From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2020

ESSAYS IN BEHAVIORAL INFORMATION ECONOMICS: OVERCONFIDENCE, SOPHISTICATION, NAÏVETÉ

Ostrizek, Franz Peter

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01n296x206j

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS IN BEHAVIORAL INFORMATION ECONOMICS: OVERCONFIDENCE, SOPHISTICATION, NAÏVETÉ

FRANZ PETER OSTRIZEK

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF ECONOMICS

ADVISOR: ROLAND BÉNABOU

SEPTEMBER 2020

© Copyright by Franz Peter Ostrizek, 2020.

All rights reserved.

Abstract:

This Dissertation comprises three chapters reconsidering questions of information economics in light of behavioral biases exhibited by workers and consumers.

In the first chapter, I analyze how firms design feedback and bonuses when workers are uncertain or even misguided about their own ability. I analyze a dynamic principal-agent setting in which both sides learn about the importance of effort. More precise performance evaluation reduces current agency costs but promotes learning, which is shown to increase future agency costs. The optimal evaluation technology is both imprecise and tough: a bad performance is always sanctioned, but a good one is not always recognized. We also study the case in which the agent has incorrect beliefs about his abilities. If the agent is overconfident, the principal uses a tough evaluation structure to preserve the agent's profitable misperception.

In the second chapter, joint with Denis Shishkin, we analyze screening with frame-dependent valuations. A monopolist principal designs an extensive-form decision problem with frames at each stage. This allows the firm to induce dynamic inconsistency and thereby reduce information rents. The optimal extensive form has a simple three-stage structure and uses only the two highest frames (high-low-high). Some types buy in the first stage, others buy at the last stage. Sophisticated consumers correctly anticipate that if they deviated, they would choose a decoy, which they want to avoid in a lower frame. This eliminates incentive compatibility constraints into types who don't buy in the first stage. With naive consumers, the principal can perfectly screen by cognitive type and extract full surplus from naifs.

In the third chapter, I consider a market where some consumers neglect follow-up fees. Why would firms increase price transparency towards such naïve consumers? We show that exploiting consumer naïveté decreases profits if naïfs are more price sensitive than sophisticates. Firms desire to be transparent towards the marginal consumer, thereby increasing the equilibrium price, while leaving their own captured consumers in the dark.

Increased transparency is most valuable to consumers when firms have the least incentives to engage in it, can increase aggregate welfare, but typically redistributes surplus from consumers to firms.

Acknowledgments:

Completing this dissertation would not have been possible without the support and encouragement of many people.

I am immensely grateful to my adviser Roland Bénabou for invaluable guidance and a seemingly endless generosity with his time over these past six years. Aside from my adviser, several professors have served an outsized role in my progress as a student here. I am particularly thankful to Pietro Ortoleva, Wolfgang Pesendorfer and Leeat Yariv. As I look at my work, I conclude that Stephen Moris has had quite an influence as well.

Thank you to my colleagues, in particular Leon Musolff, Evgeniy Safonov and Andrei Zeleneev for many a tea and chat in 200; and my coauthors Elia Sartori and Denis Shishkin. I have learnt immensely in countless conversations with them. Elia always brightens up tedious work, not least with his frenzied manipulation of Mathematica files leading to many hours of reacting with disbelief, elation and despair to the output of this ingenious and infuriating piece of software. Special thanks to Denis for putting up with a coauthor who switches between being convinced to have found a proof and a counterexample for the very same result every few minutes; and for pastries from *The Little Chef*.

Finally, I want to thank my family and friends for always being compassionate and supportive throughout my graduate school experience.

For my mother

Contents

Abstract iii

Acknowledgments v

List of Figures x

1 Vague by Design: Performance Evaluation and Learning from Wages 1

1.1 Introduction 1

1.2 The Model 9

1.3 Analysis 12

1.3.1 Transformation to Belief Space 12

1.3.2 Terminal Period 15

1.3.3 Initial Period 17

1.3.4 Analysis of the Solution and Comparative Statics 25

1.4 Preserving and Correcting Misperceptions 26

1.4.1 Contracting with Heterogeneous Priors 28

1.4.2 Terminal Period 30

1.5 Discussion and Extensions 34

1.5.1 Private Information Acquisition 34

1.5.2 Unobservable Effort 36

1.5.3 Long-Run Commitment 39

1.6 Concluding Remarks 40

References 41

1.A Proofs 47

2 Screening with Frames **59**

2.1 Introduction 59

2.2 Screening with Frames and Extensive Forms 67

2.2.1 Contracts and Frames 67

2.2.2 Extensive-Form Decision Problems 70

2.2.3 The Firm's Problem 71

2.3 Optimal Screening 72

2.3.1 Optimal Structure of the Extensive Form 73

2.3.2 Necessary and Sufficient Conditions for Implementation 76

2.3.3 Optimal Contracts 80

2.3.4 Optimal Concealed Types 83

2.3.5 Discussion 85

2.4 Extensions 88

2.4.1 Naivete 88

2.4.2 Additional Participation Constraints and Cool-off Regulation 93

2.5 Conclusion and Discussion 96

References 98

2.A Preliminary Lemmata 102

2.B Proofs 105

3 Hidden Costs: Anticompetitive Effects of Transparency **123**

3.1 Introduction 123

3.2 The Model 129

3.3 Static Equilibrium 131

3.3.1 The Distributional Impact of Consumer Naïveté 134

3.4 Price Transparency 135

3.4.1 Short-run Incentives for Transparency . . . 136
3.4.2 Long-run Transparency: Consumer Naïveté as a State Variable . . . 138
3.4.3 Long-run Transparency: Broad Campaigns . . . 140
3.4.4 Targeting Transparency . . . 143
3.5 Extensions . . . 146
3.5.1 Extensive Margin . . . 146
3.5.2 Transformative Transparency . . . 151
3.6 Discussion . . . 152
3.6.1 Transparency in Shrouded Markets . . . 153
3.6.2 Discussion of Assumptions . . . 154
3.6.3 Identification of the Summary Statistic $\alpha_0 - \alpha$ . . . 156
3.7 Conclusion . . . 157
References . . . 158
3.A Proofs . . . 162

List of Figures

1.3.1 The concavification of ℓ^* at μ. 24

1.4.1 The effect of information on the principal's continuation value. 31

2.1.1 Example 2.1 62

2.3.1 A standard EDP for $(c_{\theta^1},\ldots,c_{\theta^5})$ with $\Theta_R=\{\theta^2,\theta^3,\theta^5\}$ and $\Theta_C=\{\theta^1,\theta^4\}$. 74

2.3.2 The construction of e_θ. 79

2.3.3 Optimal Θ_C for $\theta^1=(1,3),\theta^2=(4,5),\theta^3=(5,6)$. 83

2.4.1 The optimal extensive-form decision problem in Example 2.2. 90

2.4.2 Interim and Ex-post Participation Constraints in Frame π 94

Chapter 1

Vague by Design: Performance Evaluation and Learning from Wages[1]

1.1 Introduction

Many firms motivate their workers to exert effort with incentive pay based on objective measures of performance.[2] Such measures have become richer and easier to obtain as a result of the IT revolution. For example, the availability of board computers and GPS tracking allows for better monitoring of truck drivers and time tracking software in law firms and other offices not only simplifies billing but also logs the activities of employees. Shop floor control systems monitor not only the flow of goods but also allow the tracking of workers. Improved natural language processing enables data collection in applications

[1]I am grateful to Roland Bénabou, Pietro Ortoleva, Wolfgang Pesendorfer for their continued guidance and to Ludmila Matysková, Sofia Moroni, Leon Musolff, Pellumb Reshidi, Evgenii Safonov, Denis Shishkin, Nikhil Vellodi, Can Urgun, Leeat Yariv and seminar audiences at Princeton University for helpful comments and discussions.

[2]According to data from the Bureau of Labor Statistics' National Compensation Survey, 39% of hours worked in US private sector firms in 2013 were in jobs with performance-related pay. 21% fall into a narrower classification of performance-related pay excluding, among other categories, referral bonuses which should arguably should be excluded from a theoretical perspective, but also safety bonuses which should be retained (Gittleman and Pierce, 2013).

ranging from call centers to health care.[3] Based on such statistics, firms can arrive at better objective measures of workers' contribution to profits.

Should this additional information be used to set performance pay? The theory of incentives seems to offer a simple answer to this question. Providing incentives to workers is costly and may cause inefficiencies because the underlying performance measures are only partially informative about effort. Conversely, more hard information about the workers' contribution is always helpful and should be used as a basis of performance pay (Holmström, 1979; Grossman and Hart, 1983). In particular, it is never optimal to base incentives on a noisy signal instead of the contribution to output.[4]

In this paper, we show that learning changes this conclusion fundamentally. The firm prefers workers to remain uninformed about their match-specific ability since uninformed workers are cheaper to motivate on average. Contingent wages, however, reveal information about output and hence their ability. As a result, the principal wants to base incentive pay not on the agent's performance directly but on a noisy signal instead.

To illustrate how explicit incentives invariably reveal information, suppose a worker's contract promises a bonus upon a sufficiently high customer satisfaction score. The worker does not observe this score directly. If he receives the bonus, however, he knows that he cleared the threshold while if he does not receive it, he learns that he fell short of it.

Why would it be profitable to conceal information about their performance from workers? The costs of providing such incentives itself provides a rationale. When ability and effort are complements in production, an agent who believes his ability is high only requires a small bonus to motivate him to exert high effort and vice versa. The impact of a given change in beliefs is amplified if it is large relative to the expected efficiency of effort. Therefore, it has a large impact at a low posterior, while it has a smaller impact at a high

[3] These tools allow call centers to detect the callers' mood, for example (Singer, 2013). For a survey on the use of natural language processing to extract information from health-related text, see Gonzalez-Hernandez et al. (2017).

[4] That is not optimal to add noise continues to hold across a wider class of models, including multi-tasking (Holmström and Milgrom, 1991) and linear-normal career concerns models Holmström (1999); Hörner and Lambert (forthcoming).

posterior. This implies that learning, which causes a mean-preserving spread of posterior beliefs, is costly on average: at low beliefs, the required bonus increases a lot, while at high beliefs, the required bonus decreases only a little.[5]

To capture this fundamental trade-off between incentives and information, we develop a model of twice-repeated moral hazard with learning. The agent's type affects not only mean output, but also the effectiveness of effort. The quality of output is not observed directly. Instead, in every period the principal designs not only wages but also the underlying performance evaluation. The agent observes his evaluations and wages. The evaluation structure therefore determines not only the cost of incentives this period, but also the extent of learning.[6] We analyze the model in Section 1.3, transforming the contracting problem into an information design problem with additional constraints (participation and incentive compatibility) and an additional choice variable, the wage at every posterior.

In the final period, when the continuation belief is of no importance to the principal, the optimal evaluation structure is fully informative. In the first period, however, there is a novel tradeoff. A more precise evaluation structure reduces agency costs this period, but induces more learning, thereby increasing agency costs in the next period. We solve for the optimal contract and show that it features a *binary evaluation structure*: The additional motive of shaping learning does not add complication relative to the fully informative evaluation. The optimal evaluation structure is "*tough*": The agent obtains a high evaluation and therefore a bonus only if his output is high quality. Even if output was high, however, he may receive a low evaluation and thus fail to obtain the bonus. After low quality output, the latter outcome is guaranteed. This information structure avoids inducing very low posteriors. Agents with such beliefs would be very expensive to motivate in the next period

[5]Since ability also affects the baseline probability of high output, the exact condition is slightly more complicated and is implied by log-supermodularity, as will be discussed later.

[6]Of course, this mechanism is predicated on two background conditions: First, the agent observes his wages. In particular, it is not possible to record the agents performance measures and provide incentives only at the end of the employment relationship. Second, information that is not used as the basis of explicit incentives does not have to be revealed to the agent in any other way. We discuss these issue in more detail after introducing the model formally in Section 1.2 and provide extensions in Section 1.5.

and even a small increase in their posterior belief has a large (decreasing) effect on the required bonus.

The influence of explicit incentives on the agent's learning becomes even more essential when the agent is initially misguided. Indeed, the learning environment then also shapes the evolution of mean beliefs, making it important for the principal to preserve profitable worker misconceptions and eliminate costly ones, and for the analyst to determine the persistence of such misconceptions. In particular, a substantial empirical and experimental literature suggests that people are often overconfident about their ability, the degree of control they have over their environment, or the extent to which they live in a "just world" that rewards effort in the long run. In Section 1.4, we therefore analyze the model with heterogeneous beliefs, allowing the agent to be optimistic or pessimistic about his type.[7] We show that a noisy and tough information structure remains optimal in the face of overconfidence: The best way to preserve profitable optimism is not "coddling" grade inflation, but tough evaluation. If the agent is pessimistic, the principal is still averse to a dispersion in beliefs, but wants to eliminate costly pessimism: If the latter effect dominates, the principal uses a fully informative evaluation to promote learning. If the latter effect dominates, the optimal evaluation structure is now *lenient*: Sometimes a bad performance nonetheless receives a good evaluation and is rewarded with a bonus.

In Section 1.5 we consider several extensions of our model. We show that the optimal evaluation remains partially informative and tough if the principal can acquire private information about the agent's performance, if effort is unobserved in addition to being noncontractible, and when the principal can commit across time periods. Section 1.6 concludes. The proofs not given in the text are collected in the Appendix.

[7] We retain the assumption that the agent is Bayesian. There is some evidence suggesting that individuals update their initially optimistic beliefs about their self-control in such a fashion Yaouanq and Schwardmann (2019).

Related Literature

This paper contributes to the large literature on information in moral-hazard models, focusing on firm-worker relationships. We offer a counterpoint to the classic results establishing that more precise information reduces agency costs (Holmström, 1979; Grossman and Hart, 1983; Kim, 1995) by providing a setting in which the principal prefers to base wages on a noisy information structure.

Coarse or noisy evaluations are also found by some other literatures. Several contributions show that obtaining the right measure of performance is difficult and using the wrong measure can backfire. This is the case with multitasking (Holmström and Milgrom, 1991) or when the agent has private information that would allow him to game a deterministic incentive scheme (Ederer et al., 2018). Hence, it can be optimal to leave information on the table and even introduce noise into the contract. Another strand of the literature shows that when evaluation is subjective, i.e. based on unverifiable private information of the principal, the resulting equilibrium remuneration will depend only on coarse information (MacLeod, 2003; Fuchs, 2007).We show that noisy evaluation is optimal even if verifiable information about the agent's true performance would be available.

That more, symmetric information about the technology can reduce profits in a moral hazard setting has been noted in the literature in several settings. It is well understood that ex-post incentive compatibility is more demanding than ex-ante incentive compatibility when implementing a fixed action. Lizzeri et al. (2002) show that interim performance evaluation is not optimal when there is no learning.[8] Nafziger (2009) demonstrates that it can be optimal to conceal information until after the agent's effort choice, even though this precludes the principal from adjusting the implemented action. Indeed, such situations are generic if the problem is sufficiently rich (Jehiel, 2015). In all these papers, the wage is still allowed to depend on the true realization of the signal, even if it is not revealed ex-ante. We

[8]In a tournament setting with exogenous and relative payments depending on cumulative output, the optimality of interim performance evaluation depends on the shape of the effort cost function (Ederer, 2010).

show that less information about the technology increases profits even if this implies that the wage cannot depend on the state even ex-post.[9]

To our knowledge, this is the first paper to combine the three key features of explicit incentives, learning about a persistent type, and information design. There are several literatures combining each two of these features:

A growing literature investigates the design of information structures in one-shot moral hazard problems with commitment to a wage scheme. The older literature (Dye, 1986; Feltham and Xie, 1994; Datar et al., 2001) considers the optimal acquisition and aggregation of information within a parametric class.[10] In Georgiadis and Szentes (forthcoming) and Li and Yang (forthcoming) the costs of information acquisition are assumed as part of the technology. Hoffmann et al. (2019) analyze a setting where the agent takes a single action, but information about his performance arrives over time. Information acquisition requires delayed payments, which creates endogenous costs because of impatience and imperfect risk sharing. We consider the design of the optimal information structure in a repeated setting with learning. The monitoring of output within every period affects not only incentives but also continuation beliefs, which leads to endogenous costs of information.

Learning about a persistent state and information design are combined in a growing literature. Most closely related to our moral-hazard setting are Smolin (2017) and Ely and Szydlowski (2019). In these papers, the worker updates his beliefs about the value of continuing the employment relationship compared to quitting based on information designed by the principal. The principal uses only information, which is valuable for the agent, as an incentive: she cannot commit to contingent payments. We analyze the role of information design when the principal also designs wages to provide incentives. Since the principal sets incentives and ability is match-specific, (symmetric) information itself is not

[9] Under this assumption, Fang and Moscarini (2005) show that information is detrimental if it erodes profitable overconfidence, see below.

[10] Indeed, when restricting attention to linear contracts, it can be optimal to leave information unused. (Feltham and Xie, 1994; Datar et al., 2001) This is a consequence of the restricted space of contracts, however.

valuable. Information and incentive design constrain each other, as the principal reveals at least as much information as is contained in wages.

Information and implicit incentives are also linked in models of career concerns (Holmström, 1999). A key difference is that in our paper, the principal commits to a wage payment based on the monitoring outcome, while in models of career concerns wages are determined by the belief of an observer ("the market"). As a consequence, the role of information is different. For career concerns, it is essential that skill and effort jointly affect the performance - the agent is motivated to exert effort because a decrease in output would be interpreted as low skill by the potential employers. Hörner and Lambert (forthcoming) analyze the optimal design of the information structure in a Gaussian career concerns model and show how it combines information from different sources or vintages to achieve the optimal combination of dependence on effort and on the agent's type.[11] In our setting with explicit incentives, entangled information is the source of the friction: the principal would prefer to provide incentives based on a signal that is informative about effort but independent of the agent's type in order to prevent learning.

The literature on learning in moral hazard models (Adrian and Westerfield, 2009; Giat et al., 2010; Prat and Jovanovic, 2014; Demarzo and Sannikov, 2017) studies learning based on output instead of an information structure that is designed endogenously by the principal. A key concern in this literature is belief manipulation – the fact that the agent's belief is private information after he deviated from the proposed effort. We analyze this issue as an extension and show that the basic shape of the optimal information structure is preserved.

Another important distinction is that we consider learning about the importance of effort, as opposed to learning about a state that affects only the level of output. The latter is often considerably more tractable and several of our assumptions are a result of

[11] They also show that is never optimal to introduce noise into the evaluation when implementing the highest effort. Dewatripont et al. (1999a) consider a one-shot career concerns problem when effort and the agent's type enter output in a general form and show that it can be optimal to add noise to the signal of his performance, as noise may increase the impact of effort on the realized signal.

handling this feature.[12] Notable exceptions are Bhaskar and Mailath (2019), who show that with learning about the importance of effort and spot contracts, the costs of implementing effort diverge as the number of periods grows, by establishing lower bounds on the cost of incentives.

Our extension to heterogeneous beliefs connects to the literature on contracting with overconfident agents, in particular de la Rosa (2011) who shows that overconfidence about the impact of effort relaxes the incentive constraint and is profitable for the principal. Fang and Moscarini (2005) show that if workers are sufficiently overconfident, the principal wants to conceal her private information about their true type by offering the same wage contract (which involves a fully informative evaluation of their output) to all workers. We derive how the principal shapes the performance evaluation to shape learning and preserve this misperception.

Technically, our paper relates to the literature on information design (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019), in particular the recent development of tools to handle information design problems with constraints (Boleslavsky and Kim, 2017; Le Treust and Tomala, 2019; Doval and Skreta, 2018) and additional choice variables (Georgiadis and Szentes, forthcoming) – in our case wages. We apply these tools in a setting where the information designer chooses a signal structure about one variable – output – in order to affect beliefs about another – the ability of the agent. This feature is particularly important in our extension to heterogeneous beliefs: Even though the prior of the principal and the agent are not mutually absolutely continuous, the information design problem is analyzed as if they were, using the transformation approach of Alonso and Câmara (2016).

[12] Importantly in our case, the information design problem solved by the principal cannot be written as the design of a distribution of posterior means unless the agent's type is binary.

1.2 The Model

A principal (she) employs an agent (he) for two periods. The principal is risk neutral, the agent is risk averse with utility index $u : [0,\infty) \to [0,\infty)$ which we assume to be unbounded.[13] Both share a common discount factor $\delta \in (0,1]$.

Technology Each period, the agent exerts nonverifiable effort $e_t \in \{0,1\}$ at cost ce_t, with $c > 0$. The worker has a time-invariant ability $\theta \in \{\theta_L, \theta_H\}$. For the main sections, we assume that the principal and the agent share a common prior belief μ that the agent has a high ability.[14]

The resulting output has either high or low quality, $y \in \{y_L, y_H\}$. We normalize the expected revenue from low output to zero and denote the expected revenue from high output by $Y > 0$. The probability of a high quality depends on the agent's effort and type, as follows:

type \ effort	$e_t = 0$	$e_t = 1$
$\theta = \theta_L$	a	$a+b$
$\theta = \theta_H$	$a+\Delta a$	$a+b+\Delta a+\Delta b$

Effort and ability are both productive, $b \geqslant 0$ and $\Delta a \geqslant 0$, and the technology is log-supermodular, $a\Delta b - b\Delta a > 0$. We assume that the principal wants to implement high effort in both periods and after all histories.[15]

Information, Contracts and Commitment We assume that the principal has full commitment within each period, but no commitment across periods. Within every period,

[13] This is for simplicity to avoid corner solutions.

[14] We relax this assumption in Section 1.4, where we investigate optimal evaluation and incentives when workers systematically misperceive the importance of their effort.

[15] It is easy to see that implementing high effort after all histories is optimal for the principal for a sufficiently high gain from high quality output, Y. This sharpens the trade-off between incentives and learning we aim to investigate, as the principal derives no instrumental value of information. Furthermore, implementing a given effort level is a standard focus in the contracting literature (e.g. Dittmann and Maug, 2007; Edmans and Gabaix, 2011).

timing is as follows: The principal proposes a contract, comprising a signal space S, a distribution over signals conditional on output, and a mapping from signals to wages.[16] Having observed the contract, the agent decides whether to quit and obtain outside utility U or to work, choosing effort level e_t. The outside utility is independent of the agent's type, which is assumed to be match specific, and satisfies $U > \frac{a}{b}c$.[17] At the end of the period, output, signals and wages realize.

Output is informative about the agent's type, but not directly observed by the agent or the principal.[18] The principal and the agent observe (noncontractible) effort, signals and wages and update their beliefs about the agent's type according to Bayes rule. Therefore, the evaluation designed by the principal has the dual role of providing the basis for incentive pay and determining the learning environment.

Discussion As mentioned in the introduction, it is important that the firm cannot engage in complete backloading of information while still providing incentives.[19] If the principal can record the output of the agent without revealing this information and credibly commit to contingent payments at the end of the relationship, a fully informative and fully delayed evaluation would be optimal. Our mechanism comes into effect when such complete backloading is costly or infeasible. In the main sections, the lack of intertemporal commitment ensures that incentives for effort have to be provided in the concurrent period. Wages are thus informative about output. We allow the principal to reveal more information through the evaluation, what is crucial, however, is that the agent observes at least as much information as contained in the wages. We discuss other options that preclude full

[16]This restriction to deterministic wages conditional on the signal is without loss, as the principal can simply extend the signal space to generate any desired randomness in the wage.

[17]The condition on the outside utility assures that the non-negativity constraint implicit in the utility function is never binding in the optimal contract.

[18]We can allow the firm to observe aggregate quantities. In a large organization, it is difficult to link a shortfall in aggregate outcomes to the individual worker, however. Formally, consider the model with a continuum of agents. Through is regular accounting activities, the firm observes aggregate outcomes such as profits, revenues or the average quality of output. These outcomes are not informative about the performance about an individual, infinitesimal agent.

[19]This is an assumption about the flow of information, the timing of the actual monetary transfer is not crucial.

"informational backloading" in Section 1.5.3, such as an impatient agent or a chance of information leaks.

A model of learning based on an endogenous signal distribution with hidden actions has the potential to create subtle issues of endogenous private information, both for the principal and the agent. For tractability and to focus on the main trade-off between incentives and learning, our assumptions ensure that no endogenous private information arises.

In the benchmark model, the principal does not acquire private information about the type of the agent, since she does not privately observe the quality of output directly, but only through the evaluation structure.[20] We relax this assumption in Section 1.5.1 and allow the principal to acquire noncontractible private information about the agent's performance in addition. We show that is optimal for the principal not to acquire such private information. To use it, it would have to affect the contract and would thereby be revealed to the agent in the second period. Then, however, it could have been revealed as part of the evaluation in the first period, in a way that reduced agency costs. Even if the choice of private information by the principal is unobserved, this outcome is supported as the unique equilibrium in a natural class.

On the agent side, the benchmark model ensures that his posterior belief remains common knowledge after a deviation to lower effort, since effort is noncontractible but observed. If effort is unobserved, the agent acquires private information about his belief after a deviation and double deviations to low effort in both periods will be strictly profitable. In Section 1.5.2, we derive the resulting dynamic incentive compatibility constraint, analyze this extended model, and show that our results generalize to this case.

[20] We can accommodate the principal observing the firm's aggregate output. Suppose there is a continuum of workers as described. Then, the aggregate outcome is uninformative about an individual agent's effort and type.

1.3 Analysis

A contract is a tuple (S, p, w), where S is an arbitrary measurable signal space, $p(\cdot|y) \in \Delta(S)$[21] denotes the distribution of the signal conditional on high (resp. low) output and $w : S \to \mathbb{R}$ denotes the wage promised to the agent after each signal.

1.3.1 Transformation to Belief Space

Every signal $s \in S$ induces a posterior belief

$$\mu(s) = \mu \frac{p(s|y_L) + (a + b + \Delta a + \Delta b)\,[p(s|y_H) - p(s|y_L)]}{p(s|y_L) + (a + b + (\Delta a + \Delta b)\mu)\,[p(s|y_H) - p(s|y_L)]} \tag{1.3.1}$$

by Bayes rule. Note that (1.3.1) relies on the presumption that high effort was exerted and is therefore only valid if there is no deviation from the effort proposed in the contract. The posterior is increasing in the likelihood ratio of the signal, $\frac{p(s|y_H)}{p(s|y_L)}$, since the high type is more likely to produce high output, and is fully determined by this likelihood ratio. It is bounded between $\underline{\mu}$ and $\bar{\mu}$, where

$$\underline{\mu} = \mu \frac{1 - (a + b + \Delta a + \Delta b)}{1 - (a + b + (\Delta a + \Delta b)\mu)}; \quad \bar{\mu} = \mu \frac{a + b + \Delta a + \Delta b}{a + b + (\Delta a + \Delta b)\mu} \tag{1.3.2}$$

denote the posteriors associated to a signal that realizes only after a low output ($p(s|y_H) = 0$) and only after a high output ($p(s|y_L) = 0$), respectively.

The contract in period t affects profits in that period but also the distribution over posteriors, which determines the continuation value of the principal. In the last period, this value is of course zero. In the first period, it is given by the expectation over the value of the contracting problem in the terminal period as a function of posterior beliefs. Let P^e_μ denote the expected probability of high output under belief μ if the agent exerts effort e.

[21] Throughout, we will use integral notation for expected values and understand expressions of the form $\int f(x)\,dx$ in the sense of distributions where required; no absolute continuity with respect to Lebesgue measure is assumed. With slight abuse of notation, we write $f(x) = f_x \in \mathbb{R}$ for $f(x) = f_x \delta_x$, where δ_x denotes a unit mass at x.

The optimal contract solves

$$\Pi_t(\mu) = \max_{S,p,w} P_\mu^1 Y + \int_S \left(P_\mu^1 p(s|y_H) + \left(1 - P_\mu^1\right) p(s|y_L) \right) \left(\delta \Pi_{t+1}(\mu(s)) - w(s) \right) ds \tag{1.3.3}$$

$$\text{s.t.} \int_S \left(P_\mu^1 p(s|y_H) + \left(1 - P_\mu^1\right) p(s|y_L) \right) u(w(s))\, ds - c \geqslant U \tag{P}$$

$$\int_S \left(P_\mu^1 p(s|y_H) + \left(1 - P_\mu^1\right) p(s|y_L) \right) u(w(s))\, ds - c \geqslant$$

$$\int_S \left(P_\mu^0 p(s|y_H) + \left(1 - P_\mu^0\right) p(s|y_L) \right) u(w(s))\, ds \tag{IC}$$

$$\int_S p(s|y_H)\, ds = \int_S p(s|y_L)\, ds = 1; \quad p(s|y) \geqslant 0 \tag{S}$$

This is a standard moral hazard problem, but with two added features. First, the principal *chooses an evaluation structure* and the wage cannot be more informative about the agent's output than the evaluation structure it is based on. In particular, the principal can choose to condition the wage on partially informative signals of output instead of output directly. Second, there is a *belief-dependent continuation value* $\Pi_{t+1}(\mu(s))$.

Note that the posterior $\mu(s)$ only realizes if the agent exerted effort. After a deviation, by contrast, the agent makes a different inference. This does not impact the incentive constraint since the principal observes the deviation. The agent does not gain *private* information about his belief and the principal holds him to his reservation value in the second period regardless.[22]

Proposition 1.1. *The optimal contract contains no signals that induce the same belief but are mapped to different wages. The contracting problem can be written as a choice of a distribution over posteriors* m *with mean* μ *and support on* $[\underline{\mu}, \bar{\mu}]$*, and a mapping from posteriors to wages.*

[22] We solve the problem where this belief-manipulation effect is present in the IC as an extension in Section 1.5.2.

While rewriting the choice of a signal structure as a choice of a distribution over posteriors is standard in the literature on Bayesian persuasion, applying this transformation to our contracting problem requires two adaptions. First, note that the principal designs an information structure about *output*, but the beliefs are about the agent's *type*. Since both spaces are one-dimensional and high quality output is more likely if the agent has a high type, there exists a one-to-one mapping between the two. Second, after a deviation to low effort, the distribution of signals changes. We need to be able to express this change as a function of the posterior distribution. Again, because the mapping from beliefs over output to beliefs over ability is one-to-one, we can find such a transformation (Boleslavsky and Kim, 2017).

