizzeri (2001). Strategic certification and provision of quality. *International Economic Review 42*(1), 267–283.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica 21*, 503–546.

Bagwell, K. and G. Ramey (1991). Oligopoly limit pricing. *The RAND Journal of Economics 22*(2), 155–172.

Bagwell, K. and M. H. Riordan (1991). High and declining prices signal product quality. *American Economic Review 81*(1), 224–39.

Battaglini, M. (2002). Multiple referrals and multidimensional cheap talk. *Econometrica 70*(4), 1379–1401.

Betts, J. R. and F. Grogger (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review 22*(4), 343–352.

Biglaiser, G. (1993). Middlemen as experts. *The RAND Journal of Economics*, 212–223.

Boleslavsky, R. and C. Cotton (2014). Grading standards and education quality. *American Economic Journal: Microeconomics*. Forthcoming.

Bolton, P., X. Freixas, and J. Shapiro (2012). The credit ratings game. *Journal of Finance 67*(1), 85–112.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics 127*(3), 1243–1285.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013). Salience and consumer choice. *Journal of Political Economy 121*(5), 803–843.

Cameron, J. and R. Eisenberger (1996). Detrimental effects of reward, reality or myth? *American Psychologist 51*(11), 1153–1166.

Chan, W., L. Hao, and W. Suen (2007). A signaling theory of grade inflation. *International Economic Review 48*(3), 1065–90.

Cho, I.-K. and D. M. Kreps (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics 102*(2), 179–221.

Costrell, R. M. (1994). A simple model of educational standards. *American Economic Review 84*(4), 956–971.

Crawford, V. and J. Sobel (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 1431–1451.

Dasgupta, P. and E. Maskin (1986). The existence of equilibrium in discontinuous economic games, i: Theory. *Review of Economic Studies 53*(1), 1–26.

de Garidel-Thoron, T. (2005). Welfare-improving asymmetric information in dynamic insurance markets. *Journal of Political Economy 113*(1), 121–150.

Desai, P. S. and D. Purohit (2004). "let me talk to my manager": Haggling in a competitive environment. *Marketing Science 23*(2), 219–233.

Dubey, P. and J. Geanakoplos (2010). Grading exams: 100, 99, 98,... or A, B, C? *Games and Economic Behavior 69*(1), 72–94.

Ellingsen, T. (1997). Price signals quality: The case of perfectly inelastic demand. *International Journal of Industrial Organization 16*(1), 43–61.

Farhi, E., J. Lerner, and J. Tirole (2014). Fear of rejection? Tiered certification and transparency. *The RAND Journal of Economics*. Forthcoming.

Faure-Grimaud, A., E. Peyrache, and L. Quesada (2009). The ownership of ratings. *The RAND Journal of Economics 40*(2), 234–257.

Figlio, D. N. and M. E. Lucas (2004). Do high grading standards affect student performance? *Journal of Public Economics 88*(9–10), 1815–1834.

Gal-Or, E. (1985). Information sharing in oligopoly. *Econometrica 53*(2), 329–343.

Gentzkow, M. and J. M. Shapiro (2006). Media bias and reputation. *Journal of Political Economy 114*(2), 280–316.

Gill, D. and J. Thanassoulis (2009). The impact of bargaining on markets with price takers: Too many bargainers spoil the broth. *European Economic Review 53*(6), 658–674.

Gill, D. and J. Thanassoulis (2013). Competition in posted prices with stochastic discounts. working paper 639, University of Oxford, Department of Economics.

Harbaugh, R. and E. Rasmusen (2013). Coarse grades: Informing the public by withholding information. *working paper*.

Hauswald, R. and R. Marquez (2006). Competition and strategic information acquisition in credit markets. *Review of Financial Studies 19*(3), 967–1000.

Healy, C. and S. Rojstaczer (2012). Where A is ordinary: the evolution of American college and university grading, 1940 – 2009. *Teachers College Record 114*(7), 1–23.

Inderst, R. (2008). Irresponsible lending with a better informed lender. *Economic Journal 118*(532), 1499–1519.

Inderst, R. and H. M. Mueller (2006). Informed lending and security design. *Journal of Finance 61*(5), 2137–2162.

Janssen, M. C. and S. Roy (2010). Signaling quality through prices in an oligopoly. *Games and Economic Behavior 68*(1), 192–207.

Kahnemann, D. and A. Tversky (1979). Prospect theory: an analysis of decision under risk. *Econometrica 47*, 263–292.

Kartasheva, A. and B. Yilmaz (2013). Precision of ratings.

Kőszegi, B. and A. Szeidl (2013). A model of focusing in economic choice. *The Quarterly Journal of Economics 128*(1), 53–104.

Kofman, F. and J. Lawarrée (1993). Collusion in hierarchical agency. *Econometrica 61*(3), 629–56.

Kühn, K.-U. and X. Vives (1995). *Information exchanges among firms and their impact on competition.* Number June. Office for Official Publications of the European Communities.

Levin, J. (2003). Relational incentive contracts. *American Economic Review 93*(3), 835–857.

Lizzeri, A. (1999). Information revelation and certification intermediaries. *The RAND Journal of Economics 30*(2), 214–231.

Mailath, G. and L. Samuelson (2012). *Repeated Games and Reputations: Long-run Relationships.* Oxford University Press.

Mathis, J., J. McAndrews, and J. Rochet (2009). Rating the raters: are reputation concerns powerful enough to discipline rating agencies? *Journal of Monetary Economics 56*(5), 657–674.

Milgrom, P. and J. Roberts (1986). Price and advertising signals of product quality. *Journal of Political Economy 94*(4), 796–821.

Oettinger, G. S. (2002). The effect of non-linear incentives on performance: Evidence from "econ 101". *The Review of Economics and Statistics 84*(3), 509–517.

Ostrovsky, M. and M. Schwarz (2010). Information disclosure and unraveling in matching markets. *American Economic Journal: Microeconomics 2*(2), 34–63.

Raskovich, A. (2007). Competition or collusion? Negotiating discounts off posted prices. *International Journal of Industrial Organization 25*(2), 341–354.

Rayo, L. and I. Segal (2010). Optimal information disclosure. *Journal of Political Economy 118*(5), 949–987.

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The Quarterly Journal of Economics 98*(4), 659.

Sharpe, S. A. (1990). Asymmetric information, bank lending and implicit contracts: A stylized model of customer relationships. *The Journal of Finance 45*(4), 1069–1087.

Sobel, J. (1985). A theory of credibility. *Review of Economic Studies 52*(4), 557–73.

Sobel, J. (2009). Signaling games. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science*. Springer New York.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

Spulber, D. F. (1995). Betrand Competition when Rivals' Costs are Unknown. *The Journal of Industrial Economics 43*(1), 1–11.

Strausz, R. (2005). Honest certification and the threat of capture. *International Journal of Industrial Organization 23*(1-2), 45–62.

Tirole, J. (1986). Hierarchies and bureaucracies: On the role of collusion in organizations. *Journal of Law, Economics, and Organization 2*(2), 181–214.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review 76*(1), 31–48.

Villeneuve, B. (2005). Competition between insurers with superior information. *European Economic Review 49*(2), 321–340.

Vives, X. (1984). Duopoly information equilibrium: Cournot and bertrand. *Journal of Economic Theory 34*(1), 71–94.

Vives, X. (2008). Information sharing among firms. *The New Palgrave Dictionary of Economics*.

List of Figures