Let $\mathfrak{m}$ denote the distribution over posteriors and (with slight abuse of notation) w the mapping from posteriors to utilities associated to (S, p, w). It is easy to see that

$$\int_S \left(P_\mu^1 p(s|y_H) + \left(1 - P_\mu^1\right) p(s|y_L) \right) \left(\delta \Pi_{t+1}(\mu(s)) - w(s) \right) ds \tag{1.3.4}$$

$$= \int \mathfrak{m}(\widehat{\mu}) \left(\delta \Pi_{t+1}(\widehat{\mu}) - w(\widehat{\mu}) \right) d\widehat{\mu}, \tag{1.3.5}$$

and similarly for the participation constraint. To transform the incentive constraint, note first that the original form of the incentive constraint is equivalent to

$$\int_S \left(b + \Delta b \mu \right) \left(p(s|y_H) - p(s|y_L) \right) u(w(s)) \, ds \geqslant c \tag{1.3.6}$$

An increase in effort increases the probability of high output by $b + \Delta b \mu$. This increase affects utility by shifting mass towards signals that are more likely after high output, therefore incentive compatibility requires a sufficiently strong correlation between a signal's responsiveness to high output and the utility delivered after it. Transforming the contracting problem into belief space, it reads

$$\Pi_t(\mu) = \max_{m,w} P_\mu Y + \int m(\widehat{\mu})\,(\delta\Pi_{t+1}(\widehat{\mu}) - w(\widehat{\mu}))\,d\widehat{\mu} \tag{1.3.7}$$

$$\text{s.t.} \int u(w(\widehat{\mu}))m(\widehat{\mu})\,d\widehat{\mu} - c \geqslant U \tag{P}$$

$$\int \big(b + \Delta b\mu\big)\frac{\widehat{\mu}-\mu}{(\Delta a + \Delta b)\mu(1-\mu)}u(w(\widehat{\mu}))m(\widehat{\mu})\,d\widehat{\mu} \geqslant c \tag{IC}$$

$$\int \widehat{\mu}\,m(\widehat{\mu})\,d\widehat{\mu} = \mu; \quad \text{supp}(m) \subset [\underline{\mu}, \bar{\mu}] \tag{BP}$$

The incentive constraint now requires a sufficiently strong correlation between the *posterior* and utility. This is because signals that are more likely after a good outcome are also associated with a high posterior probability that the agent is the high type. This correlation is rescaled since, depending on the parameters of the problem, this dependence may be more or less strong.

1.3.2 Terminal Period

In the second period, the principal has no continuation value from the relationship. Absent any reason to manipulate the agent's learning, the only objective in designing the signal structure is to provide incentives cheaply and there is no reason to leave information about output unused. It is optimal to use the most informative signal structure (Grossman and Hart, 1983).

Proposition 1.2. *The optimal contract in the second period uses the fully informative evaluation structure.*

The profit in the terminal period induces a continuation value

$$\int \Pi_2(\widehat{\mu})m(\widehat{\mu})\,d\widehat{\mu} \tag{1.3.8}$$

for the principal in the first period, where $\mathfrak{m}$ is the distribution over posteriors induced by learning from the evaluation in the first-period. We now show that this learning is costly for the principal, since it always reduces her continuation value.

More information about the agent's ability has two effects. On the one hand, it allows the principal to adapt the contract to the agent's ability. The contract filters out the nuisance parameter "ability" more effectively and provides incentives for effort more precisely. As a consequence, the wage can be less risky and it is cheaper to provide incentives. This effect is stronger the larger the effect of ability on the probability of high output. On the other hand, the agent also has more information when he decides whether to shirk or exert effort. Consequently, the wage has to be more risky on average in order to satisfy the IC constraint and it is more expensive to provide incentives. In other words, it is easier to satisfy the incentive compatibility constraint in expectation ("ex-ante") rather than for a more informed agent ("interim"). This effect is stronger the larger the effect of ability on the impact of effort. It dominates and learning reduces profits whenever complementarities are sufficiently strong, i.e. when the technology is log-supermodular in effort and ability.[23]

Proposition 1.3. *The value of the second period contracting problem, Π_2, is strictly concave in beliefs.*

Equivalently, consider the expected continuation value induced by distributions over beliefs, $\mathfrak{m}, \mathfrak{m}' \in \Delta([0,1])$, where $\mathfrak{m}$ is Blackwell less informative than $\mathfrak{m}'$. The principal prefers the less informative distribution

$$\int \Pi_2(\widehat{\mu})\mathfrak{m}(\widehat{\mu})\,\mathrm{d}\widehat{\mu} \geqslant \int \Pi_2(\mu)\mathfrak{m}'(\widehat{\mu})\,\mathrm{d}\widehat{\mu}. \tag{1.3.9}$$

[23] This is merely a *sufficient* condition. It is not tight for any nondegenerate utility function. Furthermore, learning is also costly if substitutability is sufficiently strong. A sufficient condition is that the probability of low output is log-supermodular in effort and ability.

To see this effect of information on the costs of incentives more concretely, consider the IC constraint in the terminal period,

$$(b + \Delta b \mu)\,(u(w_H) - u(w_L)) = c. \tag{1.3.10}$$

High effort increases the probability of high output by $P^1_\mu - P^0_\mu = b + \mu\Delta b$. The principal pays a base wage w_L and adds a bonus $w_H - w_L$ if and only if output is high (Proposition 1.2). The utility bonus is inversely proportional to the expected impact of effort, $b + \mu\Delta b$, and thus a convex function of the agent's beliefs. Consequently, a greater dispersion of beliefs causes an increase in the expected bonus. The principal wants the agent to stay uninformed, because it is cheaper to pay a bonus that is large enough in expectation than the expected bonus required by an informed agent.

1.3.3 Initial Period

In the first period, our main trade-off is in effect. By Proposition 1.2, providing incentives for the agent is cheaper in this period if the evaluation structure is more informative, while by Proposition 1.3 the resulting learning is costly as it increases the expected cost of incentives in the next period.

How is this trade-off resolved in the optimal contract? We employ the tools of information design to characterize the optimal evaluation structure without imposing any exogenous restrictions. While such restrictions, e.g. to a binary evaluation structure, may seem natural in a setting with a binary state and binary output, we know from this literature that they can be with loss of generality. Indeed, since the contracting problem has two constraints – participation and incentive compatibility, results from constrained information design suggest that the optimal evaluation structure may involve up to four signals (Le Treust and Tomala, 2019; Doval and Skreta, 2018).

To analyze this joint information and contract design problem, we make some assumptions on the utility function.

Assumption 1.1. *Let $w = u^{-1}$ denote the wage function mapping a level of utility to the wage required to provide it. It satisfies*

1. *(No incentives at infinity)* $\frac{w(x)}{x} \to \infty$ *as* $x \to \infty$.

2. *(Bounded changes in curvature)*

$$\frac{3(b+\mu\Delta b)\Delta b}{c(a\Delta b - b\Delta a)} \geqslant \frac{w'''(u_L)}{w''(u_L)} \text{ and } \frac{w'''(u_H)}{w''(u_H)} \geqslant -\frac{3(b+\mu\Delta b)\Delta b}{c((1-a)\Delta b + b\Delta a)},$$

   where $u_L = U - \frac{a+\mu\Delta a}{b+\mu\Delta b}c$ *and* $u_H = U + \frac{1-a-\mu\Delta a}{b+\mu\Delta b}c$.

3. *(Decreasing curvature)* $w''' \leqslant 0$.

All three restrictions are sufficient conditions that will be used in the proof of the main theorem. The first condition ensures that an interior solution exists. The principal doesn't find it profitable to provide an arbitrarily high payment with vanishing probability in order to incentivize the agent. The second condition ensures that the shape of the continuation value Π_2 is determined unambiguously by the technology and not by changes in the curvature of the utility function. It rules out that the curvature of the utility function changes too quickly. The third condition ensures that the information design problem is governed by the shape of the continuation value. All three conditions are satisfied for CRRA utility ($u(x) = \frac{x^{1-\gamma}}{1-\gamma}$) for $\gamma \leqslant \frac{1}{2}$ if the outside utility is sufficiently high.[24] They are always satisfied for $u(x) = \sqrt{2x}$.

Theorem 1.1. *Suppose* u *satisfies Assumption 1.1. Then, the optimal evaluation structure in the first period is (essentially) unique. It is binary and tough with* $S = \{G, B\}$ *and*

$$p(G|y_H) = 1-\sigma, \quad p(B|y_H) = \sigma, \quad p(G|y_L) = 0 \quad p(B|y_L) = 1, \tag{1.3.11}$$

[24] To see this, note that $\frac{w'''(x)}{w''(x)} = \frac{2\gamma-1}{1-\gamma}\frac{1}{x}$ for CRRA utility.

for $\sigma \in [0,1)$.

First, the motive to control learning does not increase the complexity of the evaluation structure. While the most informative evaluation is binary, a noisy evaluation can take many forms. The Theorem establishes that the optimal evaluation remains binary. The joint design of wages and information is crucial for this result, it does not necessarily hold when the wage function is fixed exogenously.[25]

Second, the principal uses a noisy binary signal of output as the basis of evaluation. The noise is asymmetric, making the evaluation "tough": A good evaluation results only if output was high. Low output always results in a bad evaluation, and the bad signal realizes also after high quality output with probability σ. In order to reduce the informativeness of the signal, the principal does not engage in "grade inflation", but instead measures performance against an "unreasonably" high standard.

The reason for this result is the shape of the continuation value of the principal. While the principal is always information averse, *the degree of information aversion is decreasing in the agent's posterior* ($\Pi_2''' > 0$). The main objective of the firm is to avoid workers from getting very pessimistic about their ability. To see why, consider again the second period IC,

$$(b + \Delta b\mu)\,(u(w_H) - u(w_L)) = c. \tag{1.3.12}$$

As we discussed previously, the impact of effort, $b + \Delta b\mu$, and the required bonus are inversely proportional, which implies that the continuation value is concave. Furthermore, this effect of learning is stronger when the posterior is low. In this case, the agent is pessimistic about the impact of his effort and even a small change in his belief has a large relative effect and causes large changes to the bonus. This leverage effect determines the shape of the continuation value if the curvature of the utility function doesn't change too much, which is guaranteed by Assumption 1.2. Therefore, the principal's information aver-

[25] Georgiadis and Szentes (forthcoming) show that the optimal information structure about effort when there are exogenous costs of acquiring information is binary when wages and the monitoring structure are designed jointly.

sion is larger at low posteriors. In order to raise the low posterior, the optimal monitoring structure pools at the bottom. Since the low evaluation might have been the result of bad luck, it is less damning.

We can provide an interpretable condition for a strictly noisy optimal evaluation in a special case.

Proposition 1.4. *Let* $u(x) = \sqrt{2x}$. *Then evaluation structure is not fully informative (*$\sigma > 0$*) if and only if*

$$\frac{1}{2}\left(c\,\frac{\Delta a + \Delta b}{b + \Delta b\mu}\,\frac{\mu(1-\mu)(\bar{\mu}-\underline{\mu})}{(\bar{\mu}-\mu)(\mu-\underline{\mu})}\right)^2 < \delta\left(\Pi_2(\underline{\mu}) + \Pi_2'(\bar{\mu})(\bar{\mu}-\underline{\mu}) - \Pi_2(\bar{\mu})\right) \tag{1.3.13}$$

The LHS corresponds to the cost of noisier evaluation in period 1, while the RHS is the cost of learning through the continuation value of the principal. A noisy evaluation structure is optimal if and only if the latter cost dominates.

Proof of Theorem 1.1

The proof of Theorem 1.1 poses the challenge of jointly designing an information structure and a wage scheme. Given a wage scheme, the information design problem can be solved by concavification (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011) taking into account the P and IC constraints (Boleslavsky and Kim, 2017; Le Treust and Tomala, 2019). The constraints make the problem multidimensional so that, although conceptually tractable, concavification is analytically difficult. Conversely, given an information structure, the problem of finding wages is a standard moral hazard problem. This tractable problem provides the starting point for a duality-based approach to such a joint information and incentive design problem, as outlined in Georgiadis and Szentes (forthcoming).

Consider the Lagrangian $\mathcal{L}$ associated to the contracting problem (1.3.7), where we retain (BP) as a constraint, and λ_P, λ_{IC} denote the Lagrange multipliers associated to the

participation and incentive constraint, respectively,

$$\begin{aligned}\mathcal{L}(m, w; (\lambda_P, \lambda_{IC})) = \int \Big\{ & P_\mu^1 Y + \delta \Pi_2(\widehat{\mu}) - w(\widehat{\mu}) \\ & + \lambda_P \left(u(w(\widehat{\mu})) - c - U \right) \\ & + \lambda_{IC} \left(\frac{b + \Delta b \mu}{(\Delta a + \Delta b)\mu(1-\mu)} (\widehat{\mu} - \mu)\, u(w(\widehat{\mu})) - c \right) \Big\} \, d\widehat{\mu}.\end{aligned} \tag{1.3.14}$$

We will write $\lambda = (\lambda_P, \lambda_{IC})$ when convenient.

As the proof relies on duality arguments, let us provide a quick summary. The contracting problem is equivalent to

$$\sup_{w, m \text{ s.t.(BP)}} \inf_{\lambda \geqslant 0} \mathcal{L}(m, w; (\lambda_P, \lambda_{IC})). \tag{1.3.15}$$

To see this, note that

$$\inf_{\lambda \geqslant 0} \mathcal{L}(m, w; \lambda) = \begin{cases} \int m(\widehat{\mu}) \left(P_\mu^1 Y + \delta \Pi_2(\widehat{\mu}) - w(\widehat{\mu}) \right) d\widehat{\mu} & \text{if (P)\&(IC) are satisfied} \\ -\infty & \text{else} \end{cases}, \tag{1.3.16}$$

the infimum simply wraps the constraints into the objective function. It is always the case that $\inf\sup \mathcal{L} \geqslant \sup\inf \mathcal{L}$, where the supremum is taken over the choice variables and the infimum over the multipliers. If this condition holds with equality, i.e. if we can exchange sup and inf, we say that the optimization problem satisfies *strong duality.*

Wage Setting Fix a distribution m satisfying (BP) and consider the problem of finding optimal wages subject to the participation and incentive constraint. This is a standard moral hazard problem and the optimal the optimal wage schedule follows from pointwise optimization of the Lagrangian. Furthermore, the problem is well behaved, so we have the following Lemma.

Lemma 1.1. *The wage setting problem satisfies strong duality, i.e.*

$$\sup_{w}\inf_{\lambda\geqslant 0}\mathcal{L}(m,w;\lambda)=\inf_{\lambda\geqslant 0}\sup_{w}\mathcal{L}(m,w;\lambda). \tag{1.3.17}$$

Information Design: Duality Given Lagrange multipliers λ, the Lagrangian at the optimal wage schedule can be written as an expectation of a function of the posterior

$$\sup_{w}\mathcal{L}(m,w;\lambda)=\int \ell^*(\widehat{\mu};\lambda)m(\widehat{\mu})\,d\widehat{\mu} \tag{1.3.18}$$

Therefore, the information design problem is of standard form. The principal maximizes the expectation of a function of posteriors,

$$\sup_{m \text{ s.t.(BP)}}\int \ell^*(\widehat{\mu};\lambda)m(\widehat{\mu})\,d\widehat{\mu}, \tag{1.3.19}$$

and can therefore be solved via concavification of ℓ^*. But note that this problem takes λ *as given.* This requires another exchange of sup and inf, which can be justified as ℓ^* is continuous and the space of beliefs is compact.

Lemma 1.2. *The information design problem satisfies strong duality, i.e.*

$$\sup_{m \text{ s.t.(BP)}}\inf_{\lambda\geqslant 0}\int \ell^*(\widehat{\mu};\lambda)m(\widehat{\mu})\,d\widehat{\mu}=\inf_{\lambda\geqslant 0}\sup_{m \text{ s.t.(BP)}}\int \ell^*(\widehat{\mu};\lambda)m(\widehat{\mu})\,d\widehat{\mu}. \tag{1.3.20}$$

Information Design: Concavification It remains to solve (1.3.19). In order to determine the concavification of ℓ^*, we need to determine its shape as a function of $\widehat{\mu}$. Using an envelope argument, it is straightforward to show[26] that

$$\begin{aligned}\frac{\partial^2}{\partial\widehat{\mu}^2}\ell^*(\widehat{\mu};\lambda)=&\lambda_{IC}^2\left[\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}\right]^2\rho'(\lambda_P+\lambda_{IC}\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\widehat{\mu}-\mu))\\&+\delta\Pi_2''(\widehat{\mu})\end{aligned} \tag{1.3.21}$$

[26] In the main text, we suppress boundary conditions related to the non-negativity constraint on wages.

where $\rho(x) := u(u'^{-1}(\frac{1}{x}))$ denotes the function that translates multipliers and scores to utilities, a function commonly encountered in moral hazard problems. The first term corresponds to the cost of providing incentives in the first period. It is positive, indicating convexity: the principal prefers the most informative evaluation structure in order to reduce agency costs. The second term corresponds to the impact of beliefs on the continuation value. It is negative: the principal wants to keep the agent uninformed in order to reduce agency costs in the next period.

Furthermore, we have that

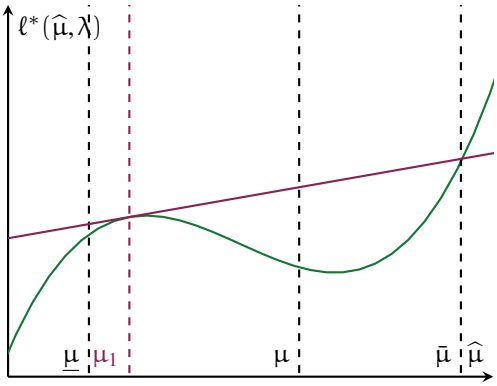
$$\begin{aligned}\frac{\partial^3}{\partial\widehat{\mu}^3}\ell^*(\widehat{\mu};\lambda) =&\lambda_{IC}^3\left[\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}\right]^3\rho''(\lambda_P+\lambda_{IC}\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\widehat{\mu}-\mu))\\ &+\delta\Pi_2'''(\widehat{\mu})>0\end{aligned} \tag{1.3.22}$$

The shape of ℓ^* has two components. The first term is the impact of the shape of the utility function. For given Lagrange multipliers, it is cheaper to provide incentives at higher posteriors as the curvature of w is decreasing (Assumption 1.3) and, equivalently, $\rho''>0$.[27] The second term is determined by the shape of the continuation value. The principal is less information averse for high posteriors.

There are three possible cases. If λ_{IC} is sufficiently small, the objective ℓ^* is strictly concave and the optimal information structure is uninformative. Clearly, this cannot be the case in the solution of (1.3.7), since the incentive constraint cannot be satisfied without any information. As λ_{IC} increases, we reach a region where the optimal information structure is partially revealing. As ℓ^* is convex for high posteriors in that case, it is fully informative at the top and uses partial pooling at the bottom (Fig. 1.3.1a). Finally, as λ_{IC} increases

[27]Note that this effect is purely "partial optimality": in the solution, the multipliers have to adjust to make sure that wage dispersion is sufficient to satisfy incentive compatibility. I conjecture that the restriction to utility functions with $\rho''>0$ is far from necessary: Instead, it is a result of the proof approach that requires solving the problem for all multipliers and relies on properties of the solution that are uniform across multipliers. $\frac{\partial^3}{\partial\widehat{\mu}^3}\ell^*(\widehat{\mu};\lambda)$ is positive *for all* multipliers only if $\rho''\geqslant 0$.

(a) Interior solution.



(b) Corner solution.

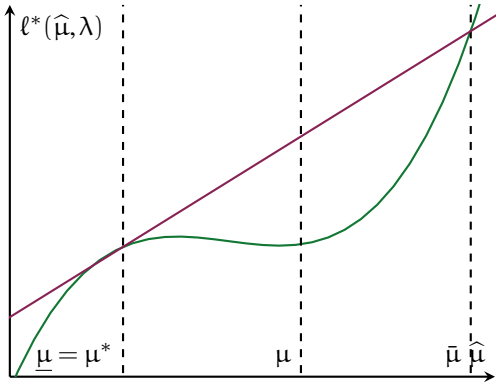


Figure 1.3.1: The concavification of ℓ^* at μ.

further, the costs of incentives overwhelm the gains from concealing information and the evaluation structure is fully informative (Fig. 1.3.1b).

Lemma 1.3. *For any λ, the optimal evaluation structure is unique and induces at most two posteriors. It induces the highest feasible posterior $\bar{\mu}$ with probability $m(\bar{\mu}) \in [0, \frac{\mu-\underline{\mu}}{\bar{\mu}-\underline{\mu}}]$ and a low posterior, $\mu^* \in [\underline{\mu}, \mu]$ with $m(\mu^*) \in [\frac{\bar{\mu}-\mu}{\bar{\mu}-\underline{\mu}}, 1]$.*

The Simplified Problem We can simplify the general problem (1.3.7) using the properties of optimal evaluation structures from the previous lemma, i.e. we can restrict attention to binary information structures where the good signal only realizes after high output. This simplified problem is

$$\max_{\mu^*, m^*, w_l, w_h} P^1_\mu Y + m^*\left[\delta\Pi_2(\mu^*) - w_l\right] + (1-m^*)\left[\delta\Pi_2(\bar{\mu}) - w_h\right] \tag{1.3.23}$$

$$\text{s.t. } m^* u(w_l) + (1-m^*)u(w_h) - c \geqslant U \tag{P^S}$$

$$\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}\left[m^*\big(\mu^*-\mu\big)u(w_l) + (1-m^*)\big(\bar{\mu}-\mu\big)u(w_h)\right] \geqslant c \tag{IC^S}$$

$$m^*\mu^* + (1-m^*)\bar{\mu} = \mu; \quad \mu^* \in [\underline{\mu}, \mu] \tag{BP^S}$$

where μ^* denotes the posterior after the bad evaluation, m^* denotes the probability of a bad evaluation and w_l, w_h denote the low and high wage, respectively.

Lemma 1.4. *The simplified contracting problem (1.3.23) has a unique solution. The optimal information structure is non-degenerate ($\mu^* < \mu$).*

The condition for an interior solution is derived in the Appendix. QED.

1.3.4 Analysis of the Solution and Comparative Statics

The properties of an interior solution are pinned down by tangency condition

$$\ell^*(\mu^*, \lambda(\mu^*)) + \frac{\partial \ell^*}{\partial \widehat{\mu}}|_{(\widehat{\mu},\lambda)=(\mu^*,\lambda(\mu^*))}(\bar{\mu} - \mu^*) = \ell(\bar{\mu}, \lambda(\mu^*)), \tag{1.3.24}$$

determining the posterior μ^* in the concavification of ℓ^* (Fig. 1.3.1). Note, however, that this condition does not correspond to the concavification of a given function. Instead, there is an additional dependence on $\lambda(\mu^*)$. This term is present because we are not solving an information design problem given payoffs, but design payoffs and information jointly, subject to a participation and and incentive compatibility constraint. For the graphical representation of our analysis this implies that, as we vary the tangent point in the figure to find the optimal μ^*, not only the tangent line but the whole function ℓ^* shifts.

Under the assumption that $u(x) = \sqrt{2x}$ we can transform (1.3.24) into a more interpretable form:

$$\frac{c^2}{2}\left(\frac{(\Delta a + \Delta b)\mu(1-\mu)}{b + \Delta b \mu}\frac{\bar{\mu} - \mu^*}{(\bar{\mu} - \mu)(\mu - \mu^*)}\right)^2 = \delta\left(\Pi_2(\mu^*) + \Pi_2'(\bar{\mu})(\bar{\mu} - \mu^*) - \Pi_2'(\bar{\mu})\right) \tag{1.3.25}$$

The LHS is the benefit from a more informative evaluation structure in period one. A more precise signal about output decreases agency costs today. This effect is larger if agency costs ($\frac{c}{b+\Delta b\mu}$) are already high and if a large dispersion of posteriors is required for a given level of information about output (since output is very informative, $\Delta b \mu(1-\mu)$ large). The

RHS is the cost of a more informative information structure through learning. A more precise signal today allows learning and thereby increases average agency costs in the next period. Indeed, the RHS is a measure of the concavity of the continuation value.

The optimal degree of shrouding, σ, is pinned down by the lower posterior belief μ^* according to

$$\sigma(\mu^*) = \frac{1-P_\mu}{P_\mu}\frac{\mu^\star - \underline{\mu}}{\bar{\mu}-\mu^*} \in [0,1], \tag{1.3.26}$$

which follows from inverting Bayes rule. It is increasing in μ^*; if the principal wants to cushion bad news, she needs to pool more on the bad signal.

Proposition 1.5. *Suppose* $u(x) = \sqrt{2x}$. *The optimal level of shrouding* σ *is*

(1) *weakly increasing in the discount factor* δ

(2) *weakly decreasing in the costs of effort in the first period,*
weakly increasing in the costs of effort in the second period, and
independent of a common increase in the cost of effort.

All comparisons are strict at interior σ.

Both comparative statics illustrate the trade-off between the cost of incentives in the first and second period. As the second period becomes more important, the evaluation structure becomes less informative. Higher costs of effort in the first period make economizing on agency costs in that period more important, thus the evaluation structure becomes more informative.

1.4 Preserving and Correcting Misperceptions

So far, we have argued that noisy and tough evaluation is the optimal way to preserve uncertainty about the agent's ability while providing incentives. We assumed that the principal and the agent agree about the situation, i.e. that they share a common prior.

There is some evidence suggesting, however, that beliefs concerning the impact of effort on outcomes – which are the driving factor of our results – may be systematically biased. Overestimation of one's abilities has been demonstrated in several laboratory contexts as well as in the workplace.[28] Overconfident workers overestimate their type and hence, given the complementarity between effort and ability, the importance of their contribution. Other biases can also affect beliefs about the impact of effort, for example the illusion of control, a tendency to overestimate the impact of individual choices on outcomes that also depend on chance, or the belief in a "just world".[29] Some individuals are also systematically underconfident and this trait is common in some groups.[30]

In this section, we analyze the optimal contract when the agent is not merely uncertain about his type, but enters the relationship with a systematic misperception. The principal now has an additional motive to shape learning, namely to affect the average posterior of the agent. An agent who overestimates his ability is more profitable because he is easier to incentivize. The principal would like to preserve this profitable misconception. Is this still achieved via tough evaluations or does she use a lenient information structure, akin to grade inflation, as the optimal way to preserve optimism?[31]

[28] See, for example Larwood and Whittaker (1977) for early evidence that individuals overestimate their abilities in a laboratory setting, (Burks et al., 2013) for a more recent incentivized study. Overconfidence is also present in tournaments (Park and Santos-Pinto, 2010) and among store managers (Huffman et al., 2019).

[29] Langer (1975) defines the illusion of control broadly as "an expectancy of a personal success probability inappropriately higher than the objective probability would warrant". The typical experiment establishes increased optimism about the outcome of a lottery in situations involving "choice, stimulus or response familiarity, passive or active involvement or competition". The fact that most experiments involve pure chance is intended as an extreme condition, suggesting that "the effects should be far greater when they are introduced into situations when there already is an element of control". But note Charness and Gneezy (2010); Filippin and Crosetto (2016), who find no evidence of illusion of control in two main experimental paradigms with monetary incentives.

According to just-world belief, effort and more generally good deeds are rewarded in the world. Such attitudes vary widely across countries and appear at best weakly related to true level of meritocracy. See Lerner (1980), and Bénabou and Tirole (2006) and the references therein for a discussion of the evidence.

[30] There is some evidence that women tend to be underconfident, for example (Niederle and Vesterlund, 2007; Hügelschäfer and Achtziger, 2014).

[31] Indeed, supporting students' self-esteem is often cited as a reason for grade inflation in schools and universities (Boretz, 2004).

1.4.1 Contracting with Heterogeneous Priors

We solve the contracting problem with heterogeneous priors. The agent again has a prior belief μ that he has high ability. The principal, by contrast, has a prior belief $\eta \in \{0,1\}$.[32] When $\eta = 0$, the principal is sure that the agent has low ability and we say that the agent is overconfident. When $\eta = 1$, by contrast, the principal is sure the agent has high ability and the agent is underconfident. The two players agree to disagree and update their priors using Bayes rule.[33]

We maintain our restrictions on the technology, namely that effort is productive ($b \geqslant 0$), the high type is more productive ($\Delta a \geqslant 0$), the technology is log-supermodular ($a\Delta b - b\Delta a > 0$) and that the outside option is sufficiently attractive to ensure an interior solution ($U > \frac{a+b}{b}c$). To focus on the effect of heterogeneous beliefs on the problem, we assume that $u(x) = \sqrt{2x}$.[34]

The Transformation to Belief Space

As before, we will transform the contracting problem and write it as the choice of a distribution of posterior beliefs and a wage function. However, the principal and the agent now have heterogeneous priors. In particular, the principal does not learn and therefore the relevant posteriors are those of the agent. In addition, the principal and the agent have different beliefs over the induced distribution of these posteriors. Let m denote the distribution according to the agent's belief and m_P this distribution according to the principal. The distribution over posteriors satisfies Bayes plausibility according to the

[32] Generally, it is reasonable to believe that the principal has better knowledge than the agent about his match-specific ability. The assumption that the principal is certain about the agent's match-specific ability is crucial for tractability in the case of heterogeneous priors. This is because the continuation value now depends both on the agents belief and the level of disagreement. With either identical priors or one degenerate prior, these two variables are simple. If these restrictions don't hold, the problem can still be rewritten in one dimension, but the information design problem is not tractable.

[33] While it would be interesting to analyze the design problem with non-Bayesian players, there is little work on information design tools for such a setting.

[34] This choice of utility function ensures that the curvature of the cost of wages, $w = u^{-1}$, (as a function of utility) is constant in the principal's problem.

agent,

$$\int \widehat{\mu}\, m(\widehat{\mu})\, d\widehat{\mu} = \mu \tag{1.4.1}$$

but, generically, not according to the principal. We can write the distribution over posteriors under the principal's prior belief, m_P, as a transformation of m, as follows. Let s be the signal inducing posterior $\widehat{\mu}(s)$.[35] Then, the probability of $\widehat{\mu}(s)$ according to the agent is

$$m(\widehat{\mu}(s)) = p(s|y_L) + (a + b + (\Delta a + \Delta b)\mu)\,[p(s|y_H) - p(s|y_L)] \tag{1.4.2}$$

According to the principal, this event has probability

$$\begin{aligned} m_P(\widehat{\mu}(s)) &= p(s|y_L) + (a + b + (\Delta a + \Delta b)\eta)\,[p(s|y_H) - p(s|y_L)] \\ &= m(\widehat{\mu}(s)) + (\eta - \mu)(\Delta a + \Delta b)\,[p(s|y_H) - p(s|y_L)] \\ &= \left[\eta \frac{\widehat{\mu}(s)}{\mu} + (1-\eta)\frac{1-\widehat{\mu}(s)}{1-\mu}\right] m(\widehat{\mu}(s)) \end{aligned} \tag{1.4.3}$$

Hence, we can follow the approach of Alonso and Câmara (2016) to Bayesian persuasion with heterogeneous priors and solve for the distribution m while the transformation factor $D^{\eta}(\mu, \widehat{\mu}) := \left[\eta \frac{\widehat{\mu}}{\mu} + (1-\eta)\frac{1-\widehat{\mu}}{1-\mu}\right]$ takes the heterogeneous priors into account.[36]

[35] This signal is unique without loss of generality by a straightforward extension of Proposition 1.1.

[36] Note that, in contrast to Alonso and Câmara (2016), the priors of the principal and the agent on the state space are *not* mutually absolutely continuous. The transformation method (as opposed to information design with surprises, Galperti, 2019) is still applicable since the posterior needs to be measurable with respect to a noisy signal of the state, namely output. This restriction keeps the belief transformation bounded. To be more precise, in our framework the principal designs an information structure about output, which implies a posterior about the type. Beliefs about the distribution of output are heterogeneous, but mutually absolutely continuous. There is a 1:1 mapping from beliefs about output to posteriors over the type.

The contracting problem with heterogeneous beliefs is thus

$$\Pi_t^\eta(\mu) = \max_{m,w} P_\eta^1 Y + \int \left(\delta \Pi_{t+1}^\eta(\widehat{\mu}) - w(\widehat{\mu})\right) D^\eta(\mu,\widehat{\mu}) m(\widehat{\mu})\, d\widehat{\mu} \tag{1.4.4}$$

$$\text{s.t.} \int u(w(\widehat{\mu})) m(\widehat{\mu})\, d\widehat{\mu} - c \geqslant U \tag{P}$$

$$\int \left(b + \Delta b \mu\right) \frac{\widehat{\mu} - \mu}{\Delta b \mu (1-\mu)} u(w(\widehat{\mu})) m(\widehat{\mu})\, d\widehat{\mu} \geqslant c \tag{IC}$$

$$\int \widehat{\mu}\, m(\widehat{\mu})\, d\widehat{\mu} = \mu; \quad \text{supp}(m) \subset [\underline{\mu}, \bar{\mu}] \tag{BP}$$

1.4.2 Terminal Period

Our results about the problem in the final period extend to the setting with heterogeneous priors. There is no reason to shape learning, so the principal prefers as much information as possible to incentivize effort as cheaply as possible. Therefore, the optimal evaluation structure in the final period is fully informative.[37]

In order to evaluate the impact of learning on the continuation value of the principal, we need to take into account the measure transform and consider

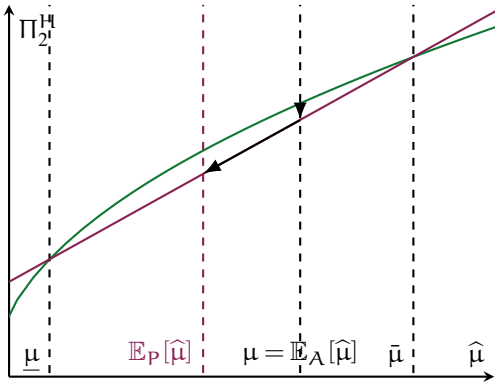
$$D^\eta(\mu,\widehat{\mu}) \Pi_2^\eta(\widehat{\mu}) \tag{1.4.5}$$

Learning creates a dispersion of the agent's posterior. This is costly for the principal, since $\frac{\partial^2}{\partial \widehat{\mu}^2} \Pi_2^\eta(\widehat{\mu}) < 0$ for the reasons discussed in the previous section. In addition, learning now also affects the expected posterior under the principal's belief. This *drift* has two effects. First, the disagreement between the principal and the agent decreases. This makes it harder to gamble on their belief difference and reduces profits. Second, the agent move towards the truth on average. Since gambling is limited due to risk aversion, this second effect dominates. If the agent is optimistic about the impact of effort ($\eta = 0$), this means he becomes less optimistic as he learns. Since optimism is profitable, therefore the principal

[37] In contrast to Proposition 1.2, this result does not follow readily from standard results about information in moral hazard problems, but requires an extension of the usual argument.

has an additional incentive to sabotage learning. If, instead, the agent is pessimistic ($\eta = 1$), he becomes less pessimistic on average, which is good for the principal.

(a) Optimism about the effect of effort.



(b) Pessimism about the effect of effort.

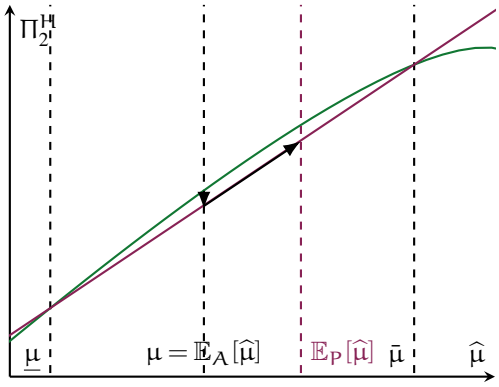


Figure 1.4.1: The effect of information on the principal's continuation value.

The total effect of learning combines the two forces of increased dispersion and drift. Therefore, learning reduces profits with optimism and has an ambiguous impact with pessimism. Let us summarize the preceding discussion.

Proposition 1.6. *Consider the contracting problem with heterogeneous beliefs. In the terminal period, the optimal evaluation structure is fully informative. The value of the second period contracting problem, Π_2^{η}, is strictly is increasing and concave in the agent's posterior.*

The impact of information is determined by

$$\frac{\partial^2}{\partial \widehat{\mu}^2}\left(D^{\eta}(\mu,\widehat{\mu})\Pi_2^{\eta}(\widehat{\mu})\right) = \underbrace{\left[\eta\frac{\widehat{\mu}}{\mu} + (1-\eta)\frac{1-\widehat{\mu}}{1-\mu}\right]\Pi_2^{\eta\prime\prime}(\widehat{\mu})}_{\textit{dispersion}} + \underbrace{2\frac{\eta-\mu}{(1-\mu)\mu}\Pi_2^{\eta\prime}(\widehat{\mu})}_{\textit{drift}} \tag{1.4.6}$$

It is negative if the agent is overconfident ($\eta = 0$). If the agent is underconfident ($\eta = 1$) the sign is ambiguous. In particular,

- *if $b > 0$, there exists a threshold $\bar{U}$ such that the principal is information loving if $U \geqslant \bar{U}$, and*

- *if* $a(1-a) > (4a-1)(\Delta a + \Delta b)$*, there exists a threshold* $\bar{b} > 0$ *such that the principal is information averse if* $b < \bar{b}$.

For high outside utility, the drift effect dominates and the principal prefers to reveal information in order to eliminate costly underconfidence. If the baseline effectiveness of effort is low (b is small), the impact of beliefs on the required bonus is highly levered, the dispersion effect dominates and the principal prefers to slow learning.

Initial Period

The shape of the optimal evaluation structure is determined by two factors: First, based on to continuation value, is the principal information averse and how does this information aversion change as a function of the posterior? This effect is similar to the common prior case, with the addition of the impact of the drift effect. Second, how do the costs of delivering utility change as a function of the posterior? This effect is not present with common priors and stems directly from the heterogeneity of beliefs.

Let us start with the familiar first effect. Taking into account the measure transform, the change in information aversion is determined by

$$\frac{\partial^3}{\partial \widehat{\mu}^3}\left(D^{\eta}(\mu,\widehat{\mu})\Pi_2^{\eta}(\widehat{\mu})\right) = \underbrace{\left[\eta\frac{\widehat{\mu}}{\mu} + (1-\eta)\frac{1-\widehat{\mu}}{1-\mu}\right]\Pi_2^{\eta'''}(\widehat{\mu})}_{\text{dispersion}} + \underbrace{3\frac{\eta-\mu}{(1-\mu)\mu}\Pi_2^{\eta''}(\widehat{\mu})}_{\text{drift}} \tag{1.4.7}$$

With an overconfident agent, both effects go in the same direction: A dispersion in beliefs has a higher leverage and is therefore more costly if the agent thinks effort is not very effective, i.e. at low posteriors. Similarly, the impact if the drift is stronger and therefore more costly at low posteriors, since a given decrease in the expected impact of effort has a higher leverage. Therefore, the information aversion of the principal decreases as she induces higher posteriors.

With an underconfident agent, there is a trade-off. A dispersion in beliefs again has a higher leverage and is therefore more costly if the agent thinks effort is not very effective,

i.e. at low posteriors. Therefore, inducing dispersion at low posteriors is more costly. Also the drift effect is also stronger with higher leverage at low posteriors, but since the drift effect is desirable with a pessimistic agent, this means that inducing drift at a low posterior is more profitable. Therefore, the total effect is ambiguous, the dispersion effect pushes towards increasing information aversion, while the drift effect pushes towards decreasing information aversion.

Let us turn to the direct effect of belief heterogeneity. It is cheaper for the principal to provide utility to the agent in states that the agent believes to be more likely than the principal. For an overconfident agent, that is a high posteriors, for an underconfident agent, that is at low posteriors.

Theorem 1.2. *The optimal evaluation structure in the first period is unique (up to renaming), binary and uses partial pooling. Let* $S = \{G, B\}$ *denote the signal space and* $\sigma \in [0,1)$ *the shrouding parameter*

- *If the agent is overconfident (*$\eta = 0$*), the optimal evaluation structure is (weakly) strict, i.e.*

$$p(G|y_H) = 1-\sigma, \quad p(B|y_H) = \sigma, \quad p(G|y_L) = 0 \quad p(B|y_L) = 1. \tag{1.4.8}$$

- *If the agent is underconfident (*$\eta = 1$*), the optimal evaluation structure is (weakly) lenient, i.e.*

$$p(G|y_H) = 1, \quad p(B|y_H) = 0, \quad p(G|y_L) = \sigma \quad p(B|y_L) = 1-\sigma. \tag{1.4.9}$$

Inducing high posteriors is always appealing for overconfident agents. All three effects work in the same direction. For underconfident agents, the drift effect and the direct effect of heterogeneous priors together are strong enough to jointly overpower the increased cost

of dispersion associated to providing information at low posteriors. To realize separation at the top and pooling at the bottom (in posterior space), the evaluation is lenient.

1.5 Discussion and Extensions

In the previous sections, we made several assumptions to ensure that the agent's posterior belief is the only state variable of the problem and that no party can acquire endogenous private information. Now, we relax those assumptions and discuss the impact on our results.

1.5.1 Private Information Acquisition

In some settings, it may be possible for the firm to privately observe additional information about the worker's output without disclosing it or using it as a basis of wages in the same period. We analyze this case and show that there exist natural equilibria that replicate the optimal contract. Furthermore, if the firm can commit not to acquire private information, it will.

Consider the model with symmetric priors. For simplicity of exposition, we assume that the principal uses a fully informative evaluation in the second period, as shown to be optimal in Proposition 1.2.[38] In the first period, the principal now also designs a *private* evaluation structure. Neither this information structure nor its realizations are observed by the agent, and we allow its distribution to depend on the realization of the public signal. Writing the problem in belief space, the principal designs a joint distribution of agent and principal posteriors $m_P(\mu_P, \widehat{\mu})$, with $\text{supp}(m_P) \subset [\underline{\mu}, \bar{\mu}]^2$. The marginal on the agent's posterior, $m(\widehat{\mu}) = \int m_p(\mu_P, \widehat{\mu})\, d\mu_P$, is observed by the agent. The distribution satisfies Bayes plausibility for both players, $\int \widehat{\mu}\, m(\widehat{\mu})\, d\widehat{\mu} = \mu$ and $\int \mu_P\, m_p(\mu_P, \widehat{\mu})\, d\mu_P = \widehat{\mu}$.

[38] This restriction is without loss on path, as a fully informative evaluation structure remains optimal for the principal. Off path, the restriction reduces the degrees of freedom for deviations, but the equilibrium can be extended naturally.

The two-period contracting problem now induces a dynamic game with incomplete information. A perfect Bayesian equilibrium consists of (1) an evaluation structure $\mathfrak{m}_P$ satisfying the above conditions, (2) a wage function $w : \widehat{\mu} \to w(\widehat{\mu}) \in \mathbb{R}_+$, (3) a first-period strategy of the worker mapping the evaluation and wage scheme to participation and effort choices, $\mathfrak{m}, w \to \{0,1\}^2$, (4) a second period contract offer, $(\mu_P, \widehat{\mu}) \to (w_L, w_H)(\mu_P, \widehat{\mu})$, (5) a belief system for the agent over his type and the information structure chosen by the principal, as a function of the posterior and the contract offer, $(\widehat{\mu}, w_L, w_H) \to \Delta\left([0,1] \times \Delta[0,1]^2\right)$, satisfying sequential rationality and consistency.

The outcome of Theorem 1.1 is achieved as the unique equilibrium in a natural class. A PBE is said to have *passive beliefs* if the second period belief of the agent is independent of the contract offer and equal to the posterior $\widehat{\mu}$ induced by the first period signal.[39]

Remark 1.1. The (essentially unique) equilibrium with passive beliefs is outcome-equivalent to the optimal contract characterized in Theorem 1.1. This equilibrium is principal preferred among all PBE of the game.

The intuition for this result is simple. For the principal facing an agent with passive belief $\widehat{\mu}$, the optimal contract in the second period satisfies both P and IC with equality. Therefore, the private information of the principal is of no use, and the continuation value induced on the first period is the same as in Propositions 1.2 and 1.3. Consequently, the principal's choice of $\mathfrak{m}_P$ is equivalent to the first-period problem.[40, 41] To see that this

[39] Orlov et al. (forthcoming) assume passive beliefs to show that the solution to their dynamic persuasion problem is robust to exogenous private information of the sender. Passive beliefs are also a common assumption in games with unobserved bilateral contracts, e.g. Hart and Tirole (1990); Brunnermeier and Oehmke (2013).

[40] Common refinements for signaling games, such as the intuitive criterion (Cho and Kreps, 1987) or D1 (Cho and Sobel, 1990), do not apply as they require the set of types of the principal to be fixed, which is not the case in our game. There are also no proper subgames to which they could be applied. Ekmekci and Kos (2019) analyze a signaling game when the sender chooses whether to acquire full information about his binary type or not, applying a form of never weak best response. Generalizing this kind of analysis to this extension is left for future research.

[41] If we nevertheless apply the reasoning of the intuitive criterion loosely to the contract offer game in the second period, it does not satisfy the requirement. This is because the principal's types with posteriors above those of the agent have a deviation that allows them to separate. This deviation, however, may not be the most intuitive psychologically. Compared to the pooling contract, the new contract features a lower bonus and delivers lower utility to the agent both under the original and under any plausible posterior belief.

equilibrium is principal preferred, note that in any PBE both the participation constraint and the incentive compatibility constraint need to be satisfied on the equilibrium path. The optimal contract is the best contract satisfying these restrictions. Any information used and thereby revealed in the second period could have been revealed in the first period, thus reducing agency costs.

The principal prefers to commit not to reveal the information in the second period. Passive beliefs provide such a form of commitment. Similarly, consider the game when the principal's choice of information structure m_P – both for private and public signals – is observed. Then, we have the following.

Remark 1.2. When the information structure is observed, any equilibrium[42] is outcome equivalent to the optimal contract characterized in Theorem 1.1.

1.5.2 Unobservable Effort

In the main sections, we assume that effort is observed but not contractible. This ensures that even after a deviation, the principal and the agent have common knowledge about their beliefs over the agent's type.

Assume instead that effort is not observed by the principal. This does not affect beliefs on equilibrium path, since the conjectured effort is correct. After a deviation to $e_t = 0$, however, the agent updates his beliefs according to

$$\tilde{\mu}(s) = \mu \frac{p(s|y_L) + (a + \Delta a)\,[p(s|y_H) - p(s|y_L)]}{p(s|y_L) + (a + \mu\Delta a)\,[p(s|y_H) - p(s|y_L)]} \tag{1.5.1}$$

while the principal continues to use the on-path updating rule (1.3.1). Hence, depending on the signal realization, the agent will be less (resp. more) optimistic about his type in the

One may conjecture that workers see such a contract offer less as a gesture of trust – as the intuitive criterion requires – but as a slight that demonstrate that the principal does not value their continued employment.

[42] Among PBE which satisfy the following natural restriction, a form of no-signaling-what-you-don't-know: After observing the information structure m_P and signal $\widehat{\mu}$, his belief is always supported on the convex hull of the support of $m_P(\cdot, \widehat{\mu})$.

second period and the contract offered by the principal will violate (resp. over-satisfy) the incentive compatibility constraint.[43] A deviation in the first period is more profitable for the agent because of this belief-manipulation effect.[44]

Let us now analyze this model, assuming that $\Delta a = 0$. This condition ensures that the agent does not learn about his type after a deviation and simplifies the problem considerably. Note that the problem in the second period is unchanged: The modification only affects continuation beliefs, which are irrelevant in the terminal period. In the first period, we need to modify the incentive-compatibility constraint in order to take the belief-manipulation effect into account.

Let $w_L(\widehat{\mu}(s))$ denote the optimal low wage in the second period problem with belief $\widehat{\mu}(s)$. The IC reads

$$\begin{aligned}\int_S \left(p(s|y_L) + (a + b + \mu\Delta b)\left[p(s|y_H) - p(s|y_L)\right]\right)\left[w(s) + U\right] ds - c \geqslant \\ \int_S \left(p(s|y_L) + a\left[p(s|y_H) - p(s|y_L)\right]\right) \cdot \\ \left[w(s) + \max\left\{w_L(\widehat{\mu}(s)) + P^1_\mu \frac{c}{b + \Delta b\widehat{\mu}(s)} - c, w_L(\widehat{\mu}(s)) + P^0_\mu \frac{c}{b + \Delta b\widehat{\mu}(s)}\right\}\right] ds\end{aligned} \tag{1.5.2}$$

This condition is now dynamic: If the agent does not deviate (first line), he will obtain his reservation utility U in the final period. If effort were observable, this would also be the case after a deviation, so this term would cancel. Since effort is not observable, he acquires private information about his type after a deviation and has a nontrivial choice in the second

[43] This assumes that the principal does not elicit the agent's belief at the beginning of the second period. For truth telling to be incentive compatible, it would need to be preferable to imitating the type that realizes on path, however. Hence, a screening mechanism in the second period cannot reduce the post-deviation payoff and therefore does not affect the optimal contract.

[44] This effect is central in the analysis of many models of moral hazard with learning, e.g. Prat and Jovanovic (2014); Demarzo and Sannikov (2017). Bhaskar and Mailath (2019) show that this motive implies that the costs of providing incentives using spot contracts grows unboundedly with the length of the time horizon in a model similar to ours, but with learning from output. It is doubtful whether the design of the information structure can reverse this conclusion and we conjecture that implementing high effort does not remain profitable for a long horizon with unobservable effort in our model.

period between exerting effort (the first term of the max) and shirking (the second term in the max). The former is optimal if he is more optimistic after the deviation ($\mu > \widehat{\mu}(s)$): The principal believes that the signal that realized is indicative of a low type and offers a correspondingly high bonus in the next period. The agent exerts effort and experiences a net gain. The latter is optimal if he is more pessimistic after the deviation ($\mu < \widehat{\mu}(s)$): The principal believes that the signal that realized is indicative of a high type and offers a correspondingly low bonus in the next period. The agent does not exert effort and thereby receives his reservation utility, avoiding the loss from the low bonus. Since the agent can reap the gain and avoid the loss, acquiring private information renders a deviation from high effort more profitable.

We can translate this dynamic IC into belief space and write

$$\int \left\{ \frac{(b+\mu\Delta b)}{\mu(1-\mu)\Delta b}(\widehat{\mu}-\mu)\,u(w(\widehat{\mu})) - \left[1 - \frac{(b+\mu\Delta b)}{\mu(1-\mu)\Delta b}(\widehat{\mu}-\mu)\right] \cdot \max\{0, c\Delta b \frac{\mu-\widehat{\mu}}{b+\widehat{\mu}\Delta b}\} \right\} m(\widehat{\mu})\,d\widehat{\mu} \geqslant c \tag{1.5.3}$$

Transformed in this fashion, the problem is amenable to an analysis along the lines of Theorem 1.1. The added complexity, however, is that kink in the incentive compatibility constraint introduces a kink in the Lagrangian of the problem.

Remark 1.3. The Lagrangian of the first period problem is concave-convex, with a concave kink at the prior, $\widehat{\mu} = \mu$. The optimal evaluation structure therefore consists of

1. a high signal that realizes only if output was good and results in the highest feasible posterior $\bar{\mu}$,

2. (possibly) a neutral signal that results in an unchanged posterior μ,

3. a low signal associated with posterior $\mu^* \in [\underline{\mu}, \mu)$.

Conditional on an informative realization from the evaluation, the signal structure is as before. The kink in the IC constraint, however, raises the possibility of a third, uninformative

signal. This is because the low signal realization is costly, as it invites belief-manipulation (1.5.3). In numerical simulations however, this possibility was never realized and we conjecture that the neutral signal is never part of the optimal contract.

1.5.3 Long-Run Commitment

In the main sections, we assumed that the principal does not have commitment across periods. This is not crucial for our results. What is crucial, however, is that the principal cannot backload all information.

To see this, suppose that the principal can commit to wages that depend on output in both periods and are revealed and paid at the end of the employment relationship. Then, informative wages do not lead to learning and hence using a fully informative evaluation is optimal. Indeed, if the principal can decide when to reveal wages, backloading information in this fashion is optimal. Common impediments against the backloading of payments other than period-by-period contracting can also break this result. Suppose for instance that the agent is less patient than the principal. In the extreme case of a myopic agent, only the current payments of the principal matter for payoffs and the period-by-period contracting problem is equivalent to the one discussed in the main sections. Noisy and (weakly) tough performance evaluation is again optimal.[45]

Our results continue to hold if the principal can postpone payments, but not information. Suppose that the first period contract specifies not only a wage this period, but also a continuation value.[46] Our results generalize to this model; the optimal evaluation structure can be noisy and is (weakly) tough.

[45] A continuity argument suggests that the result generalizes to interior impatience. A full analysis of this problem is left for future research.

[46] Note that specifying only a continuation value but not the exact way this value is delivered is potentially with loss in this setting. The outcome with commitment to a continuation value can be replicated by a two-period contract only if the principal can condition on the true posterior of the agent. This, however, would render effort contractible. For a model of moral hazard with long-run contracting, we also need to assume unobservable effort. Then, however, the continuation value is not a sufficient statistic for the influence of future payments on the current contract. In particular, the second period IC may be slack in the optimal contract after some realizations of the first-period signal.

Our results can be extended to the case where the principal can engage in partial informational backloading, as follows: The agent does not observe the wages he receives. Instead, there is a probability $\alpha \in [0,1]$ that he observes the outcome of his evaluation (e.g. the agent overhears the management talking about it). Importantly, the firm observes whether the agent did observe the evaluation or not. The utility specification and timing of contracting remains the same as in the main sections. The optimal contract in the first period is equivalent to the solution to our problem with discount factor $\alpha\delta$. A lower chance of discovery α leads to a more informative performance evaluation, lower agency costs.

1.6 Concluding Remarks

Our model demonstrates why it can be in a principal's interest to base incentives on a noisy evaluation of the agent's performance, even when the principal could measure true output and commit to contingent wages. The underlying insight is that output contains information both about effort, which she wants to ascertain and incentivize, and the agent's match-specific ability, which she would like to keep shrouded.

The optimal performance evaluation is *tough*: Good performance is not always recognized, but bad performance is always punished. Such tough evaluation ensures that even after a bad evaluation, the agent is not too pessimistic about his type. This is optimal because learning is especially costly at low posteriors, as a given change in beliefs has a large impact relative to the small expected efficiency of effort. One way a firm can commit to tough evaluation is through the selection and training of evaluators. Unreasonably strict supervisors and drill-sergeant mentality is part of the optimal organization design.

Our results inform not only the optimal evaluation of employee performance, but are also suggestive about the selection of information sources. Monitoring effort directly remains desirable. Among measures that combine information about effort and ability, the principal prefers measure that are less sensitive to ability. This is in sharp contrast with

models of implicit incentives. There, the fact that a signal combines effort and ability is the source of incentives, as the agent exerts effort to avoid being perceived as low-ability. The analysis of evaluation design when both explicit and implicit incentives are present – including the distinction between internal evaluation and externally visible evaluation – is an interesting avenue for future research.

References

Adrian, Tobias and Mark M. Westerfield, "Disagreement and Learning in a Dynamic Contracting Model," *Review of Financial Studies*, October 2009, *22* (10), 3873–3906.

Alonso, Ricardo and Odilon Câmara, "Bayesian Persuasion with Heterogeneous Priors," *Journal of Economic Theory*, September 2016, *165*, 672–706.

Aumann, Robert J. and Michael Maschler, *Repeated Games with Incomplete Information*, Cambridge, Mass: MIT Press, 1995.

_ and _ , "Belief in a Just World and Redistributive Politics," *The Quarterly Journal of Economics*, May 2006, *121* (2), 699–746.

_ and _ , "Information Design: A Unified Perspective," *Journal of Economic Literature*, March 2019, *57* (1), 44–95.

Bhaskar, V. and George J. Mailath, "The Curse of Long Horizons," *Journal of Mathematical Economics*, May 2019, *82*, 74–89.

Boleslavsky, Raphael and Kyungmin Kim, "Bayesian Persuasion and Moral Hazard," *SSRN Electronic Journal*, 2017.

Boretz, Elizabeth, "Grade Inflation and the Myth of Student Consumerism," *College Teaching*, 2004, *52* (2), 42–46.

Brunnermeier, Markus K. and Martin Oehmke, "The Maturity Rat Race," *The Journal of Finance*, 2013, *68* (2), 483–521.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini, "Overconfidence and Social Signalling," *The Review of Economic Studies*, 2013, *80* (3 (284)), 949–983.

Charness, Gary and Uri Gneezy, "Portfolio Choice and Risk Attitudes: An Experiment," *Economic Inquiry*, 2010, *48* (1), 133–146.

Cho, In-Koo and David M. Kreps, "Signaling Games and Stable Equilibria," *The Quarterly Journal of Economics*, May 1987, *102* (2), 179–221.

_ **and Joel Sobel**, "Strategic Stability and Uniqueness in Signaling Games," *Journal of Economic Theory*, April 1990, *50* (2), 381–413.

Datar, Srikant, Susan Cohen Kulp, and Richard A. Lambert, "Balancing Performance Measures," *Journal of Accounting Research*, June 2001, *39* (1), 75–92.

de la Rosa, Leonidas Enrique, "Overconfidence and Moral Hazard," *Games and Economic Behavior*, November 2011, *73* (2), 429–451.

Demarzo, Peter M. and Yuliy Sannikov, "Learning, Termination, and Payout Policy in Dynamic Incentive Contracts," *The Review of Economic Studies*, January 2017, *84* (1), 182–236.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole, "The Economics of Career Concerns, Part I: Comparing Information Structures," *The Review of Economic Studies*, January 1999, *66* (1), 183–198.

Dittmann, Ingolf and Ernst Maug, "Lower Salaries and No Options? On the Optimal Structure of Executive Pay," *The Journal of Finance*, 2007, *62* (1), 303–343.

Doval, Laura and Vasiliki Skreta, "Constrained Information Design: Toolkit," Technical Report 2018.

Dye, Ronald A., "Optimal Monitoring Policies in Agencies," *The RAND Journal of Economics*, 1986, *17* (3), 339–350.

Ederer, Florian, "Feedback and Motivation in Dynamic Tournaments," *Journal of Economics & Management Strategy*, September 2010, *19* (3), 733–769.

_ , Richard Holden, and Margaret Meyer, "Gaming and Strategic Opacity in Incentive Provision," *The RAND Journal of Economics*, December 2018, *49* (4), 819–854.

Edmans, Alex and Xavier Gabaix, "The Effect of Risk on the CEO Market," *The Review of Financial Studies*, August 2011, *24* (8), 2822–2863.

Ekmekci, Mehmet and Nenad Kos, "Signaling Covertly Acquired Information," Technical Report 2019.

Ely, Jeffrey and Martin Szydlowski, "Moving the Goalposts," *Journal of Political Economy*, May 2019, p. 704387.

Fang, Hanming and Giuseppe Moscarini, "Morale Hazard," *Journal of Monetary Economics*, May 2005, *52* (4), 749–777.

Feltham, Gerald A. and Jim Xie, "Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations," *The Accounting Review*, 1994, *69* (3), 429–453.

Filippin, Antonio and Paolo Crosetto, "Click'n'Roll: No Evidence of Illusion of Control," *De Economist*, September 2016, *164* (3), 281–295.

Fuchs, William, "Contracting with Repeated Moral Hazard and Private Evaluations," *American Economic Review*, September 2007, *97* (4), 1432–1448.

Galperti, Simone, "Persuasion: The Art of Changing Worldviews," *American Economic Review*, March 2019, *109* (3), 996–1031.

Georgiadis, George and Balazs Szentes, "Optimal Monitoring Design," *Econometrica*, forthcoming, p. 55.

_ , Steve T. Hackman, and Ajay Subramanian, "Investment under Uncertainty, Heterogeneous Beliefs, and Agency Conflicts," *Review of Financial Studies*, April 2010, *23* (4), 1360–1404.

Gittleman, Maury and Brooks Pierce, "How Prevalent Is Performance-Related Pay in the United States? Current Incidence and Recent Trends," *National Institute Economic Review*, November 2013, *226* (1), R4–R16.

Gonzalez-Hernandez, G., A. Sarker, K. O'Connor, and G. Savova, "Capturing the Patient's Perspective: A Review of Advances in Natural Language Processing of Health-Related Text," *Yearbook of Medical Informatics*, August 2017, *26* (1), 214–227.

Grossman, Sanford J. and Oliver D. Hart, "An Analysis of the Principal-Agent Problem," *Econometrica*, January 1983, *51* (1), 7.

Hart, Oliver and Jean Tirole, "Vertical Integration and Market Foreclosure," *Brookings Papers on Economic Activity. Microeconomics*, 1990, *1990*, 205.

Hoffmann, Florian, Roman Inderst, and Marcus M. Opp, "Only Time Will Tell: A Theory of Deferred Compensation," Technical Report 2019.

Holmström, Bengt, "Moral Hazard and Observability," *The Bell Journal of Economics*, April 1979, *10* (1), 74–91.

_ , "Managerial Incentive Problems: A Dynamic Perspective," *The Review of Economic Studies*, January 1999, *66* (1), 169–182.

_ **and** _ , "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, January 1991, *7*, 24–52.

_ **and Nicolas S. Lambert**, "Motivational Ratings," *Review of Economic Studies*, forthcoming.

Huffman, D., Collin Raymond, and Julia Shvets, "Persistent Overconfidence and Biased Memory: Evidence from Managers," Technical Report 2019.

Hügelschäfer, Sabine and Anja Achtziger, "On Confident Men and Rational Women: It's All on Your Mind(Set)," *Journal of Economic Psychology*, April 2014, *41*, 31–44.

Jehiel, Philippe, "On Transparency in Organizations," *The Review of Economic Studies*, April 2015, *82* (2), 736–761.

_ **and Matthew Gentzkow**, "Bayesian Persuasion," *American Economic Review*, October 2011, *101* (6), 2590–2615.

Kim, Son Ku, "Efficiency of an Information System in an Agency Model," *Econometrica*, 1995, *63* (1), 89–102.

Langer, Ellen J., "The Illusion of Control," *Journal of Personality and Social Psychology*, 1975, *32* (2), 311–328.

Larwood, Laurie and William Whittaker, "Managerial Myopia: Self-Serving Biases in Organizational Planning," *Journal of Applied Psychology*, April 1977, *62* (2), 194–198.

Lerner, Melvin J, *The Belief in a Just World: A Fundamental Delusion*, Boston, MA: Springer US : Imprint : Springer, 1980.

Li, Anqi and Ming Yang, "Optimal Incentive Contract with Endogenous Monitoring Technology," *Theoretical Economics*, forthcoming.

Lizzeri, Alessandro, Margaret A. Meyer, and Nicola Persico, "The Incentive Effects of Interim Performance Evaluations," Technical Report, Penn Economics Department September 2002.

MacLeod, W. Bentley, "Optimal Contracting with Subjective Evaluation," *The American Economic Review*, 2003, *93* (1), 216–240.

Nafziger, Julia, "Timing of Information in Agency Problems with Hidden Actions," *Journal of Mathematical Economics*, December 2009, *45* (11), 751–766.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away From Competition? Do Men Compete Too Much?," *The Quarterly Journal of Economics*, August 2007, *122* (3), 1067–1101.

Orlov, Dmitry, Andrzej Skrzypacz, and Pavel Zryumov, "Persuading the Principal To Wait," *Journal of Political Economy*, forthcoming.

Park, Young Joon and Luís Santos-Pinto, "Overconfidence in Tournaments: Evidence from the Field," *Theory and Decision*, July 2010, *69* (1), 143–166.

Prat, Julien and Boyan Jovanovic, "Dynamic Contracts When the Agent's Quality Is Unknown: Dynamic Contracts," *Theoretical Economics*, September 2014, *9* (3), 865–914.

Singer, Natasha, "In a Mood? Call Center Agents Can Tell," *The New York Times*, October 2013.

Smolin, Alex, "Dynamic Evaluation Design," Technical Report 2017.

Treust, Maël Le and Tristan Tomala, "Persuasion with Limited Communication Capacity," *Journal of Economic Theory*, November 2019, *184*, 104940.

Yaouanq, Yves Le and Peter Schwardmann, "Learning about One's Self," Technical Report 2019.

Appendix

1.A Proofs

Proof of Proposition 1.1: To see that the mapping from posteriors to wages is 1:1 and deterministic, let $m(s) := P_\mu^1 p(s|y_H) + (1-P_\mu^1)p(s|y_L)$ denote the probability of the signal under high effort. It is easy to see that the contracting problem (1.3.3) is equivalent to the utility space problem

$$\max_{S,p_H,p_L,v} P_\mu Y + \int_S \Big(\delta\Pi_{t+1}(\widehat{\mu}(s)) - w(v(s))\Big) m(s)\,ds \tag{1.A.1}$$

$$\text{s.t.} \int_S v(s)m(s)\,ds - c \geqslant U$$

$$\int_S \Big(b + \Delta b\mu\Big) \frac{\mu(s)-\mu}{\mu(1-\mu)(\Delta a + \Delta b)} v(s)m(s)\,ds \geqslant c \tag{IC}$$

$$\int_S p(s|y_H)\,ds = \int_S p(s|y_L)\,ds = 1 \tag{S}$$

where we used the representation of the IC in (1.3.6) and the fact that

$$\mu(s) - \mu = \mu(1-\mu)(\Delta a + \Delta b)(p(s|y_H) - p(s|y_L)). \tag{1.A.2}$$

Suppose there are two signals s, s' with $\mu(s) = \mu(s')$ and different utilities $v(s) \neq v(s')$. We could then set $\tilde{v} = \frac{m(s)}{m(s)+m(s')}v(s) + \frac{m(s')}{m(s)+m(s')}v(s')$ after both signals. This modification leaves all constraints unchanged, but reduces the costs of incentives since w is strictly convex.

Therefore, the payoff of any contract is pinned down uniquely by its induced distribution over posterior beliefs and mapping from posteriors to utilities, where optimality allows us to restrict attention to deterministic mappings by the above.

To see the bounds on posteriors, consider

$$\mu(s) = \mu \frac{1 + (a + \Delta a + b + \Delta b)(\frac{p(s|y_H)}{p(s|y_L)} - 1)}{1 + (a + \Delta a\mu + b + \Delta b\mu)(\frac{p(s|y_H)}{p(s|y_L)} - 1)} \tag{1.A.3}$$

This expression is maximized for $p_L = 0$, which induces the upper bound, and minimized for $p_H = 0$, which induces the lower bound. □

Proof of Proposition 1.2: Note that full information is strictly Blackwell more informative than any other information structure. Then, the result follows from Proposition 13 in Grossman and Hart (1983). Since both the Blackwell comparison as well as the concavity of the utility function are strict, uniqueness follows from an immediate generalization of their proof. □

Proof of Proposition 1.3: By standard arguments, both the participation and the incentive constraint are binding. Hence

$$\Pi_2(\mu) = P_\mu Y - P_\mu w(U - c + (1 - P_\mu)\frac{c}{b + \Delta b\mu}) - (1 - P_\mu)w(U - c - P_\mu \frac{c}{b + \Delta b\mu}). \tag{1.A.4}$$

Note that we require $U - P_\mu \frac{c}{b+\Delta b\mu} > 0$ to satisfy the implicit nonnegativity constraint in the agent's utility function. Since $\frac{\partial}{\partial \mu} P_\mu \frac{c}{b+\Delta b\mu} \propto \Delta a b - \Delta b a < 0$, this is implied by $U > \frac{a+b}{b} c$.

It is easy to verify that

$$\begin{aligned}\Pi_2''(\mu) \propto & 2(b\Delta a - a\Delta b)(b\Delta a + \Delta b(1-a))(b+\Delta b\mu) \\ & \cdot \left[w'(U + (1-P_\mu)\frac{c}{b+\Delta b\mu}) - w'(U - P_\mu \frac{c}{b+\Delta b\mu}) \right] \\ & - cP_\mu (b\Delta a + \Delta b(1-a))^2 w''(U + (1-P_\mu)\frac{c}{b+\Delta b\mu}) \\ & - c(1-P_\mu)(b\Delta a - a\Delta b)^2 w''(U - P_\mu \frac{c}{b+\Delta b\mu}) \end{aligned} \tag{1.A.5}$$

The two latter terms are clearly negative, and so is the first, since $b\Delta a - a\Delta b < 0$. The statement about the Blackwell comparison is immediate. □

We will prove the Lemmas used in the text to establish Theorem 1.1 first.

Proof of Lemma 1.1: Note that the optimal wage in the dual problem is given by

$$w^*(\lambda, \widehat{\mu}) = \max\{0, u'^{-1}(\lambda_P + \frac{b+\Delta b\mu}{(\Delta a + \Delta b)\mu(1-\mu)}(\widehat{\mu} - \mu))\} \tag{1.A.6}$$

Let λ^* denote a pair of multipliers such that the constraints are binding or the respective Lagrange multiplier is zero and the associated wage function is feasible. Then,

$$\begin{aligned} \inf_\lambda \sup_w \mathcal{L}(w,\lambda) = \inf_\lambda \mathcal{L}(w^*(\lambda,\cdot),\lambda) \leqslant & \mathcal{L}(w^*(\lambda^*,\cdot),\lambda^*) \\ = P_\mu^1 Y - \int w^*(\lambda,\widehat{\mu}) m(\widehat{\mu})\, d\mu \leqslant & \sup_w \inf_\lambda \mathcal{L}(w,\lambda) \end{aligned} \tag{1.A.7}$$

whence the problem satisfies strong duality.

To find such a λ^* consider the dual problem. Note that by an envelope argument

$$\frac{\partial \mathcal{L}(w^*(\lambda,\cdot),\lambda)}{\partial \lambda_P} = u(w^*(\lambda,\widehat{\mu})) - U - c \tag{1.A.8}$$

and similarly for λ_{IC} and hence the Hessian of the objective is given by

$$\begin{pmatrix} \int f(\widehat{\mu})\,d\widehat{\mu} & \int f(\widehat{\mu})\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\widehat{\mu}-\mu)\,d\widehat{\mu} \\ \int f(\widehat{\mu})\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\widehat{\mu}-\mu)\,d\widehat{\mu} & \int f(\widehat{\mu})\left[\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\widehat{\mu}-\mu)\right]^2 d\widehat{\mu} \end{pmatrix} \tag{1.A.9}$$

where $f(\widehat{\mu}) := \rho'(\lambda_P + \lambda_{IC}\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\widehat{\mu}-\mu))m(\widehat{\mu})$ is a positive kernel and the range of integration is over $\widehat{\mu}$ such that $u'^{-1}(\lambda_P + \frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\widehat{\mu}-\mu) \geqslant 0$. Hence, the integral is an inner product (between functions that share support with m), the objective is weakly convex by Cauchy-Schwarz.

If m is nondegenerate, we can bound the space of multipliers in the dual problem without loss. To see this, note that the infimum is bounded above by $P^1_\mu Y$ and this bound is achieved by $\lambda = 0$. For the first bound, suppose the wage is constant. Then

$$P^1_\mu Y - w + \lambda_P\,(u(w) - U - c) - \lambda_{IC} c \leqslant \mathcal{L}(w^*(\lambda,\cdot),\lambda) \leqslant 0 \tag{1.A.10}$$

Choose w such that $u(w) - U - c \geqslant 2c$. Then, for a suitable A,

$$\lambda_P \leqslant A + \frac{1}{2}\lambda_{IC} \tag{1.A.11}$$

Similarly, construct a wage function such that $\int u(w(\widehat{\mu}))\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\widehat{\mu}-\mu)m(\widehat{\mu})\,d\widehat{\mu} - c \geqslant 2(U+c)$. Then

$$\begin{aligned} P^1_\mu Y - w + \lambda_P\left(\int u(w)m(\widehat{\mu})\,d\widehat{\mu} - U - c\right) \\ + \left(\lambda_{IC}\int u(w)\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\widehat{\mu}-\mu)m(\widehat{\mu})\,d\widehat{\mu} - c\right) \leqslant \\ P^1_\mu Y - w - \lambda_P\,(U+c) + 2(U+c)\lambda_{IC} \leqslant 0 \end{aligned} \tag{1.A.12}$$

and, for suitable B,

$$\lambda_{IC} \leqslant B + \frac{1}{2}\lambda_P. \tag{1.A.13}$$

Taking both inequalities together, it is easy to see that we can bound λ and the set of λ to be compact. Hence, the dual problem has a solution. By convexity, it satisfies the FOC or comparative slackness. As the FOC is equal to the respective constraints, the wage at the solution either satisfies the constraints with equality or the constraint is slack and the multiplier is zero. We have found the desired λ^* for the case of nondegenerate m. For degenerate m, the problem is infeasible and hence both the primal and dual value are $-\infty$. Therefore the wage problem satisfies strong duality. □

Proof of Lemma 1.2: Clearly, the space of posterior distributions satisfying (BP) is compact in the weak topology, and, as ℓ^* is continuous and bounded for any λ, the problem is continuous and linear in m. Continuity in λ is immediate. To see quasi-convexity, note that by an envelope argument

$$\frac{\partial \ell^*}{\partial \lambda_P} = \int \left(\rho(\lambda_P + \lambda_{IC} \frac{b + \Delta b \mu}{(\Delta a + \Delta b)\mu(1-\mu)} (\widehat{\mu} - \mu)) - U - c \right) m(\widehat{\mu}) \, d\widehat{\mu} \tag{1.A.14}$$

and similarly for λ_{IC} and hence the Hessian of the objective is given by (1.A.9), the objective is weakly convex by Cauchy-Schwarz. Therefore, the problem satisfies the conditions of Sion's Minimax Theorem and we have

$$\inf_{\lambda \geqslant 0} \sup_{w, m \text{ s.t. (BP)}} \mathcal{L}(m, w; (\lambda_P, \lambda_{IC})) = \sup_{m \text{ s.t. (BP)}} \inf_{\lambda \geqslant 0} \sup_{w} \mathcal{L}(m, w; (\lambda_P, \lambda_{IC})). \tag{1.A.15}$$

□

Proof of Lemma 1.3: From (1.3.21), it is easy to see that

$$\frac{\partial^3}{\partial \widehat{\mu}^3} \ell^*(\widehat{\mu}; \lambda) = \delta \Pi_2'''(\widehat{\mu}) + \lambda_{IC}^3 \left[\frac{b + \Delta b \mu}{(\Delta a + \Delta b)\mu(1-\mu)} \right]^3 \rho''(\lambda_P + \lambda_{IC} \frac{b + \Delta b \mu}{(\Delta a + \Delta b)} \frac{(\widehat{\mu} - \mu)}{\mu(1-\mu)}) \tag{1.A.16}$$

It is elementary but tedious to show that

$$\begin{aligned}\Pi_2'''(\mu) =& \frac{c}{(b+\mu\Delta b)^6}\Big[6\Delta b\,(b\Delta a+(1-a)\Delta b)\,(a\Delta b-b\Delta a)\,(b+\mu\Delta b)^2\left(w_H'-w_L'\right)\\ &+3c\,(a\Delta b-b\Delta a)^2\,(b+\mu\Delta b)\,(b\Delta a+\Delta b\,(2-2a-b-\mu(\Delta a+\Delta b)))\,w_L''\\ &+3c\,(b\Delta a+(1-a)\Delta b)^2\,(b+\mu\Delta b)\,(a\Delta b-b\Delta a+\Delta b\,(a+b+\mu(\Delta a+\Delta b)))\,w_H''\\ &-c^2\,(a\Delta b-b\Delta a)^3\,(1-a-b-\mu(\Delta a+\Delta b))\,w_L'''\\ &+c^2\,(b\Delta a+(1-a)\Delta b)^3\,(a+b+\mu(\Delta a+\Delta b))\,w_H'''\end{aligned} \tag{1.A.17}$$

where $u_L = U - \frac{a+\mu\Delta a}{b+\mu\Delta b}c$ and $u_H = U + \frac{1-a-\mu\Delta a}{b+\mu\Delta b}c$ and $w_I^{(n)}$ is the nth derivative of w at u_I. Under Assumption 1.2, we have $\Pi_2''' > 0$. Hence, since $\rho'' \geqslant 0$, we have $\frac{\partial^3}{\partial\widehat{\mu}^3}\ell^*(\widehat{\mu};\lambda) \geqslant 0$ for all λ.

Let $\text{cav}f = \max_{\psi,\psi'\in[\underline{\mu},\bar{\mu}],\alpha\in[0,1]\,\text{s.t.}\,\alpha\psi+(1-\alpha)\psi'=\mu}\{\alpha f(\psi)+(1-\alpha)f(\psi')\}$ denote the concavification of function f on the interval $[\underline{\mu},\bar{\mu}]$ and consider the set of beliefs that can be used to generate the concavification of ℓ^* at the prior belief μ,

$$\Psi(\lambda_P,\lambda_{IC}) := \{\psi\in[\underline{\mu},\bar{\mu}]\,|\,\exists\psi'\in[\underline{\mu},\bar{\mu}],\alpha\in[0,1]\,\text{s.t.}\,\alpha\psi+(1-\alpha)\psi'=\mu \text{ and} \tag{1.A.18}$$

$$\text{cav}\ell^*(\mu;\lambda_P,\lambda_{IC}) = \alpha\ell^*(\psi;\lambda_P,\lambda_{IC})+(1-\alpha)\ell^*(\psi';\lambda_P,\lambda_{IC})\} \tag{1.A.19}$$

We have to show that the set is at most cardinality two and has the described structure. First, consider the case when ℓ^* is globally concave. Then it is strictly concave at μ and, clearly, $\Psi(\lambda_P,\lambda_{IC}) = \{\mu\}$. If instead ℓ^* is globally convex, then $\Psi(\lambda_P,\lambda_{IC}) = \{\underline{\mu},\bar{\mu}\}$. In all other cases, there exists a ψ such that ℓ^* is strictly concave for $\widehat{\mu} < \psi$ and strictly convex for $\widehat{\mu} > \psi$. Then, the concavification of ℓ^* is equivalent to ℓ^* up to a threshold $\mu^* < \psi$ and linear, generated by $\mu^*,\bar{\mu}$ afterwards. Hence, either $\Psi(\lambda_P,\lambda_{IC}) = \{\mu\}$, or $\Psi(\lambda_P,\lambda_{IC}) = \{\mu^*,\bar{\mu}\}$. The remaining statements are immediate from Bayes plausibility, $M(\mu^*)\mu^* + M(\bar{\mu})\bar{\mu} = \mu$ and $M(\mu^*) + M(\bar{\mu}) = 1$. □

Proof of Lemma 1.4: To show nondegeneracy in the simplified problem, we need to show that the optimal distribution of posteriors is nondegenerate. Then, existence follows from the bounds on dual multipliers in Lemma 1.1. To this purpose, we show that there exists an $\bar{\epsilon}_1$ such that $\mu^* < \mu - \bar{\epsilon}_1$. This also establishes non-degeneracy of the optimal information structure.

Suppose not, let $\mu^* = \mu - \epsilon$ and we will show that the costs of providing incentives diverge as $\epsilon \to 0$. To see this, note that

$$m(\mu^*)\mu^* + m(\bar{\mu})\bar{\mu} = \mu \tag{1.A.20}$$

$$m(\bar{\mu}) = \frac{\mu - m(\mu^*)\mu^*}{\bar{\mu} - \mu} \tag{1.A.21}$$

$$= \frac{\mu - m(\mu^*)\mu - \epsilon}{\bar{\mu} - \mu} \tag{1.A.22}$$

$$= \epsilon \frac{m(\mu^*)}{\bar{\mu} - \mu} \leqslant \epsilon \frac{1}{\bar{\mu} - \mu} \tag{1.A.23}$$

In the IC constraint, we have

$$c \leqslant \frac{b + \Delta b \mu}{(\Delta a + \Delta b)\mu(1-\mu)} \left[m^* \big(\mu^* - \mu\big) u(w^*) + (1 - m^*) \big(\bar{\mu} - \mu\big) u(\bar{w}) \right] \tag{1.A.24}$$

$$\leqslant \frac{b + \Delta b \mu}{(\Delta a + \Delta b)\mu(1-\mu)} \left[\epsilon \frac{1}{\bar{\mu} - \mu} \big(\bar{\mu} - \mu\big) \big(u(\bar{w}) - u(\underline{w})\big) \right] \tag{1.A.25}$$

Hence, as $\epsilon \to 0$, we require $u(\bar{w}) \geqslant c_0 \epsilon^{-1}$, for a suitable constant c_0. But then, the objective is $\leqslant c_1 - \epsilon \cdot c_2 w\left(\epsilon^{-1}\right) \to -\infty$ for suitable constants, which is clearly not optimal.

Hence, the optimal distribution of posteriors is nondegenerate and a solution exists. Is is unique since the problem is concave with convex constraint sets. □

Proof of Theorem 1.1: The claims about the information structure follow immediately from the previous lemma. □

Proof of Proposition 1.4: We rewrite the simplified problem, noting that $m^* = \frac{\bar{\mu}-\mu}{\bar{\mu}-\mu^*}$ and maximizing out wages. First, note that the IC constraint reads

$$\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}\left[m^*(\mu^*-\mu)u(w^*)+(1-m^*)(\bar{\mu}-\mu)u(\bar{w})\right] = \tag{1.A.26}$$

$$\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}\left[\frac{\bar{\mu}-\mu}{\bar{\mu}-\mu^*}(\mu^*-\mu)u(w^*)+\frac{\mu-\mu^*}{\bar{\mu}-\mu^*}(\bar{\mu}-\mu)u(\bar{w})\right] = \tag{1.A.27}$$

$$\frac{b+\Delta b\mu}{\Delta b\mu(1-\mu)}\frac{(\bar{\mu}-\mu)(\mu-\mu^*)}{\bar{\mu}-\mu^*}\left[u(\bar{w})-u(w^*)\right] \geqslant c \tag{1.A.28}$$

Then $u(\bar{w}) = \lambda_P + \lambda_{IC}\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\bar{\mu}-\mu)$ and $u(w^*) = \lambda_P - \lambda_{IC}\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}(\mu-\mu^*)$. The multipliers are $\lambda_P = U + c$ and

$$\lambda_{IC} = \frac{c}{\left(\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}\right)^2(\bar{\mu}-\mu)(\mu-\mu^*)}$$

By an envelope argument, the first order condition for μ^* is (writing in utility space)

$$0 = \delta\left[\frac{\bar{\mu}-\mu}{(\bar{\mu}-\mu^*)^2}\left(\Pi_2(\mu^*)-\Pi_2(\bar{\mu})\right)+\frac{\bar{\mu}-\mu}{\bar{\mu}-\mu^*}\Pi_2'(\mu^*)\right]+$$
$$\frac{1}{2}\left[\frac{\bar{\mu}-\mu}{(\bar{\mu}-\mu^*)^2}\left(u^{*2}-\bar{u}^2\right)+\frac{\bar{\mu}-\mu}{\bar{\mu}-\mu^*}\lambda_{IC}\left(\frac{b+\Delta b\mu}{\Delta b\mu(1-\mu)}\right)u^*\right] = \tag{1.A.29}$$

$$\delta\left[\frac{\bar{\mu}-\mu}{(\bar{\mu}-\mu^*)^2}\left(\Pi_2(\mu^*)-\Pi_2(\bar{\mu})\right)+\frac{\bar{\mu}-\mu}{\bar{\mu}-\mu^*}\Pi_2'(\mu^*)\right]-\frac{1}{2}\lambda_{IC}^2\left(\frac{b+\Delta b\mu}{(\Delta a+\Delta b)\mu(1-\mu)}\right)^2(\bar{\mu}-\mu) \tag{1.A.30}$$

as is straightforward but tedious to show. Plugging in for the multiplier and multiplying through, we arrive at the condition

$$\delta\left[\Pi_2(\mu^*)-\Pi_2(\bar{\mu})+(\bar{\mu}-\mu^*)\Pi_2'(\mu^*)\right] = \frac{1}{2}\left(\frac{(\Delta a+\Delta b)\mu(1-\mu)}{b+\Delta b\mu}c\right)^2\frac{(\bar{\mu}-\mu^*)^2}{(\bar{\mu}-\mu)^2(\mu-\mu^*)^2} \tag{1.A.31}$$

Note that this condition holds for an interior solution. As $\mu^* \to \mu$, the RHS diverges while LHS stays bounded, so there will never be a corner solution at this limit. As $\mu^* \to \underline{\mu}$, LHS grows as $\Pi_2'' < 0$ and RHS shrinks, but both stay bounded. We therefore have a corner solution at $\mu^* = \underline{\mu}$ if (1.3.13) is violated. □

Proof of Proposition 1.6: To see the first statement, consider the problem in signal/utility space. Then, the cost of incentives is

$$\int ((a+b+\eta(\Delta a+\Delta b))p_H(s)+(1-a-b-\eta(\Delta a+\Delta b))p_L(s))\,w(u(s))\,ds \quad (1.A.32)$$

and the constraints depend on

$$\int ((a+b+\mu(\Delta a+\Delta b))p_H(s)+(1-a-b-\mu(\Delta a+\Delta b))p_L(s))\,u(s)\,ds \quad (1.A.33)$$

and similar for IC. For a given p_L, p_S, u, we will construct a cheaper fully informative contract. Consider providing $\int p_H(s)u(s)\,ds$ for certain after high output and $\int p_L(s)u(s)\,ds$ after low output. The constraints are unchanged, so this contract is feasible. It is also cheaper by the convexity of w, strictly so if p_L, p_H were not degenerate.

To see monotonicity and concavity, note that

$$\begin{aligned}\Pi_2^\eta(\mu) =&(a+b+\eta(\Delta a+\Delta b))Y-(a+b+\eta(\Delta a+\Delta b))\frac{1}{2}\left(U+(1-P_\mu)\frac{c}{b+\Delta b\mu}\right)^2\\ &-(1-a-b-+\eta(\Delta a+\Delta b))\frac{1}{2}\left(U-P_\mu\frac{c}{b+\Delta b\mu}\right)^2 \quad (1.A.34)\end{aligned}$$

$$\Pi_2^{0\prime}(\mu) = \frac{c}{(b+\Delta b\mu)^3}\left[ac\Delta b(1-a-b) + bc\Delta b\,(1-a-b-\mu(\Delta a+\Delta b))\right.$$

$$\left. -bc\Delta a(\Delta a+\Delta b)\mu + b(\Delta a+\Delta b)(b+\Delta b\mu)U\right] \tag{1.A.35}$$

$$> \frac{c}{(b+\Delta b\mu)^3}\left[c\Delta b\,(1-a-b-\mu(\Delta a+\Delta b))\,(a+b) + b(\Delta a+\Delta b)(b+\Delta b\mu)U\right] \tag{1.A.36}$$

$$> 0 \tag{1.A.37}$$

$$\Pi_2^{0\prime\prime}(\mu) = -\frac{c}{(b+\Delta b\mu)^4}\left[c\Delta b^2\,(3a(1-a-b) + b(3-3a-2b-2\Delta b\mu)\right.$$

$$\left. +2b\Delta b(\Delta a+\Delta b)(b+\Delta b\mu)U - cb(\Delta a^2+2\Delta a\Delta b)\,(2\Delta b\mu-b)\right] \tag{1.A.38}$$

$$< -\frac{c}{(b+\Delta b\mu)^4}\left[c\Delta b^2\,(3a(1-a-b) + b(3-3a-2b-2\Delta b\mu)\right.$$

$$\left. +2b\Delta b(\Delta a+\Delta b)(b+\Delta b\mu)U - cb(\Delta a^2+2\Delta a\Delta b)\,(2\Delta b\mu-b)\right] \tag{1.A.39}$$

$$< 0 \tag{1.A.40}$$

using the fact that either $(2\Delta b\mu - b)$ is negative or we can use log-supermodularity. The results for overconfidence also follow from straightforward but tedious computation.

To see the result on information, note that information corresponds to a MPS of m, which the principal evaluates as an integral of $\left[\eta\frac{\widehat{\mu}}{\mu} + (1-\eta)\frac{1-\widehat{\mu}}{1-\mu}\right]\Pi_2^{\eta}(\widehat{\mu})$.

<u>If $\eta = 0$:</u> Since $U > \frac{a+b}{b}c$, $\frac{\partial^2}{\partial\widehat{\mu}^2}\left(\left[\eta\frac{\widehat{\mu}}{\mu} + (1-\eta)\frac{1-\widehat{\mu}}{1-\mu}\right]\Pi_2^{\eta}(\widehat{\mu})\right) > 0$ follows from direct computation.

<u>If $\Delta b < 0$:</u> The sign is ambiguous,

$$\frac{\partial^2}{\partial\widehat{\mu}^2}\left(\left[\eta\frac{\widehat{\mu}}{\mu} + (1-\eta)\frac{1-\widehat{\mu}}{1-\mu}\right]\Pi_2^{\eta}(\widehat{\mu})\right) \propto c\left[b^2\Delta a\,(2b - \Delta b(6+4\mu) - \Delta a(6+3\mu))\right.$$

$$+(2b-\Delta b\mu)\,(\Delta b(1-a)(b+\Delta a+\Delta b)$$

$$\left. +\ 4b\Delta a(\Delta a+\Delta b) + \Delta ba(1-a-b-\Delta a-\Delta b))\right]$$

$$+2b(b+\Delta b)(\Delta a+\Delta b)(b+\Delta b\mu)(U-c) \tag{1.A.41}$$

In particular, the expression is increasing in U. Furthermore, let $U = \frac{a+b}{b}c - \delta$ in order to ensure that the constraint is satisfied as we change b. For $b = 0$, we get

$$\frac{\partial^2}{\partial \widehat{\mu}^2}\left(\left[\eta\frac{\widehat{\mu}}{\mu} + (1-\eta)\frac{1-\widehat{\mu}}{1-\mu}\right]\Pi_2^{\eta}(\widehat{\mu})\right) \propto -[a(1-a) + (\Delta a + \Delta b)(1-4a)] \tag{1.A.42}$$

and if this expression is negative, we get the cutoff by continuity. □

Proof of Theorem 1.2: As this proof closely follows the same template as the proof of Theorem 1.1, we will be brief. All functions relate to Section 1.4, we refrain from using decorators to mark this association.

(Optimal Wages) The pointwise optimal wage schedule in the Lagrangian associated with (1.4.4) is

$$w^*(\widehat{\mu},\lambda) = \frac{1}{2}\left(\frac{1}{\eta\frac{\widehat{\mu}}{\mu} + (1-\eta)\frac{1-\widehat{\mu}}{1-\mu}}\right)^2\left(\lambda_P + \lambda_{IC}\frac{b + \Delta b\mu}{(\Delta a + \Delta b)\mu(1-\mu)}(\widehat{\mu} - \mu)\right)^2 \tag{1.A.43}$$

(Info Design) The Lagrangian is additively separable and

$$\begin{aligned}\ell^*(\widehat{\mu};\lambda) =& P_0 Y + \left[\eta\frac{\widehat{\mu}}{\mu} + (1-\eta)\frac{1-\widehat{\mu}}{1-\mu}\right][\delta\Pi_2(\widehat{\mu}) - w^*(\widehat{\mu},\lambda)] + \lambda_P\left(u(w^*(\widehat{\mu},\lambda)) - c - U\right) \\ &+ \lambda_{IC}\left(\frac{b + \Delta b\mu}{(\Delta a + \Delta b)\mu(1-\mu)}(\widehat{\mu} - \mu)\,u(w^*(\widehat{\mu},\lambda)) - c\right)\end{aligned} \tag{1.A.44}$$

<u>If $\eta = 0$:</u> Then,

$$\frac{\partial^2}{\partial \widehat{\mu}^2}\ell^*(\widehat{\mu};\lambda) = \frac{1-\mu}{(1-\widehat{\mu})^3}\left(\lambda_P + \lambda_{IC}\frac{b + \Delta b\mu}{(\Delta a + \Delta b)\mu}\right)^2 + \delta\frac{(1-\widehat{\mu})\Pi_2^{0\prime\prime}(\widehat{\mu}) - 2\Pi_2^{0\prime}(\widehat{\mu})}{1-\mu} \tag{1.A.45}$$

$$\frac{\partial^3}{\partial \widehat{\mu}^3}\ell^*(\widehat{\mu};\lambda) = 3\frac{1-\mu}{(1-\widehat{\mu})^4}\left(\lambda_P + \lambda_{IC}\frac{b + \Delta b\mu}{(\Delta a + \Delta b)\mu}\right)^2 + \delta\frac{(1-\widehat{\mu})\Pi_2^{0\prime\prime\prime}(\widehat{\mu}) - 3\Pi_2^{0\prime\prime}(\widehat{\mu})}{1-\mu} \tag{1.A.46}$$

and $\frac{\partial^3}{\partial\widehat{\mu}^3}\ell^*(\widehat{\mu};\lambda) > 0$. Lemma 1.3 goes through. We can apply the proof of Lemmas 1.4 and 1.2 mutatis mutandis and arrive at the Theorem.

If $\eta = 1$: Then, we have

$$\frac{\partial^2}{\partial\widehat{\mu}^2}\ell^*(\widehat{\mu};\lambda) = \frac{\mu}{\widehat{\mu}^3}\left(\lambda_P - \lambda_{IC}\frac{b+\Delta b\mu}{(\Delta a+\Delta b)(1-\mu)}\right)^2 + \delta\frac{\widehat{\mu}\Pi_2^{1\prime\prime}(\widehat{\mu}) + 2\Pi_2^{1\prime}(\widehat{\mu})}{\mu} \tag{1.A.47}$$

$$\frac{\partial^3}{\partial\widehat{\mu}^3}\ell^*(\widehat{\mu};\lambda) = -3\frac{\mu}{\widehat{\mu}^4}\left(\lambda_P - \lambda_{IC}\frac{b+\Delta b\mu}{(\Delta a+\Delta b)(1-\mu)}\right)^2 + \delta\frac{\widehat{\mu}\Pi_2^{\eta\prime\prime\prime}(\widehat{\mu}) + 3\Pi_2^{1\prime\prime}(\widehat{\mu})}{\mu} \tag{1.A.48}$$

It is straightforward but tedious to show that $\frac{\partial^2}{\partial\widehat{\mu}^2}\ell^*(\widehat{\mu};\lambda) = 0 \implies \frac{\partial^3}{\partial\widehat{\mu}^3}\ell^*(\widehat{\mu};\lambda) < 0$ and therefore the Lagrangian is either convex or convex to concave (it cannot be globally concave by incentive compatibility). Hence, a lenient information structure is optimal and the lemmas generalize. □

Chapter 2

Screening with Frames[1]

2.1 Introduction

Ample evidence, casual empiricism and introspection suggest that framing effects are common in choice.[2] In particular, the way a product is presented and the setting of the sales interaction can have a strong impact on consumer valuations.[3] Concordantly, many firms go to great expenses to improve the presentation of their product in largely non-informative and payoff-irrelevant ways through packaging, in-store presentation, and the emotions invoked by the sales pitch.

[1]We are grateful for valuable discussions and suggestions to Roland Bénabou, Elliot Lipnowski, Stephen Morris, Pietro Ortoleva, Wolfgang Pesendorfer, Doron Ravid, and Leeat Yariv; and for helpful comments to seminar participants at Princeton University, the Econometric Society Summer School 2018 in Singapore, Young Economists Symposium 2018, ESWM 2018, the NOeG Winter Workshop 2018, and participants of the RSF Summer Institute 2018. Ostrizek gratefully acknowledges financial support from the Philip Bennett *79 Public Policy and Finance Fund.

[2]For example, decision makers overvalue the impact of certain product attributes if they vary strongly in the choice set (see Bordalo et al., 2013, and references therein) and tend to be risk averse in decisions framed as gains and risk seeking for losses (Tversky and Kahneman, 1981).

[3]Consumer decisions are affected by the framing of insurance coverage (Johnson et al., 1993), the description of a surcharge (Hardisty et al., 2010), whether discounts are presented in relative or absolute terms (DelVecchio et al., 2007), prices as totals or on a per-diem basis (Gourville, 1998), and by background music (Areni and Kim, 1993; North et al., 1997, 2003, 2016). Large effects of framing on consumer valuation are also found in incentivized lab experiments and across policy discontinuities (Bushong et al., 2010; Schmitz and Ziebarth, 2017).

Most of the literature focuses on framing in static decision situations. However, many economic interactions including sales unfold in several stages. For instance, when buying a car, a consumer is first exposed to a manufacturer's marketing material, contemplates his purchase decision at home, and is then affected by the way the product is presented by the dealer. Even the sales pitch itself unfolds sequentially. As a result, firms have the opportunity to frame the options offered to the consumers differently at different stages of the decision and to use such changes of framing strategically. What is the optimal structure of a sales interaction? In particular, is it always best to present a product in the most favorable light? In general, how can a principal leverage the power to affect agents' preferences throughout a sequential interaction?

We investigate these questions by adding framing and extensive forms to a classic screening problem. The interaction of framing and sequential mechanisms allows the principal to exploit dynamic inconsistency to reduce information rents, whether consumers are sophisticated or not. While a growing literature analyzes this possibility assuming that the pattern of taste changes is given by the consumer's preferences (e.g. temptation or β-δ), in our model this pattern is *endogenous*. The monopolist induces changing tastes by varying framing throughout the interaction. She designs not only the contracts, but also the structure of the sequential decision problem along with a frame at each stage.

Our main result characterizes the structure of an optimal extensive-form decision problem (EDP) in a setting with quasi-linear single crossing utility and any finite number of types and frames under regularity conditions. The optimum is achieved in *three stages* using *two frames* (Theorem 2.1). If the consumers are sophisticated, the optimal EDP has the following key features:

1. *Short Interaction.* All types make at most three choices, and some types make only one choice.

2. *Natural Structure: approach–"cool-off"–close.* First, the agent is presented with a range of choices under a "hard sell" condition (highest valuation frame) and ei-

ther buys now or expresses interest in one of the contracts, but is given time to consider. Then he is allowed to "cool-off" (second highest valuation frame) and decides whether to continue or take the outside option. Finally, again in the "hard sell" frame, he is presented with the contract he expressed interest in and a range of decoy contracts designed to throw off agents that misrepresented their type initially.

3. *Gains from Framing vs. Rent Extraction.* The principal can either "reveal" or "conceal" each type. Therefore, she faces a trade-off between maximizing the valuation by using only the highest frame (revealed types) and reducing information rents by using frames in a high-low-high pattern to induce dynamic inconsistency (concealed types).

We illustrate these features and the main construction for sophisticated consumers in the following example.

Example 2.1. There are two equally likely types θ^1 (low) and θ^2 (high) and two frames, low and high. Preferences of type θ^i in frame $f \in \{l,h\}$ are represented by $u_f^i(p,q) = \theta_f^i q - p$, where the marginal utility θ_f^i depends both on the type and the current frame (see Fig. 2.1.1a). The monopolist principal (she) produces a good of quality q at cost $\frac{1}{2}q^2$ and maximizes profits.

If the principal offers a menu, this is a standard screening problem with an additional choice of a frame. It is easy to see that it is optimal to pick the high frame h and offer contracts so that θ_h^1's participation constraint and θ_h^2's incentive compatibility constraint bind, which yields a profit of 20.[4]

The principal can do better. Consider the allocation that would arise if the principal could make types observable at the cost of always putting the low type in the low frame. Then, she could implement the efficient full-extraction contract for θ_l^1, $c^1 = (9,3)$, and

[4]In particular, the optimal contracts are $(p^1,q^1) = (8,2)$ and $(p^2,q^2) = (32,6)$. Note that with these functional forms, $q^i = \theta_f^i$ is efficient for frame f and the quality of type 2 is distorted downward compared to the efficient quality for *both* frames.

θ^2_h, $c^2 = (36,6)$, obtaining a profit of $\Pi_{\text{extensive}} = 22.5 > 20$. We show that this is indeed possible in an EDP by varying the frames: $h \to l \to h$.

(a) Payoff types.

	$f = l$	$f = h$
θ^1	3	4
θ^2	5	6

(b) The optimal extensive-form decision problem implementing c^1, c^2.

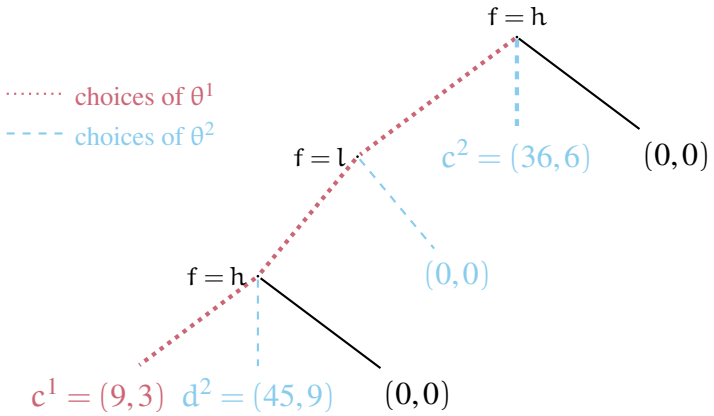


Figure 2.1.1: Example 2.1

To see how the principal achieves this, consider Fig. 2.1.1b. It is easy to check that the low type prefers c^1 to any other contract in the EDP in both frames and therefore proceeds through the tree to c^1. What about the high type? Because c^1 is preferable to c^2 for him in both frames, we need to show that such a deviation is *infeasible* in this extensive form. To deviate to c^1, at the root the high type needs to choose the continuation problem leading to this contract. As he is sophisticated, he correctly anticipates his future choices but cannot commit. That is, at the second stage he anticipates that at the final stage he would pick the decoy d^2 (in the high frame). But according to his taste at the second stage (in the low frame), the decoy is very unappealing, so he would choose the outside option. Hence, at the root the choice of the continuation problem is effectively equivalent to the outside option, thus, making the deviation to c^1 impossible.

By placing a decoy contract as a "tempting poison pill" in the extensive form, the principal effectively removes the incentive compatibility constraint. Hence, the high type doesn't obtain any information rent as the low type is concealed. This comes at the cost of adding an additional participation constraint, namely for the low type in the low frame, who

has to pass through the low frame on the path to his contract. Consequently, the maximal surplus that can be extracted from the low type is lower than in the static menu. There is a trade-off between concealing the contract intended for the low type in the continuation problem and thereby eliminating information rents and extracting surplus from this type.△

In general, the profit maximization problem is an optimization over the set of all extensive-form decision problems. However, based on the structure of the optimal EDP established in Theorem 2.1, we identify an equivalent simple optimization problem in contract space (Theorem 2.2). The principal partitions the set of types into revealed and concealed. This partition determines the participation and incentive constraints: Concealing a type eliminates incoming IC constraints at the cost of a tighter participation constraint. In contrast to the classic setting, it is never optimal to exclude any type, as it is strictly better to sell a strictly positive quality to every type and conceal some of them instead (Proposition 2.3).[5]

For the main sections, we assume that consumers are sophisticated. They correctly anticipate their choices, but cannot commit to a course of action.[6] As the optimal sales interaction has a simple 3-stage structure, correctly anticipating behavior in this extensive form is relatively easy. Moreover, consumers are exposed to sales pitches on a daily basis, they are experienced and understand the flow of the interaction. Sophistication reflects the idea that consumers understand that they are more prone to choose premium option when under pressure from the salesperson (high frame), and (in a low frame) avoid putting themselves in such situations that lead to excessive purchases. In addition, sophistication serves as a benchmark, by making it difficult for the principal to extract surplus. Even if consumers are fully strategically sophisticated and can opt out of the sales interaction at

[5]This is in line with Corollary 2 in Salant and Siegel (2018), which states that there is no exclusion with two types, when the principal offers a framed menu under a participation constraint in a neutral frame. A related result is in Eliaz and Spiegler (2006). They show that there is no exclusion when the principal screens by the degree of sophistication. We show that no-exclusion holds when the principal screens by payoff type.

[6]Sophistication is a common modeling choice in the domain of time preference following the seminal work of Strotz (1955); Laibson (1997).

any point, framing in extensive forms affects the sales interaction and its outcomes. Indeed, the principal turns consumers' sophistication against them.

We also consider naive consumers. They fail to anticipate that their tastes may change and choose a continuation problem as if their choice from this problem would be made according to their current tastes. For naive consumers, the principal can implement the efficient quantities in the highest frame and extract all surplus with a three-stage decision problem. She does so using decoy contracts in a bait-and-switch: Naive consumers expect to choose a decoy option tailored to them and reveal their type by choosing the continuation problem containing it at the root (bait), but end up signing a different contract due to the preference reversals induced by a change of frame (switch). When both naive and sophisticated consumers are present in arbitrary proportions and this cognitive type is not observable to the firm, our results generalize (Theorem 2.3).[7] The optimal extensive-form still has three stages and implements the same contracts as if the cognitive type were observable. There are no cross-subsidies from naive to sophisticated consumers.

Many jurisdictions mandate a right to return goods and cancel contracts, especially when the sale happened under pressure (e.g. door to door). This gives consumers the option to reconsider their purchase in a calm state of mind, unaffected by the immediate presence of the salesperson. We analyze such regulation and find that, while the principal can no longer use framing to exaggerate surplus, she can still use the resulting dynamic inconsistency to fully eliminate the information rent of all types. Sophisticated consumers do not require protection by a right to return if they can decide to avoid the seller, e.g. by not visiting the store, but naive consumers would benefit even in this case.

Beyond the setting of framing in screening problems, we view our results as steps towards understanding the impact of behavioral choice patterns (both framing and choice set dependence) when a principal (or mechanism designer) can offer extensive-form decision

[7] Spiegler (2011) notes that the principal can costlessly screen by cognitive type in a setting without taste heterogeneity.

problems in order to exploit the resulting violations of dynamic consistency and demand for commitment. We return to this discussion in the conclusion.

We set up the model in Section 2.2. In Section 2.3, we show that the optimal extensive-form decision problem is of a simple three-stage structure. We find a relaxed problem in price-quality space that characterizes the optimal vector of contracts. In Section 2.4, we construct the optimal extensive-form decision problem if some consumers are naive about the effect of framing. We also consider the case when the principal's choice of extensive form is restricted to account for a participation decision (e.g. a right to return the product) in an exogenous "neutral" frame. We conclude with discussions. Proofs are collected in the Appendix.

Related Literature

A growing literature studies the manipulation of framing by firms. Piccione and Spiegler (2012) and Spiegler (2014) focus on the impact of framing on the comparability of different products. Salant and Siegel (2018) study screening when framing affects the taste for quality, as in our setting. In this paper, the principal chooses a framed menu, while we study the optimal design of an extensive-form decision problem to exploit the dynamic inconsistency generated by choice with frames and make predictions about the structure of interactions. In addition, our model makes different predictions for the use of framing and efficiency in the setting where the two are closely comparable:[8] Using extensive forms, it is *always* more profitable to use framing (not only when it is sufficiently weak) and framing removes *all* distortions created by second-degree price discrimination (not only some) in our setting.

Our article is also related to behavioral contract theory more generally (for a recent survey, see Kőszegi 2014, for a textbook treatment, see Spiegler 2011), in particular to

[8]That is, comparing their Section 3 with our Section 2.4.2, where we impose a right to return the product in an exogenously given "neutral" frame. They also consider a model without returns but with a "basic" product that has to be offered and an insurance problem in which the monopolist can highlight one of the options, turning it into a reference point relative to which consumers experience regret.

screening problems with dynamically inconsistent agents (Eliaz and Spiegler, 2006, 2008; Esteban et al., 2007; Esteban and Miyagawa, 2006a,b; Zhang, 2012; Galperti, 2015; Heidhues and Kőszegi, 2010, 2017; Yu, 2018; Moser and Olea de Souza e Silva, 2019).[9] These papers consider situations when the taste changes are *given* by the preferences of the agents (e.g. Gul and Pesendorfer (2001) or β-δ) and consequently design a 2-stage decision problem as induced by the natural time structure of the problem. We study how a principal *chooses* the sequence of frames and an extensive form of arbitrary (finite) length to induce dynamic inconsistency and we show that a 3-stage mechanism is optimal.

Given the optimal sequence of frames, this mechanism employs techniques similar to those in the literature. In particular, it involves off-path options that remain unchosen by every type ("decoys"). In Esteban and Miyagawa (2006a) and Galperti (2015) such decoys make deviations less attractive and are thus analogous to the decoy contracts we introduce in the optimal extensive form for sophisticated agents. Heidhues and Kőszegi (2010) show that credit contracts for partially sophisticated quasi-hyperbolic discounters feature costly delay of the payment which the consumer fails to expect when signing the credit contract. Immediate repayment is hence an unused option analogous to the "bait" decoys we introduce to screen naive consumers.

Glazer and Rubinstein (2012, 2014) consider models where the principal designs a procedure such that misrepresenting their type is beyond the boundedly rational agents' capabilities. While their decision problems are based on hypothetical questions about the agent's type, we show that it is possible to structure a choice problem with framing to make it impossible to imitate certain types.

[9] Eliaz and Spiegler (2006, 2008) screen dynamically inconsistent agents by their degree of sophistication and optimistic agents by their degree of optimism, respectively. Esteban et al. (2007); Esteban and Miyagawa (2006a,b) study screening when agents are tempted to over- or underconsume. Zhang (2012) studies screening by sophistication when consumption is habit inducing. Galperti (2015) studies screening in the provision of commitment contracts to agents with private information on their degree of time inconsistency, Heidhues and Kőszegi (2017) study selling credit contracts in this setting. Yu (2018) and Moser and Olea de Souza e Silva (2019) study optimal taxation problem, where agents are also heterogeneous in the degree of present-bias.

There is a large literature on endogenous context effects, e.g. through focusing the attention of the decision maker on attributes that vary strongly or are exceptional within the choice set (Bordalo et al., 2013; Kőszegi and Szeidl, 2013). We consider the case of framing through features of the choice situation, such as the sales pitch or the presentation format. Thus, consumers in our model fit into the choice with frames framework of Salant and Rubinstein (2008).

The presence of different frames and extensive forms places our screening setting close to implementation. If we reinterpret our decision maker as a group of individuals with common knowledge of their type but different tastes, one individual corresponding to each frame, the principal applies implementation in backward induction (Herrero and Srivastava, 1992). While they give abstract conditions for implementability in a very general setting, we characterize the structure of the optimal decision problem for our screening model and derive properties of the optimal contracts.

2.2 Screening with Frames and Extensive Forms

We build on the classic model of price discrimination (Mussa and Rosen, 1978; Maskin and Riley, 1984), extending the framework in two ways. Instead of simple menus, firms design an extensive-form decision problem. Furthermore, for every decision node the firm picks a frame affecting the valuation of consumers. Our results are driven by the interaction of both ingredients.

2.2.1 Contracts and Frames

The firm produces a good with a one-dimensional characteristic $q \geqslant 0$, interpreted as quantity or quality. Throughout the exposition, we maintain the latter interpretation. A contract c is a pair of a price p and a quality q, the space of contracts is $\mathcal{C} = \mathbb{R} \times \mathbb{R}_+$.

There is a finite set of frames F with $|F| \geqslant 2$ and a finite type space Θ endowed with a full support prior μ. Each type is a function $\theta : F \to \mathbb{R}$ that maps frames into payoff types, denoted as $\theta_f \equiv \theta(f)$. For a given payoff type θ_f the consumer is maximizing the utility function

$$u_{\theta_f}(p, q) = v_{\theta_f}(q) - p. \tag{2.2.1}$$

where $v : \mathbb{R} \times \mathbb{R}_+$ is a thrice differentiable function, that satisfies

$$\frac{\partial v}{\partial q} > 0, \qquad \frac{\partial v}{\partial \theta} \geqslant 0, \qquad \frac{\partial^2 v}{\partial \theta \partial q} > 0, \qquad \frac{\partial^2 v}{\partial q^2} < 0, \qquad \frac{\partial^3 v}{\partial q^2 \partial \theta} > 0. \tag{2.2.2}$$

For convenience, we normalize $\forall \theta_f, v_{\theta_f}(0) = 0$. Note that we assumed that utility is quasi-linear in money and frames affect a consumer's taste for quality. This is consistent with framing effects on price perception, as long as these effects are multiplicatively separable.

For a given vector of contracts $\mathbf{c} = (c_\theta)_{\theta \in \Theta}$, we refer to the constraints

$$u_{\theta_f}(c_\theta) \geqslant 0, \text{ and} \tag{P^f_θ}$$

$$u_{\theta_f}(c_\theta) \geqslant u_{\theta_f}(c_{\theta'}) \tag{$IC^f_{\theta\theta'}$}$$

as the participation constraint for θ and the incentive compatibility constraint (IC) from θ to θ' in frame f.

First, we require a non-triviality condition.

Assumption 2.1 (Relevant Frames)**.** *For any $f, f' \in F$ there exists a type $\theta \in \Theta$ such that $\theta_f \neq \theta_{f'}$, i.e. the two frames induce a different valuation for this type.*

Second, we also assume additional structure on payoff types across frames in order to ensure that the problem remains one-dimensional despite the addition of frames.

Assumption 2.2 (Comonotonic Environment)**.** *For any* $\theta, \theta' \in \Theta$*,* $f, f' \in F$ *:*

$$\theta_f > \theta_{f'} \implies \theta'_f > \theta'_{f'} \quad \textit{and} \quad \theta_f > \theta'_f \implies \theta_{f'} > \theta'_{f'}. \tag{2.2.3}$$

The first part of the assumption implies that frames can be ordered by their impact on the valuation. There is a lowest frame, i.e. a frame inducing the lowest valuation for every type and a highest frame, i.e. a frame inducing the highest valuation for every type. The second part implies that types can also be ordered by their valuation independently of the frame. With slight abuse of notation, we denote the order on frames and types using regular inequality signs.

In many cases, a frame has a similar impact on different consumer types. The more effectively a seller emphasizes quality, for instance, the higher a consumer values quality irrespective of their type. The first part of our assumption is satisfied as long as the *direction* of the impact of a given frame is the same for all types. The second part is satisfied as long as the *size* of the effect is not too different between types relative to their initial difference in valuation. In particular, suppose there is a neutral frame f_n. The assumption is satisfied when the absolute impact of an enthusiastic frame $f_e > f_n$ is greater for high valuation consumers and the absolute impact of a pessimistic frame $f_p < f_n$ is greater for low valuation consumers: The firm can amplify the initial feelings of consumers and make all types more or less interested in quality, but it cannot manipulate them to the degree that the order is reversed.

Assumption 2.2 precludes any frame from impacting the valuations of different types in a different direction. For example, focusing a car buyers attention on emissions may increase the valuation of a "green" car for some buyers while reducing the valuation of all cars, including the "green" car, for others. Similarly, it rules out cases where the order of types by their payoff parameter depends on the frame. For example, the demand for health insurance coverage may be lower among smokers than nonsmokers if they are not reminded

about the long run effects of their habit, but is higher for smokers than nonsmokers if the effects of smoking are made salient during the sale of insurance. Together with Assumption 2.2, it also rules out that certain frames are specific to certain types. We discuss how we can relax our assumptions in Section 2.3.5.

2.2.2 Extensive-Form Decision Problems

We model the sales interaction as an extensive-form decision problem (single-player game), with a frame attached to each decision node. For example, the following situation can be represented by a two-stage extensive-form decision problem. First, the consumer contemplates whether to visit the store and then purchases a product in the store. Perhaps, the consumer is initially affected by marketing materials (frame at the root) and then the consumer is affected by the sales pitch in the store (frame at the second stage).

We define extensive-form decision problems (EDPs) by induction. Call an extensive decision problem with k stages a k-EDP. For any set S, let $\mathcal{P}(S)$ denote the set of all finite subsets of S containing the outside option $0 := (0,0)$. The set of 1-EDPs is $\mathcal{E}^1 := \mathcal{P}(C) \times F$, that is, a 1-EDP $e = (A, f)$ is a pair of a finite menu A and a frame f. For each $k > 1$, the set of k-EDPs is $\mathcal{E}^k := \mathcal{P}(\cup_{l=0}^{k-1} \mathcal{E}^l) \times F$, so that $e = (E, f) \in \mathcal{E}^k$ is a pair of a finite set of EDPs E and a frame f.[10] Finally, the set $\mathcal{E}$ of all finite EDPs is given by

$$\mathcal{E} := \bigcup_{k=1}^{\infty} \mathcal{E}^k. \tag{2.2.4}$$

In other words, an EDP is a finite tree with a frame assigned to each decision node. Moreover, the outside option is available for consumers at each stage.

Choice from Extensive-Form Decision Problems The preferences of consumers are represented by a utility function defined on the set of contracts. Therefore, the choice of

[10] Set $\mathcal{E}^0 = C$ to allow for terminal choices at all stages. Also, note that the choice of frame at each stage is unrestricted. In particular, the frame at the root of an EDP does not place any constraints on subsequent frames. We discuss the role of this assumption in Section 2.3.5.

consumers with type θ is well defined on the set $\mathcal{E}^1$ of 1-EDPs which are simply menus with frames. To define consumer choice for any EDP we assume that the consumers are *sophisticated*. Presented with a choice between several decision problems, the consumer correctly anticipates future choices, and chooses the continuation problem according to her current frame. The current self has no commitment power other then the choice of a suitable continuation problem.

Formally, we define the sophisticated consumer's choice in an EDP by induction. Call $\sigma : \Theta \to \mathcal{C}$ an *outcome* of a 1-EDP $(A, f) \in \mathcal{E}^1$ if $\sigma(\theta)$ maximizes u_{θ_f} on A. Suppose the consumer is facing a k-EDP $(E, f) \in \mathcal{E}^k$. Choosing between continuation problems in E, she anticipates her choice $\sigma^e(\theta) \in \mathcal{C}$ for each $e \in E$, but evaluates the contracts $\{\sigma^e(\theta)\}_{e\in E}$ in the current frame f. Let Σ^e be the set of outcomes of an EDP $e \in \cup_{l=0}^{k-1} \mathcal{E}^l$ with $\Sigma^e = \{\theta \mapsto e\}, \forall e \in \mathcal{E}^0$. Then σ is an outcome of (E, f) if there is a solution $\sigma^e \in \Sigma^e$ for every $e \in E$, such that $\forall \theta \in \Theta$

$$\sigma(\theta) \in \underset{\{\sigma^e(\theta):e\in E\}}{\operatorname{argmax}} u_{\theta_f}(\sigma^e(\theta)). \tag{2.2.5}$$

2.2.3 The Firm's Problem

The monopolist produces goods of quality q at convex cost $\kappa(q)$, that is twice-differentiable and satisfies boundary conditions to ensure interior efficient quantities: $\kappa(0) = 0$, $\kappa' > 0$, $\kappa'' > 0$ and $\forall \theta_f \in \mathbb{R}$, $v'_{\theta_f}(0) - \kappa'(0) > 0$, $\lim_{q\to\infty} v'_{\theta_f}(q) - \kappa'(q) < 0$.

Given a vector of contracts $\mathbf{c} = (c_\theta)_{\theta\in\Theta} = (p_\theta, q_\theta)_{\theta\in\Theta}$, the profit of the firm is given by

$$\Pi(\mathbf{c}) := \sum_{\theta\in\Theta} \mu_\theta (p_\theta - \kappa(q_\theta)). \tag{2.2.6}$$

Finally, the firm designs an EDP to maximize profits

$$\Pi^\star := \max_{e\in\mathcal{E},\mathbf{c}\in\Sigma^e} \Pi(\mathbf{c}) \tag{GP}$$

In analogy with the mechanism design literature, we say a vector of contracts $\mathbf{c}$ is *implemented* by an EDP e if it has a solution $\sigma = \mathbf{c}$. We then call $\mathbf{c}$ *implementable*. In these terms, the principal maximizes profits over the set of implementable contracts.[11]

We denote the efficient quality for a payoff parameter θ_f by $\widehat{q}_{\theta_f}$ with

$$v'_{\theta_f}(\widehat{q}_{\theta_f}) = \kappa'(\widehat{q}_{\theta_f}). \tag{2.2.7}$$

The efficient quality is unique, positive and strictly increasing in the payoff parameter by our assumptions on v and κ. We denote the contract offering this quality and extracting all surplus from the corresponding payoff type by $\widehat{c}_{\theta_f} := \left(v_{\theta_f}(\widehat{q}_{\theta_f}), \widehat{q}_{\theta_f}\right)$.

2.3 Optimal Screening

Before we analyze the general problem (GP), we analyze two special cases. In both cases, the problem collapses to a simple static screening problem.

Consider a simpler problem, where the firm can only choose a 1-EDP, i.e. a menu and a frame. In this case, it is optimal to choose the highest frame $h := \max F$, maximizing consumer valuation. Alternatively, suppose there is only one frame: $F = \{h\}$. Consequently, any EDP must use the same frame at every stage. The extensive-form structure does not matter in this case: As consumers are perfectly rational and dynamically consistent, they pick the most preferred contract from the extensive form. Hence, an extensive form is equivalent to an unstructured menu offering the same set of contracts.

[11] Note that (GP) does not require $\mathbf{c}$ to be the unique outcome of e. We only require partial implementation, as is customary in contract theory to ensure the compactness of the principal's problem. It can be shown that for any $\epsilon > 0$ the firm can design an EDP of the same structure with a unique outcome that achieves $\Pi^* - \epsilon$.

In both cases, the optimal menu corresponds to the solution of the classic monopolistic screening problem with the set of types $\{\theta_h\}_{\theta\in\Theta}$.[12]

Remark 2.1. Let $\mathbf{c}^*$ be the vector of contracts obtained by maximizing profits subject to the participation constraint for the lowest type and all IC constraints, all in frame h. Then the 1-EDP $\left(\emptyset \cup \{c^*_\theta\}_{\theta\in\Theta}, h\right)$ solves (GP) if

1. the firm is constrained to 1-EDPs, or
2. there is only one frame: $F = \{h\}$.

This shows that framing or extensive forms alone are not sufficient for our results. Only both features together allow the principal to use different frames at different stages of the decision and thereby generate violations of dynamic consistency that can be exploited.

2.3.1 Optimal Structure of the Extensive Form

In this section, we show that the optimal EDP has a simple three-stage structure. Towards this result, let us define a class of EDPs which share these structural features. Let h and l denote the highest and second highest frame:

$$h := \max F, \tag{2.3.1}$$

$$l := \max F \setminus \{h\}. \tag{2.3.2}$$

For any standard EDP the set of types Θ is partitioned into two sets, as there are two ways to present the contract associated to a given type: Contracts c_θ for revealed types ($\theta \in \Theta_R$) are presented at the root, while contracts for concealed types ($\theta \in \Theta_C$) are presented in separate continuation problems e_θ. Then, the three stages are (see Fig. 2.3.1):

1. Root: $f = h$; available choices: contracts c_θ for $\theta \in \Theta_R$ and EDPs e_θ for $\theta \in \Theta_C$.

[12] This is in contrast to (Salant and Siegel, 2018), where there is an ex-post participation constraint in an exogenously given frame, or a default option, i.e. a restricted menu choice problem for the principal.

2. Continuation problem: $f = l$; available choices: outside option and continue.

3. Terminal choice: $f = h$; available choices: c_θ and decoys $d^\theta_{\theta'}$ for $\theta' > \theta$.

Formally, we have the following definition.

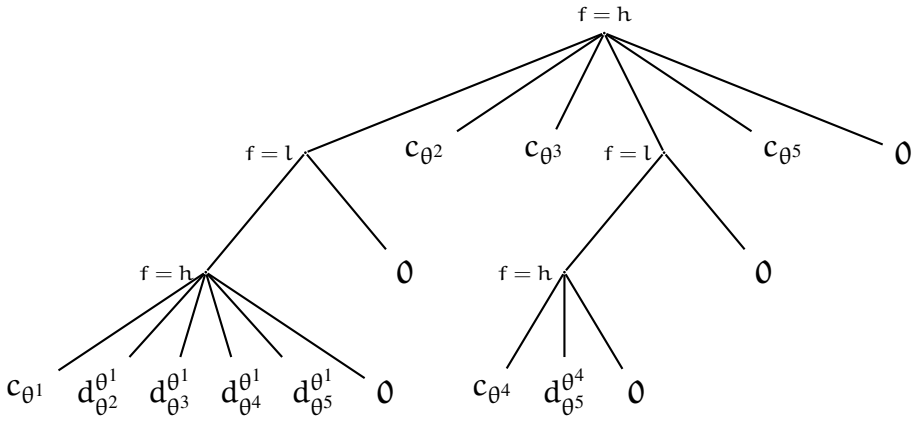
Definition 2.1. An EDP e is a *standard* EDP for a vector of contracts $\mathbf{c}$ if there exists a partition $\{\Theta_C, \Theta_R\}$ of Θ, and decoy contracts $\{d^\theta_{\theta'}\}_{\theta \in \Theta_C, \theta' > \theta}$, such that

$$e = \left(\{e_\theta\}_{\theta \in \Theta_C} \cup \{c_\theta\}_{\theta \in \Theta_R} \cup \{0\}, h \right), \text{ where} \tag{2.3.3}$$

$$e_\theta = \left(\left\{ \left(\{c_\theta, 0\} \cup \{d^\theta_{\theta'}\}_{\theta' > \theta}, h \right), 0 \right\}, l \right), \forall \theta \in \Theta_C. \tag{2.3.4}$$

The extensive form in Example 2.1 is a standard EDP. Type θ^1 is concealed – his contract is available only after a continuation problem – while type θ^2 is revealed – his contract is available immediately at the root.

Figure 2.3.1: A standard EDP for $(c_{\theta^1}, \ldots, c_{\theta^5})$ with $\Theta_R = \{\theta^2, \theta^3, \theta^5\}$ and $\Theta_C = \{\theta^1, \theta^4\}$.



Note that the notion of standard EDP is solely about the structure of the EDP. It puts no restrictions on the decoy contracts and is silent about choice. In particular, a standard EDP for $\mathbf{c}$ may not implement $\mathbf{c}$.[13]

Standard EDPs are sufficient to achieve the optimum.

[13] Whenever we state that a vector of contracts $\mathbf{c}$ is implemented by a standard EDP, however, it is understood that it is implemented by a standard EDP for $\mathbf{c}$.

Theorem 2.1. *If* $\mathbf{c}$ *is an optimal vector of contracts in (GP), then it is implemented by a standard EDP.*

Several observations follow from this result. First, the optimum can be achieved in *three stages for an arbitrary number of agent types*, even though the principal has arbitrarily complicated and long extensive forms at her disposal. As the number of types increases, the structure and length of the decision problem stays the same, only the number of available contracts increases. Furthermore, the optimal EDP has a simple structure that we interpret as follows: At the beginning, the consumer is presented a range of contracts $\{c_\theta\}_{\theta\in\Theta}$ while the salesperson focuses their attention on quality (high frame). Some of those contracts (those intended for revealed types) can be signed immediately, some others (those intended for concealed types) are only available after an additional procedure that gives the consumer some time to consider, while sales pressure is reduced (lower frame). This can be an explicit wait period, where the consumer is asked to think about the contract and recontact the seller. Alternatively, the change in frame could be achieved by a change in the salesperson or by acquiring a confirmation that this type of offer is even available for the consumer. If the consumer is still interested after this ordeal, she is presented with additional offers, the decoy contracts. On path, these offers remain unchosen, the consumer chooses the contract she initially intended to obtain.

Second, types are *separated at the root*. The principal does not use the extensive-form structure to discover the type of a consumer piecemeal, it is an implementation device to screen contracts against imitation.

Third, only the *two highest frames* are used. As we have seen in Remark 2.1, the principal desires to put everyone in the highest frame if there is no extensive-form structure. On the other hand, if every decision node uses the same frame, the extensive-form structure is irrelevant for agents choice. Consequently, the principal uses at least two frames in order to induce violations of dynamic consistency. As long as the principal induces such violations, the decoys can be constructed irrespective of the number of or cardinal differences between

the frames used. Hence, two frames are sufficient for the principal to reap *all potential gains* from such violations. Finally, only the highest two are used in the optimal EDP in order to maximize valuations.

2.3.2 Necessary and Sufficient Conditions for Implementation

In order to provide foundations for Theorem 2.1, we proceed in two steps. First, we identify an upper bound on profits in any EDP by providing necessary conditions every implementable vector of contracts has to satisfy. Then, returning to standard EDPs, we provide sufficient conditions on a vector of contracts ensuring that it can be implemented in this class. In particular, we explicitly construct decoy contracts and show that the principal can thereby eliminate downward IC constraints into concealed types.

Necessary Conditions for Implementation by General EDPs

Consider an arbitrary EDP implementing a vector of contracts $\mathbf{c} = (c_\theta)_{\theta\in\Theta}$. Denote the frame at the root by f_R. Extending the notion of revealed and concealed types from standard EDPs, for each type θ there are two possibilities: If there exists a path from the root to c_θ with all decision nodes set in f_R, then θ is called revealed. Alternatively, if every path from the root to c_θ involves at least one $f_\theta \neq f_R$, then θ is called concealed. As usual, we will denote the sets of revealed and concealed types by Θ_R and Θ_C, respectively.

First, consider participation constraints. If the path from the root to c_θ passes through a node in frame f, then, since the outside option is always available, c_θ needs to satisfy the corresponding participation constraint P_θ^f. In particular, every contract has to satisfy the constraint at the root $P_\theta^{f_R}$.

We now turn to incentive compatibility constraints. If θ is revealed, c_θ can be reached by any type from the root, as consumers are dynamically consistent when the frame does

not change along the path. Consequently, for any θ', c_θ must not be an attractive deviation:

$$u_{\theta'_{f_R}}(c_{\theta'}) \geqslant u_{\theta'_{f_R}}(c_\theta) \quad \forall \theta' \in \Theta. \tag{$IC^{f_R}_{\theta\theta'}$}$$

If θ is concealed, there is a change of frame along the path to c_θ. This induces a violation of dynamic consistency, which may make deviations into c_θ impossible. As we are looking for necessary conditions, we impose no incoming IC constraint in this case.

So far, we identified a family of conditions indexed by $(f_R, \{f_\theta\}_{\theta\in\Theta}, \Theta_C)$, such that a vector of contracts is implementable only if it satisfies at least one of them. The following proposition shows that without loss of generality, we can set $f_R = h$ and $f_\theta = l$. This is because the exact frame only matters for participation, while the change of frame affects IC. Consequently, the contract must satisfy the least restrictive participation constraints, i.e. in the highest and second highest frame.

Proposition 2.1. *If* $\mathbf{c}$ *is implemented by an EDP, then it satisfies constraints* $\{P^h_\theta\}_{\theta\in\Theta_R}$, $\{P^l_\theta\}_{\theta\in\Theta_C}$, *and* $\{IC^h_{\theta\theta'}\}_{\theta\in\Theta,\theta'\in\Theta_R}$ *for some partition* $\{\Theta_C, \Theta_R\}$ *of* Θ.

The necessary conditions illustrate the trade-off between using framing to increase consumer valuation and its use to reduce information rents. For revealed types, the participation constraint needs to be satisfied only in the highest frame, the frame resulting in the least restrictive constraint. This results in the greatest surplus. For concealed types, the participation constraint needs to be satisfied in the second highest frame. This reduces the surplus from the interaction. The principal is compensated for this reduction through the removal of IC constraints into concealed types.

Sufficient Conditions for Implementation by Standard EDPs

To construct a standard EDP that implements a vector of contracts $\mathbf{c}$, we proceed in two steps. First, we need to determine the set of concealed types. Then, we construct decoys for the continuation problems of these types. Clearly, type θ can be concealed in a standard

EDP only if c_θ satisfies the participation constraint P^l_θ, since otherwise he would prefer to opt out in the second stage.

If θ is concealed, the principal can design decoys in order to make *some* deviations into c_θ impossible. Whereas decoys cannot rule out all upward deviations, they can rule out *all downward* deviations into c_θ. Consequently, a vector of contracts is implementable by a standard EDP even if it does not satisfy the downward IC constraints, as long as the types that are attractive to imitate can be concealed.

Proposition 2.2. *If* $\mathbf{c}$ *satisfies the constraints* $\{P^h_\theta\}_{\theta\in\Theta_R}$, $\{P^l_\theta\}_{\theta\in\Theta_C}$, $\{IC^h_{\theta\theta'}\}_{\theta<\theta'}$, *and* $\{IC^h_{\theta\theta'}\}_{\theta\in\Theta,\theta'\in\Theta_R}$ *for some partition* $\{\Theta_C,\Theta_R\}$ *of* Θ, *then* $\mathbf{c}$ *is implemented by a standard EDP.*

As in Example 2.1, the principal constructs decoys to render downward deviations into concealed types impossible in the extensive form. This construction is the central step in our results and we therefore present it in the text. The construction ensures that (ii) if θ is concealed, no type $\theta' > \theta$ can imitate θ. The decoys don't interfere with the choices of any type at the root, as they will not be chosen from the continuation problem (i). In particular, θ chooses the intended contract (iii).

Lemma 2.1 (Decoy Construction)**.** *For any* $\theta \in \Theta$, c_θ *satisfies* P^l_θ *if and only if there exist decoys* $(d^\theta_{\theta'})_{\theta'>\theta}$, *such that the corresponding* e_θ *in (2.3.4) has an outcome* σ *that satisfies*

(i) $\sigma(\theta') \in \{0, c_\theta\}$ *for all* $\theta' \in \Theta$,

(ii) $\sigma(\theta') = 0$ *for all* $\theta' > \theta$, *and*

(iii) $\sigma(\theta) = c_\theta$.

Construction. The construction of the decoys and the continuation problem e_θ is illustrated in Fig. 2.3.2. At the terminal stage, agents are presented with the choice between the contract c_θ, the outside option and a set of decoys $\{d^\theta_{\theta'}\}_{\theta'>\theta}$, one for every type greater

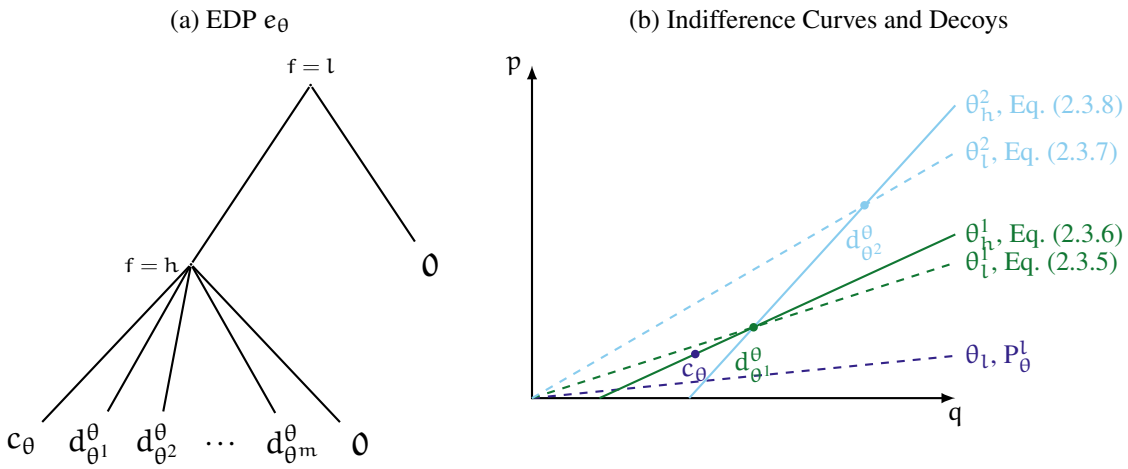


Figure 2.3.2: The construction of e_θ.

than θ. Given a contract c_θ, the decoy $d^\theta_{\theta^1}$ for the next largest type θ^1 is implicitly defined by the system

$$u_{\theta^1_l}(0) = u_{\theta^1_l}(d^\theta_{\theta^1}). \tag{2.3.5}$$

$$u_{\theta^1_h}(c_\theta) = u_{\theta^1_h}(d^\theta_{\theta^1}) \tag{2.3.6}$$

Then, decoy $d^\theta_{\theta^2}$ for the next type θ^2 solves

$$u_{\theta^2_l}(0) = u_{\theta^2_l}(d^\theta_{\theta^2}). \tag{2.3.7}$$

$$u_{\theta^2_h}(d^\theta_{\theta^1}) = u_{\theta^2_h}(d^\theta_{\theta^2}) \tag{2.3.8}$$

Proceeding by induction, we construct decoys for all $\theta' > \theta$. Now we define an outcome σ as follows. The single-crossing property ensures that each type $\theta' \geqslant \theta$ chooses their corresponding (decoy) contract out of the menu $\{c_\theta, d^\theta_{\theta_1}, \ldots, d^\theta_{\theta_m}, 0\}$ in frame h. At the root, θ will choose its contract since it satisfies P^l_θ and any $\theta' > \theta$ will choose the outside option. Finally, single crossing ensures that types $\theta' < \theta$ prefer the outside option over the decoys as well. We formally verify the construction in Section 3.A. □

2.3.3 Optimal Contracts

We show that the principal's problem (GP) over the space of extensive forms is equivalent to a two-step maximization problem based on the necessary conditions for implementation (Proposition 2.1). This relaxed problem characterizes the optimal vector of contracts.

An Equivalent Problem in Price-Quality Space

Let us summarize the necessary condition in an optimization problem. Recall that these conditions are indexed by the set of concealed types. This set is an additional choice variable for the principal in the relaxed problem.

$$\Pi^R = \max_{\Theta_C \subseteq \Theta} \max_{(c_\theta)_{\theta\in\Theta}} \Pi(\mathbf{c}) \tag{RP}$$

$$\text{s.t.} \quad u_{\theta_h}(c_\theta) \geqslant 0, \quad \forall \theta \in \Theta_R := \Theta \setminus \Theta_C \tag{P^h_θ}$$

$$u_{\theta_l}(c_\theta) \geqslant 0, \quad \forall \theta \in \Theta_C \tag{P^l_θ}$$

$$u_{\theta_h}(c_\theta) \geqslant u_{\theta_h}(c_{\theta'}), \quad \forall \theta \in \Theta, \theta' \in \Theta_R \tag{$IC^h_{\theta\theta'}$}$$

While it is not true that *every* vector of contracts satisfying the necessary conditions is implementable, the solution of (RP) is implementable, as it satisfies the sufficient conditions of Proposition 2.2.

Theorem 2.2. *A pair* $(\Theta_C, \mathbf{c})$ *solves (RP) if and only if* $\mathbf{c}$ *solves (GP). Moreover, such a solution exits and* $\mathbf{c}$ *can be implemented by a standard EDP with a set of concealed types* Θ_C.

In other words, the general problem (GP) attains the upper bound given by (RP)

$$\Pi^* = \Pi^R, \tag{2.3.9}$$

and the necessary conditions together with optimality is sufficient for implementation, even in the restricted class of standard EDPs.

Without the equivalent formulation, even verifying the existence of a solution to (GP) can be troublesome. Theorem 2.2 shows that instead of a complex optimization problem defined over extensive forms, the principal can solve well-behaved contracting problems over a menu of price-quality pairs, one for each potential set of concealed types and compare the attained values to find the optimum.[14] Once the principal found the (RP) optimal concealed types and vector of contracts, it is easy to construct a standard EDP implementing it using Lemma 2.1.

No Shut-down

In the classic model of screening, it is sometimes optimal for the monopolist to exclude low types by selling the outside option to them. In our model, this is never the case, because concealing a type is always strictly better for the monopolist than excluding it.

Proposition 2.3. *The optimal contract* (p_θ, q_θ) *for a type* θ *satisfies* $0 < \underline{q}_\theta \leqslant q_\theta \leqslant \widehat{q}_{\theta_h}$, *where* $v_{\theta_h}(\underline{q}_\theta) - \kappa(\underline{q}_\theta) := v_{\theta_l}(\widehat{q}_{\theta_l}) - \kappa(\widehat{q}_{\theta_l})$. *In particular, every type of consumer buys positive quality.*

Indeed, concealing a type can be interpreted as a soft form of shut-down. In order to eliminate information rents, the principal reduces the revenue extracted from a type. The key difference is that it can be achieved at a strictly positive quality, while extracting revenue from this type.

Optimal Contracts for Concealed Types

For concealed types, we can provide an additional lower bound on quality in the optimal contract. The contract for concealed types is subject to constraints in two frames: a

[14] Indeed, a stronger result holds. The (RP) optimal contracts for any, even suboptimal, set of concealed types is implementable. The problem can be further simplified by noting that only local IC - those into the nearest revealed types - are binding. See Section 3.A.

participation constraint in the lower frame l and an IC constraint in the higher frame h. Since concealed types cannot be imitated, there is no reason to distort their quality downward below the efficient quality in the lower frame, $\widehat{q}_{\theta_l}$. It can be optimal, however, to increase the quality above this level in order to deliver rent more cost-effectively in order to satisfy the IC constraint.

Proposition 2.4. *Consider a concealed type $\theta \in \Theta_C$. Then, the optimal quality is bounded between the efficient quality in frame l and h: $\widehat{q}_{\theta_l} \leqslant q_\theta \leqslant \widehat{q}_{\theta_h}$. In particular, the optimal contract is*

$$(p_\theta, q_\theta) = \begin{cases} \widehat{c}_{\theta_l}, & \text{if } \Delta_\theta \leqslant v_{\theta_h}(\widehat{q}_{\theta_l}) - v_{\theta_l}(\widehat{q}_{\theta_l}), \\ \left(v_{\theta_l}(q^\star), q^\star\right), & \text{if } \Delta_\theta \in \left[v_{\theta_h}(\widehat{q}_{\theta_l}) - v_{\theta_l}(\widehat{q}_{\theta_l}), v_{\theta_h}(\widehat{q}_{\theta_h}) - v_{\theta_l}(\widehat{q}_{\theta_h})\right], \\ \left(v_{\theta_h}(\widehat{q}_{\theta_h}) - \Delta_\theta, \widehat{q}_{\theta_h}\right), & \text{if } \Delta_\theta \geqslant v_{\theta_h}(\widehat{q}_{\theta_h}) - v_{\theta_l}(\widehat{q}_{\theta_h}), \end{cases} \tag{2.3.10}$$

where $q^\star$ solves $v_{\theta_h}(q^\star) - v_{\theta_l}(q^\star) = \Delta_\theta$, and $\Delta_\theta := \operatorname{argmax}_{\theta' \in \Theta_R} u_{\theta_h}(c_{\theta'})$ denotes the rent delivered to type $\theta \in \Theta_C$, and $\mathbf{c}$ is the optimal contract.

If the required rent is low, only the participation constraint in the low frame binds and the optimal contract is the efficient contract for the type in the low frame. As more rent needs to be delivered in the high frame, it becomes optimal to increase the quality of the product up to the efficient quality in the high frame.

The contract further illustrates the cost of concealing a type. From the perspective of the high frame, a concealed type always receives at least the minimum rent $v_{\theta_h}(\widehat{q}_{\theta_l}) - v_{\theta_l}(\widehat{q}_{\theta_l})$, reducing the payoff of the principal. The cost of concealing a type is decreasing in the information rent Δ. If it is sufficiently high (in the third regime of (2.3.10)), it is costless to conceal the type.

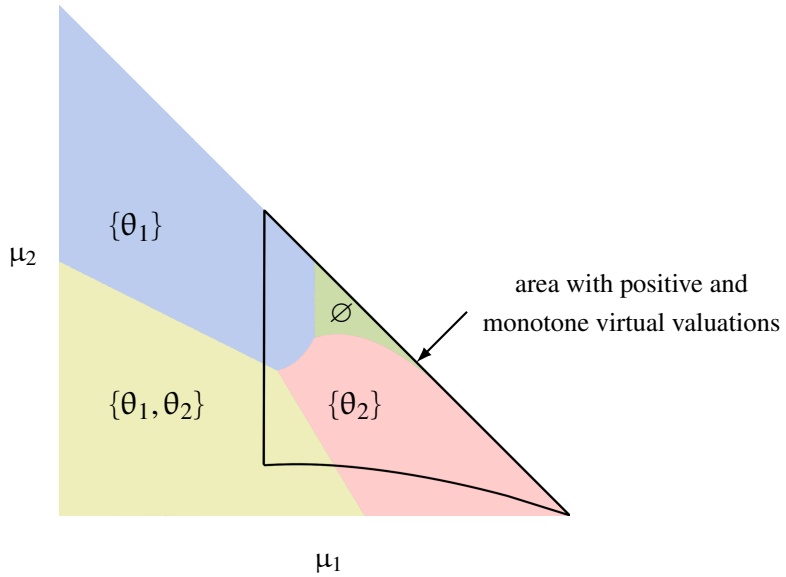
2.3.4 Optimal Concealed Types

One might conjecture that it is optimal for the principal to conceal low types and reveal high types. Even though this does not hold in general, this statement has a grain of truth: Revealing the highest type is always optimal. This is because types are concealed in order to eliminate downward deviations into them, which is not a concern for the highest type.

Remark 2.2. Suppose $(\Theta_C^*, \mathbf{c}^*)$ solves (RP) and the highest type $\overline{\theta} = \max \Theta$ is concealed, $\overline{\theta} \in \Theta_C^*$. Then $(\Theta_C^* \setminus \overline{\theta}, \mathbf{c}^*)$ also solves (RP).

In general, there are no other restrictions on the optimal set of concealed types, as the following linear-quadratic three-type example illustrates. In Fig. 2.3.3 we plot the regions of the probability simplex where particular sets of concealed types are optimal. All four cases are realized for some probabilities. In addition, the restriction to monotone virtual values that ensures monotonicity in the classic screening model doesn't rule out any configuration.

Figure 2.3.3: Optimal Θ_C for $\theta^1 = (1,3), \theta^2 = (4,5), \theta^3 = (5,6)$.



Loosely speaking, the concealed types are substitutes for the principal. Consider two types $\theta < \theta'$. By concealing θ, the principal reduces the rent θ' obtains, increasing the

costs of concealing θ' (as it is more costly to conceal a type if it has a low information rent; Proposition 2.4). In addition, a lower rent implies that concealing θ' has a lower gain as well, as information rents compound. Simlarly, concealing θ' reduces the benefit of concealing θ. This pattern of substitutability is reflected in Fig. 2.3.3 as the regions $\Theta_C = \{\theta_1\}$ and $\Theta_C = \{\theta_2\}$ touch.

Sufficiently Likely Types Are Revealed It is not profitable to conceal very likely types, since the gain from the reduction of information rents for other types is outweighed by the loss of profits that can be extracted from them directly.

Proposition 2.5. *For any type θ there exists a probability threshold $\bar{\mu}_\theta \in (0,1)$, such that for any $\mu_\theta \in [\bar{\mu}_\theta, 1]$, an optimal set of revealed types contains θ.*

This proposition suggests interpreting the contracts of revealed types as standard options that are relevant for common types of consumers and available immediately in the store, and the contracts for concealed types as specialty options relevant for rare types of consumers and available only on order.

High θ_l Favors Concealing The difference between the valuations in frames h and l determines the cost of concealing. If we fix all types, but *increase* the l-frame valuation of a concealed type, this cost is reduced and this type remains concealed.

Proposition 2.6. *Let Θ_C be an optimal sets of concealed types for (Θ, μ) and let $\theta \in \Theta_C$. Define $\tilde{\theta}$ such that $\tilde{\theta}_l \geqslant \theta_l, \tilde{\theta}_h = \theta_h$. Then, for the set of types $(\Theta \setminus \{\theta\}) \cup \{\tilde{\theta}\}$ there exists a solution of the principal's problem (RP) with the set of concealed types $\tilde{\Theta}_C := (\Theta_C \setminus \{\theta\}) \cup \{\tilde{\theta}\}$.*

Fixing the highest valuation the principal can achieve for each type, the cost of concealing is low if θ_l is high. We can interpret this as a more precise control of the principal over consumer valuations. With sufficient control, she will conceal all types except for the highest.

Proposition 2.7. *Fix* $\theta_h, \forall \theta \in \Theta$. *There exists an* $\varepsilon > 0$ *such that if* $\theta_h - \theta_l < \varepsilon, \forall \theta \in \Theta$, *then* $\Theta_C = \Theta \setminus \max \Theta$.

2.3.5 Discussion

Commitment and Direct Mechanisms Consumers are sophisticated but lack commitment. This is crucial, as the power of the principal to relax IC constraints by concealing types relies on the resulting dynamic inconsistency. In particular, this implies that our contracts cannot be implemented by a direct mechanism. Restricting to direct mechanisms effectively gives commitment as single-stage interaction does not allow for dynamic inconsistencies. As observed by Galperti (2015), with dynamically inconsistent agents, the revelation principle doesn't apply directly. Instead, agents need to resubmit their complete private information at every stage. In our setting, an indirect mechanism is more convenient. Alternatively, one could construct an equivalent "quasi-direct" mechanism in which a reported type is mapped to a menu of menus instead of a contract.

The principal, by contrast, as the designer of the single-agent mechanism, has and requires commitment.[15]

Weakening the Comonotonicity Assumption Our assumptions can be relaxed at the cost of transparency. Suppose (i) *there exists a unique highest frame*, i.e.

$$\exists h \in F, \{h\} = \bigcap_{\theta \in \Theta} \operatorname*{argmax}_{f \in F} \theta_f \tag{2.3.11}$$

[15] To see why, consider the terminal decision problem of a concealed type. Without commitment, the principal would increase quality on the contracts of concealed types and thereby violate the participation constraint in the low frame in the previous stage. If consumers anticipate this, the extensive-form decision problem unravels. Characterizing the outcome without commitment is beyond the scope of this paper.

and (ii) *comonotonicity holds locally*, i.e. for each θ

$$\exists l(\theta) \in \underset{f \in F \setminus \{h\}}{\operatorname{argmax}} \theta_f, \text{ such that } \forall \theta', \tag{2.3.12}$$

$$\theta'_h < \theta_h \implies \theta'_{l(\theta)} \leqslant \theta_{l(\theta)}, \tag{2.3.13}$$

$$\theta_h < \theta'_h \implies \theta_{l(\theta)} < \theta'_{l(\theta)} < \theta'_h. \tag{2.3.14}$$

Then our results generalize.[16] In particular, it is sufficient that there is an unambiguously highest and second highest frame. Note that we do not require the lowest type to be sensitive to framing, as framing is only used to place decoys.

Moreover, if $\theta_{l(\theta)} = \theta_h, \forall \theta \in \Theta$, then the principal achieves the first-best. This is because $l(\theta)$ will be used to eliminate IC constraints, without tightening participation constraints as $\theta_{l(\theta)} = \theta_h$.

Consider the following example that violates Assumption 2.2, but satisfies the assumption above. A product has n flaws and there are n types of consumers, such that for type i flaw i is irrelevant. The sales person can either avoid discussing the flaws (high frame), or focus the attention on one of them. Formally, denote the types $\theta^1, \ldots, \theta^n$ and $n+1$ and frames by $h, l_1, \ldots, l_n$ and suppose that

$$\forall i, \theta^i_h = \theta^i_{l_i}, \tag{2.3.15}$$

$$\forall i, j, \theta^i_h > \theta^i_{l_j}. \tag{2.3.16}$$

The principal can implement the first-best using a standard EDP with $\Theta_C = \Theta$ and type-specific low frame in the second stage.

Participation Constraint At Every Stage We assume that the agent can opt-out and choose the outside option at every stage of the decision problem. This is crucial for the

[16] The principal only uses the frames $h, \{l(\theta)\}_{\theta \in \Theta}$. Furthermore, suppose the set of concealed types is Θ_C. As long as local comonotonicity holds for all concealed types, our results generalize.

trade-off between value exaggeration and rent extraction. A weaker restriction would be to require the outside option to be available only somewhere in the decision problem, for example at the root or in every terminal decision node. In this case, the principal can achieve full extraction at the efficient quantities in the high frame for some parameters.

Restriction on the Choice of Frames We assume that the principal is unrestricted in the choice of frames and, in particular, that a change of framing is effective.[17] One might suppose that framing effects are instead partially "sticky". That is, if the principal is choosing f' after f, then the consumer's payoff type will be $\alpha\theta_{f'} + (1-\alpha)\theta_f$ for some $\alpha \in (0.5, 1]$. Then our results generalize.

Random Mechanisms We restrict the principal to use a deterministic extensive-form mechanism. One can show in examples that the principal can do strictly better by randomizing within the standard mechanism. Randomization allows the principal to smooth out the concealment of types. To see this, consider a situation with three types where it is optimal to conceal only the intermediate type and the P^l-constraint is binding in his contract. Then, the IC constraint from the highest to the intermediate type is slack at the root, as the intermediate type is concealed and the highest type obtains a strictly positive rent (from the IC to the lowest type). Consider a modification of the mechanism where the intermediate type is concealed with probability $1-\epsilon$ and revealed otherwise, obtaining the contract that is optimal ignoring the IC constraint of the highest type. The uncertainty resolves after the agent makes his decision at the root, but before the frame-change to l. In this mechanism, the highest type still strictly prefers not to imitate the intermediate type at the root if ϵ is sufficiently small. Furthermore, ex-ante profit is strictly greater as the "revealed" contract for the intermediate type is more profitable than concealing him.

[17] This is in line with evidence showing that framing effects, such as gain-loss, are observed within-subject (Tversky and Kahneman, 1981) and even among philosophers who claim to be familiar with the notion of framing, to have a stable opinion about the answer to the manipulated question and were encouraged to consider a different framing from the one presented (Schwitzgebel and Cushman, 2015).

2.4 Extensions

2.4.1 Naivete

Naive consumers understand the structure of the extensive-form decision problem and the choices available to them, but they fail to anticipate the effect of framing. Faced with an EDP, they pick the continuation problem containing the contract they prefer in their current frame.[18] They fail to take account of the fact that in this continuation problem, they may be in a different frame and end up choosing a different contract.

Setup

Towards the definition of a naive outcome, let $\mathcal{C}(e)$ denote the set of contracts in an EDP e. That is, letting $\mathcal{C}(e) = e$ for $e \in \mathcal{E}^0$, define

$$\mathcal{C}(e) := \cup_{e' \in E} \mathcal{C}(e'), \text{ for } e = (E, f). \tag{2.4.1}$$

Now call $s_\theta : \mathcal{E} \cup \mathcal{E}^0 \to \mathcal{E} \cup \mathcal{E}^0$ a naive strategy for θ if

$$s_\theta(e) = e \ \forall e \in \mathcal{E}^0 \tag{2.4.2}$$

$$s_\theta(E, f) \in E \ \forall (E, f) \in \mathcal{E} \tag{2.4.3}$$

$$\mathcal{C}(s_\theta(E, f)) \cap \underset{\mathcal{C}(e)}{\operatorname{argmax}}\, u_{\theta_f} \neq \varnothing. \tag{2.4.4}$$

Put differently, when facing $e = (E, f)$, a consumer identifies the f-optima in the set of all contracts in e, $\mathcal{C}(e)$, and chooses a continuation problem containing an optimum.

[18] A related idea is projection bias. (Loewenstein et al., 2003). The main difference is that our construction depends on the consumers' ability to forecast their future actions, not tastes. In this general sense, sophisticated consumers exhibit no projection bias, while naive consumers exhibit complete projection bias.

We call $\nu : \Theta \to \mathcal{C}$ a naive outcome of an EDP e if there exists a naive strategy profile s such that any type θ arrives at $\nu(\theta)$ by following s_θ, i.e. $\nu(\theta) = (\underbrace{s_\theta \circ \cdots \circ s_\theta}_{k \text{ times}})(e)$ for $e \in \mathcal{E}^k$. Let N^e be the set of all naive outcomes to an EDP e.

We consider the case when there are both naive and sophisticated consumers and the principal cannot observe their cognitive type. Let $\Theta = \Theta_S \sqcup \Theta_N$ be the disjoint union of the set of sophisticated types Θ_S and the set of naive types Θ_N. That is, we allow for the existence of $\theta^s \in \Theta_S$ and $\theta^n \in \Theta_N$ which differ only in their sophistication, but not in their tastes conditional on any frame. Define the optimal profits similarly to (GP) as

$$\begin{aligned} \Pi^* := \max_{e \in \mathcal{E}} & \Pi(\mathbf{c}) \\ & \text{s.t. } c_\theta \in \Sigma^e(\theta), \forall \theta \in \Theta_S, \\ & \quad c_\theta \in N^e(\theta), \forall \theta \in \Theta_N. \end{aligned} \tag{2.4.5}$$

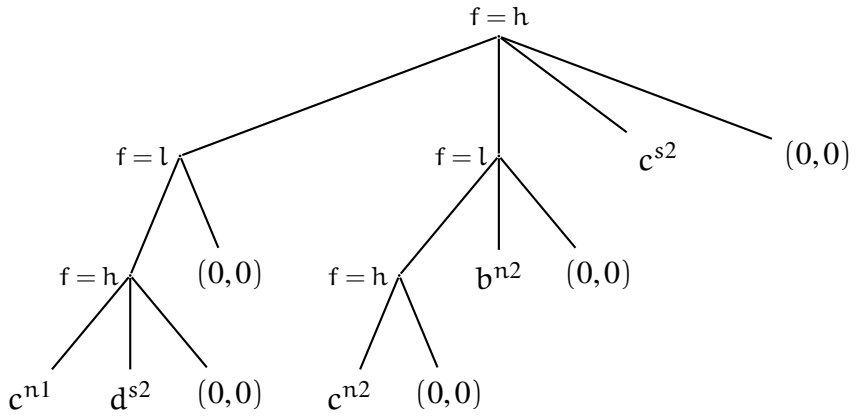
Optimal Structure and Contracts

We illustrate in an example how the principal can use decoys to screen when naive types are present.

Example 2.2. Recall from Example 2.1 that there are two frames, $\{l, h\}$, and two payoff types, $\{\theta^1, \theta^2\}$. The key construction can be illustrated using three equally likely types, two naive and one sophisticated. There is a naive version of both payoff types, and a sophisticated high type, formally $\Theta = \Theta_S \sqcup \Theta_N = \{\theta^{s2}\} \sqcup \{\theta^{n1}, \theta^{n2}\}$. In this setting, the principal can sell the h-efficient quality to naive consumers and fully extract their surplus. This creates no information rents for the sophisticated type – screening by cognitive type is free. As a result, she can also implement the high-frame full-extraction contract for θ^{s2}. The optimal EDP is given in Fig. 2.4.1. It implements $c^{n1} = (16, 4)$, $c^{n2} = (36, 6)$, $c^{s2} = (36, 6)$.

First, consider the sophisticated type. As in Example 2.1, the contract c^{n1} is more attractive than the implemented c^{s2}, but it is concealed using the decoy d^{s2}.

Figure 2.4.1: The optimal extensive-form decision problem in Example 2.2.



Let's turn to the naive types. The leftmost continuation problem is intended for θ^{n1}. Even though θ^{n1} is concealed, the principal extracts full surplus in the high frame. How is this possible? At the second stage in frame l, he indeed prefers the outside option over c^{n1}. But, he wrongly believes that he will choose the outside option after continuing. Hence, he continues and – back in frame h – chooses c^{n1}.

In order to implement the contract for θ^{n2}, the principal needs to use a decoy. At the root, he strictly prefers c^{n1} to c^{n2}. In order to lure him into the middle continuation problem, the principal introduces a decoy b^{n2}. This decoy works differently from the decoys used with sophisticated consumers.[19] It serves as bait and is the most preferred contract out of the whole decision problem for θ^{n2}. As a consequence, he continues into the middle continuation problem. There, the switch happens: b^{n2} is unattractive from the perspective of the low frame and θ^{n2} continues, expecting to pick the outside option in the continuation problem. Like θ^{n1} he reconsiders at the terminal node and ends up with c^{n2}. $\triangle$

This construction generalizes. The optimal EDP achieves the same outcome as if the principal knows which consumers are naive and the types of the naive consumers. Naive

[19] In this simple example, the two decoys, $d^{s2} = (40,8)$ and $b^{n2} = (40,8)$, coincide. This is the case because they are designed to distract from the same option, c^2, and there are no other contracts in the decision problem. It doesn't hold true in general, even if naive and sophisticated consumers share the same payoff type.

types don't receive information rents, they obtain the full extraction contract in the high frame. Sophisticated consumers obtain the optimal contract according to Theorem 2.2.

Theorem 2.3. *Let* e *an optimal EDP with the set of types* Θ *and* σ *and* ν *be its firm-preferred sophisticated and naive outcomes, respectively. Then there exists an EDP* e_S *that is optimal for the set of types* Θ_S *with conditional prior and its firm-preferred sophisticated outcome* σ_S*, such that*

$$\sigma(\theta) = \sigma_S(\theta), \quad \forall \theta \in \Theta_S \tag{2.4.6}$$

$$\nu(\theta) = \widehat{c}_{\theta_h}, \quad \forall \theta \in \Theta_N. \tag{2.4.7}$$

The optimal extensive-form decision problem retains the simple three-stage structure, we only add a continuation problem for each naive type to the extensive form described in Theorem 2.1. Consequently, the optimum can be achieved by a three-stage EDP with $|\Theta|$ continuation choices at the root, similar to a standard EDP, but with additional second-stage decoys.

The principal also uses decoy contracts for naive consumers, but their role is reversed: In the construction for sophisticated consumers, we placed decoys in continuation problems to make sure that no other type wants to enter the continuation problem, as they correctly anticipate that they would choose the decoy. The construction for naive consumers is a mirror image: Instead of decoys to repel imitators, we introduce decoys in order to lure types into their corresponding continuation problems. Agents wrongly believe that they will choose their respective decoy, which is the most attractive contract in the whole EDP for them in their current frame. Once types are separated at the root of the decision problem, the dynamic inconsistency introduced by changing frames allows the decision problem to reroute consumers from their decoy to the intended contract.

Welfare Gains from Sophistication

Are consumers better off if they are sophisticated? Welfare statements in the presence of framing are generally fraught with difficulty. Still, we can rank the contracts obtained by sophisticated and naive agents from a consumer perspective without taking a stand on the welfare-relevant frame.[20] In the following sense sophistication partially protects consumers from exploitation through the use of framing.

Remark 2.3. For all types, the contract under sophistication is weakly preferred to the contract under naivete from the perspective of every frame.[21]

From an efficiency perspective, the two cases are not unambiguously ranked. For naive consumers, the principal implements the efficient quality from the perspective of the highest frame. Quality is lower for sophisticated consumers, an efficiency gain from the perspective of all frames except the highest one.

Discussion: Partial Naivete

We can also extend our results to partial (magnitude) naivete. Denote the parameter determining the intensity of naivete by $\alpha \in [0,1]$, with $\alpha = 0$ representing full sophistication. Suppose a consumer with current payoff type θ anticipates a future choice that will actually be made according to payoff type θ'. Let $\widehat{\theta}(\theta, \theta', \alpha)$ denote what he currently perceives to be his future payoff type. Assume $\widehat{\theta}$ is increasing in the first two arguments and satisfies $\widehat{\theta}(\theta, \theta, \alpha) = \theta$ for all α. Under full sophistication we have $\widehat{\theta}(\theta, \theta', 0) = \theta'$, under full naivete $\widehat{\theta}(\theta, \theta', 1) = \theta$. This structure ensures that predictions satisfy comonotonicity (Assumption 2.2). Whenever $\alpha < 1$, we can extend the sophisticated construction by replacing θ_l^i by $\widehat{\theta}(\theta_h^i, \theta_l^i, \alpha)$ in (2.3.5) etc. Similarly, we can adjust the naive construction by modifying the decoy construction whenever $\alpha > 0$. In both cases, moving away from the baseline case

[20] The observation remains true if the choices in none of the frames are deemed welfare-relevant, as long as the welfare-relevant payoff parameters are weakly smaller than those induced by the highest frame.

[21] This can be interpreted as a weak improvement in the sense of Bernheim and Rangel (2009) if the two contracts are not identical.

increases the level of quality required in the decoys. If – contrary to what we assumed – providing very high quality decoys is not entirely costless or quality is bounded, we expect to see the sophisticated construction for agents with low α and the naive construction for individuals with high α.

2.4.2 Additional Participation Constraints and Cool-off Regulation

In many jurisdictions, regulation mandates a right to return a product for an extended period of time after the purchase. The express purpose of such regulation is to allow consumers to cool off and reconsider the purchase in a calm state of mind unaffected by manipulation by the seller.[22] Interestingly, such legislation typically only applies to door-to-door sales and similar situations of high sales pressure to which consumers did not decide to expose themselves. If consumers decide to enter a store or contact a seller, they are not protected by the law. This suggests that legislators consider the option to avoid the firm's sales pressure entirely to be sufficient to protect consumers. Our framework allows us to evaluate this intuition.

Consider a situation when consumers decide whether or not to go to the store in the neutral frame. One can interpret this decision as an additional *interim* participation decision at the root. Alternatively, suppose that there is a regulation that allows consumers to return a product if they wish to do so ex-post in the neutral frame (as in Salant and Siegel, 2018). One can interpret this decision as an additional *ex-post* participation decision at every terminal decision stage.

Formally, denote the neutral frame by $\mathfrak{n} \in \mathsf{F}$, $\mathfrak{n} < \mathfrak{h}$.[23] This is the frame the consumer is in when unaffected by direct sales pressure by the firm.[24] We call $\bar{\mathfrak{e}} := (\{\mathfrak{e}, \mathfrak{0}\}, \mathfrak{n})$ an

[22] E.g. directive 2011/83/EU: "the consumer should have the right of withdrawal because of the potential surprise element and/or psychological pressure".

[23] It is immediate that additional participation decisions in frame $\mathfrak{h}$ don't change the result, as all implemented contracts satisfy them.

[24] One possible effect of marketing is influencing this neutral frame, but we won't consider this margin.

interim modification[25] of e. Then e is an *EDP with an interim participation constraint* if it is an interim modification of some EDP. To define an *EDP with an ex-post participation constraint*, we define an *ex-post-modification* $\underline{e}$ of an EDP e recursively. First, for any $e \in \mathcal{E}^0$, let $\underline{e} := \bar{e}$. Having defined an ex-post modification on $\mathcal{E}^j, \forall j = 0, \ldots, k$, for any $e = (E, f) \in \mathcal{E}^{k+1}$, define its ex-post modification as $\underline{e} := (\{\underline{e}'\}_{e' \in E}, f)$.

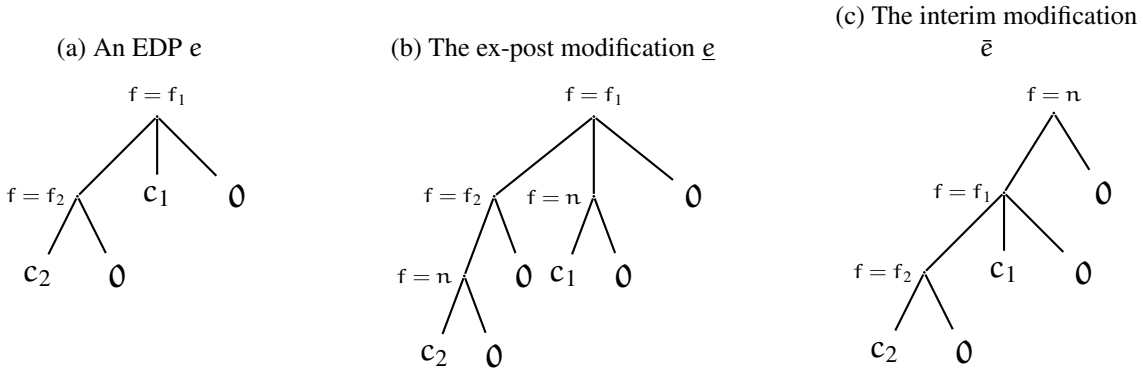


Figure 2.4.2: Interim and Ex-post Participation Constraints in Frame π

Sophisticated Consumers If consumers are sophisticated, both constraints are equivalent and imply that if a contract is chosen by type θ, then it must satisfy the additional participation constraint P^{π}_{θ}. The following observation shows that the firm implements the efficient allocation associated with frame π and leaves no information rent to consumers.

Remark 2.4. Suppose $\Theta = \Theta_S$. Let $\bar{e}^*$ and $\underline{e}^*$ be optimal EDPs with interim and ex-post participation constraints. Then their firm-preferred outcomes $\bar{\sigma}$ and $\underline{\sigma}$, respectively, are such that

$$\bar{\sigma}(\theta) = \underline{\sigma}(\theta) = \widehat{c}_{\theta_\pi}. \tag{2.4.8}$$

This observation is immediate from Theorem 2.1. The principal can remove all incoming IC constraints at the cost of an additional participation constraint in a lower frame. As

[25] Here, the notion of interim modification is defined on $\mathcal{E} \cup \mathcal{E}^0 \setminus \{0\}$. For simplicity, let $\bar{0} := 0$

such a constraint is introduced in any case with interim or ex-post participation constraints in a neutral frame, the principal can conceal all types without additional cost.[26]

Both restrictions protect against overpurchases relative to the preferences in the neutral frame, but cannot protect against the extraction of all information rents by exploiting induced violations of dynamic consistency. If sophisticated consumers can avoid the interaction with the firm, they indeed do not require additional protection by a right to return. They correctly anticipate their future actions and hence – given a choice – only interact with a seller, if the result will be acceptable to them from their current frame of reference.

Naive Consumers With naive consumers, we now need to distinguish between an interim choice to initiate the interaction and an ex-post right to return in the same neutral frame. While a right to return is still effective, naive consumers cannot protect themselves by avoiding the seller altogether.

Remark 2.5. Suppose $\Theta = \Theta_N$. Let $\bar{e}^*$ and $\underline{e}^*$ be optimal EDPs with interim and ex-post participation constraints. Then their firm-preferred naive outcomes $\bar{v}$ and $\underline{v}$ satisfy $\forall \theta \in \Theta$,

$$\bar{v}_\theta = \widehat{c}_{\theta_h}, \tag{2.4.9}$$

$$\underline{v}_\theta = \widehat{c}_{\theta_n}. \tag{2.4.10}$$

The intuition underlying the design of regulation does not apply for naive consumers. They are overly optimistic about the outcome of their interaction with the seller. As a result, the option to avoid the seller entirely is not sufficient to protect them from over-purchasing. In the optimal EDP, all consumers regret the purchase from the perspective of the neutral frame. A right to return even for in-store sales would offer them additional protection.

[26] Salant and Siegel (2018) show that the principal may not use framing when such a constraint is added to the problem of designing a framed menu. In particular, the principal cannot necessarily extract all rents without the use of an extensive form.

2.5 Conclusion and Discussion

We analyze the effect of framing in a model of screening. The principal can frame consumer decisions in several ways, affecting consumer valuations as expressed by their choices. Such a setting naturally leads to extensive-form decision problems. The firm uses framing not only to increase consumers valuations at the point of sale, but mainly to induce dynamic inconsistency and thereby reduce information rents, despite strategic sophistication of consumers. Our main result is that the optimal contracts can be implemented by an extensive-form decision problem with only three stages and two frames. At the initial interaction, only some contracts are immediately available, others are only available after the consumers' frame is lowered – which we interpret as a cool-off period. Upon recall, the consumer is presented with an extended menu, but chooses the expected option.[27]

This simple extensive form allows the firm to eliminate information rents at the cost of lower surplus and thereby achieve a payoff that is strictly larger than full surplus extraction at all but the highest frame. Even if consumers are protected by a shop-entry decision or right to return the product in an exogenously given neutral frame, they are not protected against the full extraction of their information rents.

We also characterize the outcome with naive consumers. The structure of the optimal extensive form and the contracts of sophisticated agents are robust to the presence of naive types. Naive types can be screened without generating any additional information rents.

Beyond Framing Throughout the analysis, we assumed that choice depends on exogenous factors of the presentation of the product that are chosen by the principal (i.e. the frame), but satisfies the axioms of utility maximization given every frame. If framing affects choice through focusing the attention of consumers on certain attributes, for example, we consider the case where these attributes are emphasized by the salesperson or the information material and assume that the properties of the choice set (such as an attribute

[27] This simple structure also supports the assumption of sophistication. In the optimum, consumers only need to grasp relatively short and intuitively understandable extensive forms.

being widely dispersed) are not relevant. Our construction induces and exploits a violation of dynamic consistency.

Such violations can also be caused by other factors, such as seasonal shifts in tastes or context effects. Hence, our construction can in principle be extended to such a setting. Consider, for example, the sale of a convertible. The current weather affects the valuation consumers have for convertibles. It constitutes an exogenous frame that cannot be manipulated directly by the principal. Analogously to Assumption 2.2, assume that all consumer types have a higher valuation for convertibles if the weather is nice (comonotonic frames). Furthermore, if one consumer type has a higher valuation for quality convertibles than another when the sun is shining, this is still true when it rains – albeit the valuation of both types is reduced (comonotonic types). Sophisticated consumers expect these shifts but consider tastes different from their current ones as mistakes. In such a setting, one could ask which pattern of taste changes is required to achieve the optimum. Our results imply that a simple pattern of taste changes (high-low-high) is sufficient.[28]

We expect that similar ideas can be applied to a setting with endogenous frames, e.g. the model of focusing (Kőszegi and Szeidl, 2013). There, a change of frame corresponds to introducing an option to the choice set that directs the focus more towards quality.[29] There is an important caveat, however. Even if it is possible to extend our construction into an analogous setting with endogenous frames, the resulting EDP may not be optimal. While in our setting, the frames have to be fixed for every decision node independently of the agent's type, context effects depend on the choice set. The choice set is generated by backward induction and hence type dependent. In effect, the frame can be type dependent.

[28] In particular, the car dealer can find the optimal EDP using our results and implement it as follows. The cars intended for revealed types can be bought immediately when the sun is shining. Cars for concealed types need to be pre-ordered. The pre-order period is sufficiently long to contain a sustained period of rain and the order can be canceled at any time. When the car is ready, it can be picked up only when the sun is shining. The consumer is offered a range of (decoy) options at this point, which are immediately available.

[29] This is possible by introducing a high quality-high price option that remains unchosen by every type Kőszegi and Szeidl (2013) argue that products that are extremely bad on all attributes are typically not taken into consideration. We don't require such products, a high quality product that is too expensive for every consumer type is sufficient.

Consequently, screening with choice-set dependent preferences is a considerably richer setting and left for future research.

References

Areni, Charles S. and David Kim, "The Influence of Background Music on Shopping Behavior: Classical Versus Top-Forty Music in a Wine Store," *ACR North American Advances*, 1993, *NA-20*.

Bernheim, B. Douglas and Antonio Rangel, "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *The Quarterly Journal of Economics*, January 2009, *124* (1), 51–104.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience and Consumer Choice," *Journal of Political Economy*, 2013, *121* (5), 803–843.

Bushong, Benjamin, Lindsay M. King, Colin F. Camerer, and Antonio Rangel, "Pavlovian Processes in Consumer Choice: The Physical Presence of a Good Increases Willingness-to-Pay," *American Economic Review*, 2010, *100* (4), 1556–71.

DelVecchio, Devon, H. Shanker Krishnan, and Daniel C. Smith, "Cents or Percent? The Effects of Promotion Framing on Price Expectations and Choice," *Journal of Marketing*, 2007, *71* (3), 158–170.

Eliaz, Kfir and Ran Spiegler, "Contracting with Diversely Naive Agents," *The Review of Economic Studies*, January 2006, *73* (3), 689–714.

_ and _ , "Consumer Optimism and Price Discrimination," *Theoretical Economics*, 2008, *3* (4), 459–497.

Esteban, Susanna and Eiichi Miyagawa, "Optimal Menu of Menus with Self-Control Preferences," *mimeo*, 2006.

_ **and** _ , "Temptation, Self-Control, and Competitive Nonlinear Pricing," *Economics Letters*, March 2006, *90* (3), 348–355.

_ **,** _ **, and Matthew Shum**, "Nonlinear Pricing with Self-Control Preferences," *Journal of Economic Theory*, July 2007, *135* (1), 306–338.

Galperti, Simone, "Commitment, Flexibility, and Optimal Screening of Time Inconsistency," *Econometrica*, 2015, *83* (4), 1425–1465.

Glazer, Jacob and Ariel Rubinstein, "A Model of Persuasion with Boundedly Rational Agents," *Journal of Political Economy*, December 2012, *120* (6), 1057–1082.

_ **and** _ , "Complex Questionnaires," *Econometrica*, 2014, *82* (4), 1529–1541.

Gourville, John T., "Pennies-a-Day: The Effect of Temporal Reframing on Transaction Evaluation," *Journal of Consumer Research*, March 1998, *24* (4), 395–403.

Gul, Faruk and Wolfgang Pesendorfer, "Temptation and Self-Control," *Econometrica*, 2001, *69* (6), 1403–1435.

Hardisty, David J., Eric J. Johnson, and Elke U. Weber, "A Dirty Word or a Dirty World? Attribute Framing, Political Affiliation, and Query Theory," *Psychological Science*, 2010, *21* (1), 86–92.

Heidhues, Paul and Botond Kőszegi, "Exploiting Naïvete about Self-Control in the Credit Market," *American Economic Review*, December 2010, *100* (5), 2279–2303.

_ **and** _ , "Naivete-Based Discrimination," *The Quarterly Journal of Economics*, May 2017, pp. 1019–1054.

Herrero, Maria José and Sanjay Srivastava, "Implementation via Backward Induction," *Journal of Economic Theory*, 1992, *56* (1), 70–88.

Johnson, Eric J., John Hershey, Jacqueline Meszaros, and Howard Kunreuther, "Framing, Probability Distortions, and Insurance Decisions," *Journal of Risk and Uncertainty*, 1993, *7*, 35–51.

Kőszegi, Botond, "Behavioral Contract Theory," *Journal of Economic Literature*, December 2014, *52* (4), 1075–1118.

_ **and Adam Szeidl**, "A Model of Focusing in Economic Choice," *The Quarterly Journal of Economics*, February 2013, *128* (1), 53–104.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, May 1997, *112* (2), 443–478.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin, "Projection Bias in Predicting Future Utility," *The Quarterly Journal of Economics*, 2003, *118* (4), 1209–1248.

Maskin, Eric and John Riley, "Monopoly with Incomplete Information," *The RAND Journal of Economics*, 1984, *15* (2), 171–196.

Moser, Christian and Pedro Olea de Souza e Silva, "Optimal Paternalistic Savings Policies," working paper, 2019.

Mussa, Michael and Sherwin Rosen, "Monopoly and Product Quality," *Journal of Economic Theory*, August 1978, *18* (2), 301–317.

North, Adrian C., Amber Shilcock, and David J. Hargreaves, "The Effect of Musical Style on Restaurant Customers' Spending," *Environment and Behavior*, September 2003, *35* (5), 712–718.

_ **, David J. Hargreaves, and Jennifer McKendrick**, "In-Store Music Affects Product Choice," *Nature*, November 1997, *390* (6656), 132–132.

_ **, Lorraine P. Sheridan, and Charles S. Areni**, "Music Congruity Effects on Product Memory, Perception, and Choice," *Journal of Retailing*, March 2016, *92* (1), 83–95.

Piccione, Michele and Ran Spiegler, "Price Competition Under Limited Comparability," *The Quarterly Journal of Economics*, January 2012, *127* (1), 97–135.

Salant, Yuval and Ariel Rubinstein, "(A, f): Choice with Frames," *The Review of Economic Studies*, 2008, *75* (4), 1287–1296.

_ **and Ron Siegel**, "Contracts with Framing," *American Economic Journal: Microeconomics*, August 2018, *10* (3), 315–346.

Schmitz, Hendrik and Nicolas R. Ziebarth, "Does Price Framing Affect the Consumer Price Sensitivity of Health Plan Choice?," *Journal of Human Resources*, 2017, *52* (1), 88–127.

Schwitzgebel, Eric and Fiery Cushman, "Philosophers' Biased Judgments Persist despite Training, Expertise and Reflection," *Cognition*, August 2015, *141*, 127–137.

Spiegler, Ran, *Bounded rationality and industrial organization*, Oxford University Press, 2011.

_ , "Competitive Framing," *American Economic Journal: Microeconomics*, 2014, *6* (3), 35–58.

Strotz, R. H., "Myopia and Inconsistency in Dynamic Utility Maximization," *The Review of Economic Studies*, 1955, *23* (3), 165.

Tversky, Amos and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, 1981, *211* (4481), 453–458.

Yu, Pei Cheng, "Optimal Retirement Policies with Present-Biased Agents," working paper, 2018.

Zhang, Wei, "Endogenous Preferences and Dynamic Contract Design," *The B.E. Journal of Theoretical Economics*, January 2012, *12* (1).

Appendix

2.A Preliminary Lemmata

Whenever types are enumerated by subscript i, we use notation $u_f^i := u_{\theta_f^i}$. For each θ_f^i, let $\succcurlyeq_f^i$ and (sometimes) $\succcurlyeq_{\theta_f^i}$ denote the corresponding preference relation. For any EDP e, let $\mathcal{C}(e)$ denote the set of all contracts, available in e. Formally, for $e = (A, f) \in \mathcal{E}^1$, $\mathcal{C}(e) := A$ and, recursively, for $e = (E, f) \in \mathcal{E}^k$, $\mathcal{C}(e) := \cup_{e' \in E} \mathcal{C}(e')$.

First, note that the consumer's preferences exhibit the single-crossing property, which is established in the following

Lemma 2.2 (Single-crossing property)**.** *For any two payoff types $\bar{\theta}, \underline{\theta} \in \mathbb{R}$, such that $\bar{\theta} \geqslant \underline{\theta}$, and contracts $x, y \in \mathcal{C}$, such that $q^y \geqslant q^x$, we have*

$$u_{\bar{\theta}}(y) \leqslant u_{\bar{\theta}}(x) \implies u_{\underline{\theta}}(y) \leqslant u_{\underline{\theta}}(x), \tag{2.A.1}$$

$$u_{\underline{\theta}}(y) \geqslant u_{\underline{\theta}}(x) \implies u_{\bar{\theta}}(y) \geqslant u_{\bar{\theta}}(x). \tag{2.A.2}$$

Proof. Take any $x, y \in \mathcal{C}$, such that $q^y \geqslant q^x$ and $u_{\underline{\theta}}(y) - u_{\underline{\theta}}(x) \geqslant 0$. Note that the increasing differences property ($v_{\theta q} > 0$) imply that

$$\begin{aligned} u_{\bar{\theta}}(y) - u_{\bar{\theta}}(x) &= v(\bar{\theta}, y) - v(\bar{\theta}, x) + p^y - p^x && \text{(2.A.3)} \\ &= \int_x^y \frac{\partial v_{\bar{\theta}}}{\partial q}(q)\,dq + p^y - p^x && \text{(2.A.4)} \\ &= \int_x^y \left(\frac{\partial v_{\underline{\theta}}}{\partial q}(q) + \int_{\underline{\theta}}^{\bar{\theta}} v_{\theta q}(\theta, q)\,d\theta \right) dq + p^y - p^x && \text{(2.A.5)} \\ &\geqslant v_{\underline{\theta}}(y) - v_{\underline{\theta}}(x) + p^y - p^x && \text{(2.A.6)} \\ &= u_{\underline{\theta}}(y) - u_{\underline{\theta}}(x) \geqslant 0. && \text{(2.A.7)} \end{aligned}$$

The proof of the first implication is analogous. □

Second, we prove the following result that ensures existence of optimal EDP.

Lemma 2.3. *For any prices $\bar{p} > \underline{p} \geqslant 0$ and payoff types $\bar{\theta}_{f'} > \underline{\theta}_f$, there exist $q \geqslant 0$, such that*

$$u_{\underline{\theta}_f}(\underline{p}, q) = u_{\bar{\theta}_f}(\bar{p}, q). \tag{2.A.8}$$

Or, equivalently, the function $q \mapsto v_{\bar{\theta}_{f'}}(q) - v_{\underline{\theta}_f}(q)$ is unbounded.

Proof. Take any $\bar{\theta}_{f'} > \underline{\theta}_f$ and set $\phi(q) := v_{\bar{\theta}_{f'}}(q) - v_{\underline{\theta}_f}(q)$. Note that ϕ is thrice differentiable and strictly increasing. Our assumption $\frac{\partial^3 v}{\partial q^2 \partial \theta} > 0$ implies that ϕ is strictly convex as

$$\phi''(q) = v''_{\bar{\theta}_{f'}}(q) - v''_{\underline{\theta}_f}(q) = \int_{\underline{\theta}_f}^{\bar{\theta}_{f'}} \frac{\partial^3 v_\theta}{\partial q^2 \partial \theta}(q)\,d\theta > 0. \tag{2.A.9}$$

Now, take any $\widehat{q} > 0$, and note that ϕ is weakly greater than its positively-sloped affine support function at $\widehat{q}$, which is unbounded.

Finally, since ϕ is unbounded, for any prices $\bar{p} > \underline{p} \geqslant 0$, there exists $q \geqslant 0$, such that

$$\bar{p} - \underline{p} = \phi(q) = v_{\bar{\theta}_f'}(q) - v_{\underline{\theta}_f}(q) \Leftrightarrow u_{\underline{\theta}_f}(\underline{p}, q) = u_{\bar{\theta}_f}(\bar{p}, q). \tag{2.A.10}$$

□

Lemma 2.4. *The efficient quality for payoff type* θ_f *defined as*

$$\widehat{q}_{\theta_f} := \underset{q \geqslant 0}{\operatorname{argmax}}\, v_{\theta_f}(q) - \kappa(q) \tag{2.A.11}$$

exists, is unique and increasing in θ_f.

Proof. Define the surplus function $\zeta_{\theta_f} : \mathbb{R}_+ \to \mathbb{R}$ as

$$\zeta_{\theta_f}(q) := v_{\theta_f}(q) - \kappa(q) \tag{2.A.12}$$

and note that it is continuous, twice differentiable and satisfies

$$\zeta_{\theta_f}(0) = 0 \tag{2.A.13}$$

$$\zeta'_{\theta_f}(0) > 0 \tag{2.A.14}$$

$$\lim_{q \to \infty} \zeta'_{\theta_f}(q) < 0 \tag{2.A.15}$$

$$\zeta''_{\theta_f} > 0, \tag{2.A.16}$$

$$\frac{\partial \zeta'_{\theta_f}(\widehat{q}_{\theta_f})}{\partial \theta_f} > 0. \tag{2.A.17}$$

Properties (2.A.14) and (2.A.15) together with the Mean Value Theorem ensure that there exists a unique global efficient quantify $\widehat{q}_{\theta_f}$ defined as

$$\zeta'_{\theta_f}(\widehat{q}_{\theta_f}) = 0 \Leftrightarrow v'_{\theta_f}(\widehat{q}_{\theta_f}) = \kappa'(\widehat{q}_{\theta_f}). \tag{2.A.18}$$

Note that (2.A.16) implies

$$\operatorname{sgn}(\zeta'_{\theta_f}(q)) = \operatorname{sgn}(\widehat{q}_{\theta_f} - q). \tag{2.A.19}$$

In addition, (2.A.17) implies that $\widehat{q}_{\theta_f}$ is increasing in θ_f. □

2.B Proofs

Proof of Remark 2.1 on page 73: First, consider a 1-EDP: By the usual arguments, the revenue maximal menu satisfies monotonicity, participation at the bottom and local downward IC constraints, the latter two with equality. Conversely, these constraints jointly imply the full set of constraints. We index types in an increasing order, i.e. $\Theta = \{\theta_1, \ldots, \theta_n\}$ with $\theta_i < \theta_{i+1}$. Suppose towards a contradiction that the revenue under frame $f < h$ is maximal with an optimal menu $\{(p_i, q_i)\}_{i\in\{1,\ldots,n\}}$. Consider the menu $\{(p'_i, q_i)\}_{i\in\{1,\ldots,n\}}$ set in frame h with

$$p'_i = p_i + \Delta_i \tag{2.B.1}$$

$$\Delta_1 = v\left(\theta_h^1, q_1\right) - v\left(\theta_f^1, q_1\right) \tag{2.B.2}$$

$$\Delta_i = \Delta_{i-1} + v(\theta_h^i, q_i) - v(\theta_f^i, q_i) - \left[v(\theta_h^i, q_{i-1}) - v(\theta_f^i, q_{i-1})\right] \tag{2.B.3}$$

This menu still satisfies monotonicity, participation at the bottom and the local downward IC constraints are still binding, as

$$\begin{aligned}
v(\theta_h^i, q_i) - p_i' &= v(\theta_f^i, q_i) + v(\theta_h^i, q_i) - v(\theta_f^i, q_i) - p_i - \Delta_i && \text{(2.B.4)}\\
&= v(\theta_f^i, q_{i-1}) + v(\theta_h^i, q_i) - v(\theta_f^i, q_i) - p_{i-1} - \Delta_i && \text{(2.B.5)}\\
&= v(\theta_f^i, q_{i-1}) - p_{i-1} - \Delta_{i-1} + \left[v(\theta_h^i, q_{i-1}) - v(\theta_f^i, q_{i-1})\right] && \text{(2.B.6)}\\
&= v(\theta_h^i, q_{i-1}) - p_{i-1} - \Delta_{i-1} && \text{(2.B.7)}\\
&= v(\theta_h^i, q_{i-1}) - p_{i-1}' && \text{(2.B.8)}
\end{aligned}$$

Hence all other IC and P are satisfied by the usual arguments. Note that $\Delta_i \geqslant 0$ by single crossing, hence expected revenue is strictly (weakly if all types pool at $q = 0$, but this is never optimal) higher under the modified contract in the higher frame.

The claim for $F = \{h\}$ follows from the following: $\forall \theta \in \Theta, \forall e \in \mathcal{E}^k, \Sigma^e(\theta) = \operatorname{argmax}_{\mathcal{C}(e)} u_{\theta_h}$ by induction on k. The base $k = 1$ is by definition. Take any $e = (E, h) \in \mathcal{E}^{k+1}$. σ is an outcome to e iff

$$\sigma(\theta) \in \underset{\{\sigma^j(\theta)\}_{e^j \in E}}{\operatorname{argmax}} u_{\theta_h}(\sigma^j(\theta)) \text{ with } \sigma^j \in \Sigma^{e^j}(\theta) = \underset{\mathcal{C}(e^j)}{\operatorname{argmax}} u_{\theta_h}. \tag{2.B.9}$$

Therefore, σ is an outcome to e if and only if

$$\sigma(\theta) \in \underset{\cup_{e^j \in E} \operatorname{argmax}_{\mathcal{C}(e^j)} u_{\theta_h}}{\operatorname{argmax}} u_{\theta_h} = \underset{\mathcal{C}(e)}{\operatorname{argmax}} u_{\theta_h} \tag{2.B.10}$$

Hence $e' := (\mathcal{C}(e), h)$ is outcome equivalent to e and the optimal EDP is equivalent to the optimal menu. □

The proof of Theorem 2.1 relies on many arguments that are required for the proofs of the following results as well. To avoid repetition, we prove Theorem 2.1 and Theorem 2.2 together.

Proof of Proposition 2.1 on page 77: The necessity of the constraints for a given set of frames $f_R, \{f_\theta\}_{\theta \in \Theta^C}$ is derived in the main text. In particular, we saw that the incentive compatibility constraints are determined by the frames used on the path to the contract of the imitated type, not the imitating type and frames used on the path to $c_{\theta'}$ cannot eliminate the IC constraints from θ' to θ for any $\theta', \theta \in \Theta$

To prove the proposition, it remains to show that we can assume that $f_R = h$ and $f_\theta = l$ for all concealed types. Suppose $f_R \neq h$. But then we can set $\Theta'_C = \Theta$. As all contracts satisfy the participation constraint in $f_R < h$, they satisfy the participation constraint in l. As there are no incentive compatibility constraints with $\Theta'_C = \Theta$, all constraints associated to this set of hidden types are satisfied. Suppose instead that $f_R = h$ but for some type $\theta' \in \Theta_C$ we have $f_{\theta'} \neq l$. But then the participation constraint for $f_{\theta'} = l$ is satisfied and hence the set of contracts is feasible in the relaxed problem. □

Before we proceed, we prove a more detailed decoy construction lemma.

Lemma 2.5. *For any ordered vector of types* $(\theta^j)_{j=0}^n$ *and contract* x*, there exists a sequence of decoys* $\mathcal{D}(x, (\theta^j)_{j=0}^n) := (d_j)_{j=1}^n$*, such that* $\forall j \geqslant 1, \forall k \neq j$*, we have*

1. $q_j^d \geqslant q_{j-1}^d$ *(decoy quantities are increasing),*
2. $d_j \sim_j^h d_{j-1}$*,*
3. $d_j \succcurlyeq_j^h x$ *and* $d_j \succcurlyeq_j^h d_k$ *(*θ^j *chooses* d_j*),*
4. $x \succcurlyeq_0^h d_j$ *(*θ^0 *chooses* x*),*
5. $0 \succcurlyeq_j^l d_j$ *(decoys are undesirable in* l*).*

Proof. Let $d_0 = x$ for brevity. We construct the decoys $d_i = (p_i, q_i)$ recursively.[30] For $i \in \{1, \ldots, n\}$, pick

$$d_i \sim_{\theta_l^i} 0 \tag{2.B.11}$$

$$d_i \sim_{\theta_h^i} d_{i-1} \tag{2.B.12}$$

or equivalently

$$v(\theta_l^i, q_i) - p_i = 0 \tag{2.B.13}$$

$$v(\theta_h^i, q_i) - p_i = v(\theta_h^i, q_{i-i}) - p_{i-i} \tag{2.B.14}$$

Existence of such a d_i follows from Lemma 2.3. To verify this construction, we proceed through a series of claims.

Claim 2.1. Decoy quantities are increasing: $q_i \geqslant q_{i-1}$.

Proof of Claim 2.1: By the two defining relations

$$v(\theta_l^i, q_i) = p_i \tag{2.B.15}$$

$$v(\theta_h^i, q_i) - p_i = v(\theta_h^i, q_{i-1}) - p_{i-1} \tag{2.B.16}$$

Hence

$$v(\theta_h^i, q_i) - v(\theta_l^i, q_i) = v(\theta_h^i, q_{i-1}) - v(\theta_{i-1}^l, q_{i-1}) > v(\theta_h^i, q_{i-1}) - v(\theta_l^i, q_{i-1}) \tag{2.B.17}$$

and $q_i > q_{i-1}$ is established as it is implied by single crossing from

$$\int_{\theta_l^i}^{\theta_h^i} v_\theta(\theta, q_i) d\theta > \int_{\theta_l^i}^{\theta_h^i} v_\theta(\theta, q_{i-1}) d\theta \tag{2.B.18}$$

[30] The present proof can be extended to the case without ordered types (but maintaining ordered frames).

$\triangle$

This also shows that all (p, q) are positive, as $q_i \geqslant q_0 = q_x$.

Claim 2.2. The decoy intended for type θ_i is chosen by this type: $d_i \in \arg\max_{\preccurlyeq_{\theta_h^i}} d_j$.

Proof of Claim 2.2: We will show that for all j we have $d_i \succcurlyeq_{\theta_h^i} d_j$. First, suppose $j < i$. Then for all $k \in [j, i]$ we have

$$d_k \sim_{\theta_h^k} d_{k-1} \tag{2.B.19}$$

and $q_k > q_{k-1}$. This implies

$$d_k \succcurlyeq_{\theta_h^k} d_{k-1} \tag{2.B.20}$$

and by since $\theta_h^i \geqslant \theta_h^k$

$$d_k \succcurlyeq_{\theta_h^i} d_{k-1} \tag{2.B.21}$$

The desired result follows by transitivity.

Second, suppose $j > i$. Again for all $k \in [i, j]$,

$$d_k \sim_{\theta_h^k} d_{k-1} \tag{2.B.22}$$

and $q_k > q_{k-1}$. But then

$$d_k \preccurlyeq_{\theta_h^i} d_{k-1} \tag{2.B.23}$$

for every decoy since $\theta_h^i \geqslant \theta_h^k$. The desired result again follows by transitivity. $\triangle$

$\square$

Proof of Lemma 2.1 on page 78: A continuation problem for type $\theta \in \Theta$ with contract c_θ satisfying all three properties is given by $e_\theta = (\{0, (\{0, c_\theta\} \cup \{d_{\theta'}\}_{\theta' > \theta}, h)\}, l)$, where the contracts $\{d_{\theta'}\}_{\theta' > \theta}$ are constructed in Lemma 2.5 as $(d_{\theta'})_{\theta' \geqslant \theta} := \mathcal{D}(c_\theta, (\theta')_{\theta' \geqslant \theta})$.

By construction, type θ chooses c_θ from the terminal problem and since c_θ satisfies the participation constraint in the low frame, $c_\theta \in \Sigma^{e_\theta}(\theta)$. For higher types, the terminal

decision problem resolves to the menu $\{d_{\theta'}, 0\}$ and by construction the outside option is weakly preferred in the low frame. Having established (ii) and (iii), it remains to show (i). Consider a type $\theta' < \theta$. In the terminal decision problem, we have $d_0 \succcurlyeq_{\theta_h} d_i$ and $q_i \geqslant q_0$, hence by single crossing $d_0 \succ_{\theta'_h} d_i$, which establishes that a lower type never chooses any of the decoys. □

Proof of Proposition 2.2 on page 78: Suppose that $\mathbf{c}$ satisfies $\{P^h_\theta\}_{\theta\in\Theta_R}$, $\{P^l_\theta\}_{\theta\in\Theta_C}$, $\{IC^h_{\theta\theta'}\}_{\theta<\theta'}$, $\{IC^h_{\theta\theta'}\}_{\theta\in\Theta,\theta'\in\Theta_R}$ for some partition $\{\Theta_C, \Theta_R\}$ of Θ. Then let e_θ for each $\theta \in \Theta_C$ be constructed as in Lemma 2.1 and consider a standard EDP

$$e^* = \Big(\{e_\theta\}_{\theta\in\Theta_C} \cup \{c_\theta\}_{\theta\in\Theta_R} \cup \{0\}, h\Big). \tag{2.B.24}$$

Notice that from Lemma 2.1 it follows for each $\theta \in \Theta_C$ that there exist an outcome σ^θ of e_θ, such that

$$\sigma^\theta(\theta') \in \{0, c_\theta\}, \forall \theta' \in \Theta, \tag{2.B.25}$$

$$\sigma^\theta(\theta') = 0, \forall \theta' > \theta, \tag{2.B.26}$$

$$\sigma^\theta(\theta) = c_\theta. \tag{2.B.27}$$

Now let σ be such that $\sigma(\theta) = c_\theta$. To show that σ is an outcome of e^*, notice that constraints $\{IC^h_{\theta\theta'}\}_{\theta\in\Theta,\theta'\in\Theta_R}$ and $\{P^h_\theta\}_{\theta\in\Theta_R}$ imply that $\forall\theta \in \Theta$,

$$\sigma(\theta) = c_\theta \in \underset{\{c_{\theta'}\}_{\theta'\in\Theta_R}}{\operatorname{argmax}}\ u_{\theta_h}. \tag{2.B.28}$$

Similarly, constraints $\{IC^h_{\theta\theta'}\}_{\theta<\theta'}$ and $\{P^h_\theta\}_{\theta\in\Theta_R}$ imply that $\forall\theta \in \Theta$,

$$\sigma(\theta) = c_\theta \in \underset{\{c_{\theta'}\}_{\theta\leqslant\theta'}}{\operatorname{argmax}}\ u_{\theta_h}. \tag{2.B.29}$$

Therefore, σ satisfies

$$\sigma(\theta) \in \underset{\{\sigma^{\theta'}(\theta)\}_{e_\theta} \cup \{c_{\theta'}\}_{\theta' \in \Theta_R}}{\operatorname{argmax}} u_{\theta_h}, \tag{2.B.30}$$

which means that it is an outcome of e^*. □

Proof of Theorem 2.1 on page 75 and Theorem 2.2 on page 80: Let $(c_\theta)_{\theta \in \Theta}, \Theta_C$ be a solution to the relaxed problem and e^* a standard EDP with decoys for all $\theta \in \Theta_C$ constructed as in Lemmata 2.5 and 2.1. We need to show that e^* implements $(c_\theta)_{\theta \in \Theta}$.

First, note that Σ^e is rectangular, i.e. if $\sigma, \sigma' \in \Sigma^e$ with $\sigma(\theta) \neq \sigma'(\theta)$ and $\sigma(\theta') \neq \sigma'(\theta')$, there exists a $\sigma^* \in \Sigma^e$ with $\sigma^* = \sigma$ except $\sigma^*(\theta') = \sigma'(\theta')$.

It follows from the IC constraints that there is no strictly profitable deviation into contracts of revealed types, i.e. $\Sigma^{e^*}(\theta) \cap \{c'_\theta\}_{\theta' \in \Theta_R \setminus \theta} \neq \emptyset$ implies that $c_\theta \in \Sigma^{e^*}(\theta)$. From Lemma 2.1, it follows that type θ cannot deviate downwards into concealed types and that no decoys are chosen, i.e. $\Sigma^{e^*}(\theta) \subset \{c_{\theta'}\}_{\Theta \setminus \{\theta' < \theta | \theta' \in \Theta_C\}}$. It remains to show that there are no strictly profitable upwards deviations in e^* to complete the proof, establishing $c_\theta \in \Sigma^{e^*}(\theta)$ for all $(c_\theta)_{\theta \in \Theta}$.

As the proof relies on properties of the solution to (RP), we start by simplifying the relaxed problem. Define

$$\eta(\theta) \coloneqq \max\left\{\theta' \in \Theta_R | \theta' < \theta\right\} \tag{2.B.31}$$

the closest revealed type below a given type θ, and

$$\chi(\theta) \coloneqq \min\left\{\theta' \in \Theta_R | \theta' > \theta\right\} \tag{2.B.32}$$

the closest revealed type above a given type θ. We now define the doubly relaxed problem, where we remove all but the downward IC constraints into the closest revealed type and the

upwards IC constraints into the next largest revealed type.

$$\max_{\Theta_C \subseteq \Theta} \max_{\{(p_\theta, q_\theta)\}_{\theta \in \Theta}} \sum_{\theta \in \Theta} \mu_\theta \left(p_\theta - \kappa(q_\theta) \right) \tag{2.B.33}$$

$$\text{s.t. } v_{\theta_h}(q_\theta) - p_\theta \geqslant 0 \quad \forall \theta \in \Theta_R$$

$$v_{\theta_l}(q_\theta) - p_\theta \geqslant 0 \quad \forall \theta \in \Theta_C$$

$$v_{\theta_h}(q_\theta) - p_\theta \geqslant v_{\theta_h}(q_{\eta(\theta)}) - p_{\eta(\theta)} \quad \forall \theta \in \Theta \tag{2.B.34}$$

$$v_{\theta_h}(q_\theta) - p_\theta \geqslant v_{\theta_h}(q_{\chi(\theta)}) - p_{\chi(\theta)} \quad \forall \theta \in \Theta \tag{2.B.35}$$

We have the following

Lemma 2.6. *The solution to the doubly relaxed problem satisfies* R*-monotonicity*

$$\theta, \theta' \in \Theta_R, \, \theta > \theta' \implies q_\theta \geqslant q_{\theta'} \tag{2.B.36}$$

and solves the relaxed problem.

Proof. Consider $\theta \in \Theta_R$, $\eta := \eta(\theta) < \theta$. Then $\theta = \chi(\eta)$ and we have

$$v_{\theta_h}(q_\theta) - p_\theta \geqslant v_{\theta_h}(q_\eta) - p_\eta \tag{2.B.37}$$

$$v(\eta_h, q_\eta) - p_\eta \geqslant v(\eta_h, q_\theta) - p_\theta \tag{2.B.38}$$

and hence

$$v_{\theta_h}(q_\theta) - v(\eta_h, q_\theta) \geqslant v_{\theta_h}(q_\eta) - v(\eta_h, q_\eta) \tag{2.B.39}$$

$$\int_{\eta_h}^{\theta_h} v_\theta(t, q_\theta) dt \geqslant \int_{\eta_h}^{\theta_h} v_\theta(t, q_\eta) dt \tag{2.B.40}$$

which implies $q_\theta > q_\eta$, establishing R-monotonicity by transitivity.

Then, we need to show that all IC are implied by the local IC. Let us proceed by induction on the number of types in Θ_R between the source of the $IC^h_{\theta\theta'}$ constraint θ and

it's target θ'. If there are no revealed types between, then $\theta' = \eta(\theta)$ (resp. $\chi(\theta)$) and we are done. Suppose that all constraints with up to n intermediate revealed types are implied and let $\theta > \theta'$, $\theta' \in \Theta_R$ with $n+1$ intermediate revealed types. The argument for $\theta' > \theta$ is identical. Then

$$
\begin{aligned}
v_{\theta_h}(q_\theta) - p_\theta &\geqslant v_{\theta_h}(q_{\eta(\theta)}) - p_{\eta(\theta)} && \text{(2.B.41)}\\
&= \big(v_{\theta_h}(q_{\eta(\theta)}) - v(\eta(\theta)_h, q_{\eta(\theta)})\big) + v(\eta(\theta)_h, q_{\eta(\theta)}) - p_{\eta(\theta)} && \text{(2.B.42)}\\
&\geqslant \big(v_{\theta_h}(q_{\eta(\theta)}) - v(\eta(\theta)_h, q_{\eta(\theta)})\big) + v(\eta(\theta)_h, q_{\theta'}) - p_{\theta'} && \text{(2.B.43)}\\
&\geqslant \big(v_{\theta_h}(q_{\theta'}) - v(\eta(\theta)_h, q_{\theta'})\big) + v(\eta(\theta)_h, q_{\theta'}) - p_{\theta'} && \text{(2.B.44)}\\
&= v_{\theta_h}(q_{\theta'}) - p_{\theta'} && \text{(2.B.45)}
\end{aligned}
$$

where we used the local IC, the induction hypothesis and monotonicity. Hence all constraints of (RP) are implied and hence satisfied at the solution to (DRP). △

Lemma 2.7. *In the optimal contract of the relaxed problem the IC from any revealed type* θ *to the closest lower revealed type* $\eta(\theta)$ *is active.*

Proof. As the relaxed problem and the doubly relaxed problem are equivalent, it is sufficient to show that local downward IC between revealed types are active in the doubly relaxed problem. Suppose towards a contradiction that one of them is not active, say from type θ to $\eta(\theta)$. Suppose we increase the price in the contract of all revealed types greater than θ including θ by some $\epsilon > 0$. Note that this change doesn't affect any constraints between the affected types. Furthermore, θ isn't the lowest revealed type, hence the participation constraint of all revealed type is implied by the IC and not active since IC-$\theta \to \eta(\theta)$) isn't active. As we can pick epsilon sufficiently small, this IC is still slack and we strictly increased revenue, contradiction the optimality of the initial contract. △

Lemma 2.8. *In the optimal contract of the relaxed problem,* $q_\theta \leqslant \widehat{q}_{\theta_h}$ *for all* $\theta \in \Theta$.

Proof. As the relaxed problem and the doubly relaxed problem are equivalent, we can work on the doubly relaxed problem. The result follows from Proposition 2.4 for concealed types. Suppose towards a contradiction that this property is violated for some subset of revealed types. Pick the smallest revealed type for which this is the case and denote it as θ. Note that $q_\eta(\theta) \leqslant \widehat{q}_{\eta(\theta)_h} < \widehat{q}_{\theta_h}$ and denote the rent given to type θ as $\Delta := v(\theta^h, q_{\eta(\theta)}) - p_{\eta(\theta)}$. (This is the correct expression, because the local downward IC is active by the above Lemma.) Consider the set of contracts where we replaced the initial contract for type θ by $\left(\widehat{q}_{\theta_h}, v_{\theta_h}(\widehat{q}_{\theta_h}) - \Delta\right)$. As θ receives the same utility in both contracts, no participation constraint is violated and all IC from α are still satisfied. The upward IC $\eta(\theta) \to \theta$ is still satisfied as it is implied by R-monotonicity (which is maintained) and the corresponding downward IC. Consider any higher type imitating θ. The amended contract gives the same utility to θ at a lower quality, hence it gives a strictly lower deviation payoff to higher types. In particular, all IC are satisfied. The revised contract is also more profitable for the principal as the most profitable way to transfer rent to type θ in frame h is using quality $\widehat{q}_{\theta_h}$. Hence, the initial set of contracts wasn't optimal. $\triangle$

Now we can show that there are no profitable feasible upward deviations in e^*. We proceed by induction. Order the types such that $\{\theta^1, \dots, \theta^n\} = \Theta$, $\theta^i < \theta^{i+1}$. Clearly, the highest type has no feasible upward deviations. Suppose all upward deviations are either infeasible or unprofitable for types θ^i into types θ^j for $j > i > m$. We need to show that the required upward IC constraints out of type θ^m are satisfied. We will proceed case by case, in addition showing that the upward IC from concealed to revealed types are always slack:

1. *Deviations into a concealed type with rent $\Delta_{\theta^i} \leqslant v(\theta^i_h, \widehat{q}_{\theta^i_h}) - v(\theta^i_l, \widehat{q}_{\theta^i_h})$:* Then the participation constraint of type θ^i is binding at the intermediate stage in frame l. But by single crossing

$$c_{\theta^i} \sim_{\theta^i_l} 0 \implies c_{\theta^i} \prec_{\theta^m_l} 0, \tag{2.B.46}$$

an imitation is infeasible.

2. *Deviations into a concealed type with rent* $\Delta_{\theta^i} > v(\theta^i_h, \widehat{q}_{\theta^i_h}) - v(\theta^i_l, \widehat{q}_{\theta^i_h})$*:* Note that in this case $q_{\theta^i} = \widehat{q}_{\theta^i_h}$ and this rent has to be the result of a possible deviation that is discouraged by a constraint of the problem and hence by the induction hypothesis this is a downward deviation into a revealed type. Hence $\Delta_{\theta^i} = v(\theta^i_h, q_\eta) - p_\eta$ for some $\eta < \theta^i$, $\eta \in \Theta_R$. But then the upward deviation isn't profitable unless the deviation into η is profitable, since $q_\eta \leqslant \widehat{q}_{\eta_h} \leqslant \widehat{q}_{\theta^i_h} = q_{\theta^i}$ and by single crossing

$$c_\eta \sim_{\theta^i_h} c_{\theta^i} \implies c_\eta \succ_{\theta^m_h} c_{\theta^i} \tag{2.B.47}$$

so all we have to show is that deviations into revealed types are not profitable. If $\eta < \theta^m$, this is achieved already by the maintained IC constraints, if $\theta^m \in \Theta_R$ it is by the upward IC. The case we need to consider are deviations from concealed types upwards into revealed types.

3. *Deviations from a concealed into a revealed type:* Consider a concealed type θ^m with a profitable upwards deviation into a revealed type. As the set of types is finite, there has to exist a lowest revealed type into which θ^m has a strictly profitable deviation. Furthermore, since we impose downward incentive compatibility constraints, this lowest target type has to be greater than θ^m. We will show that such a lower bound cannot exist, hence there can be no profitable upward deviation.

   Suppose such a lower bound exists, $\underline{\theta} = \min\{\theta \in \Theta_R | \theta^m_h q_\theta - p_\theta > \theta^m_h q_{\theta^m} - p_{\theta^m}\}$. But then, consider type $\eta(\underline{\theta})$. A deviation into this type is also strictly profitable since $c_{\eta(\underline{\theta})} \sim_{\underline{\theta}_h} c_{\underline{\theta}}$ and by R-monotonicity $q_{\eta(\underline{\theta})} \leqslant q_{\underline{\theta}}$, but then by single crossing $c_{\eta(\underline{\theta})} \succcurlyeq_{\theta^m_h} c_{\underline{\theta}} \succ_{\theta^m_h} c_{\theta^m}$, contradicting the minimality of $\underline{\theta}$. Hence there can be no strictly profitable upward deviation.

And we established that there can be no upward deviation by type θ^m. By induction, no type prefers any attainable contract offered to higher types in e^* and hence we found an EDP that attains the upper bound to the solution of (GP) and therefore (RP)=(GP). □

Proof of Proposition 2.3 on page 81: Let $\mathbf{c} = (c_\theta)_\theta = ((p_\theta, q_\theta))_\theta$ be an optimal vector of contracts implemented by some EDP. By Theorem 2.1, we can construct a standard EDP e with that implements it. Let Θ_C and Θ_R be the sets of revealed and concealed types in e. If $\theta \in \Theta_C$, the statement follows from Proposition 2.4. We proved that $q_\theta < \widehat{q_{\theta_h}}$ as Lemma 2.8. Therefore there is only one case left to consuder. Assume that $\theta \in \Theta_R$ and towards a contradiction that $q_\theta < \underline{q}_\theta$, where $\underline{q}_\theta$ satisfies $\zeta_{\theta_h}(\underline{q}_\theta) = \zeta_{\theta_l}(\widehat{q}_{\theta_l})$. Denote the rent in this contract by $\Delta := v_{\theta_h}(q_\theta) - p_\theta$.

We will construct a vector of contracts with strictly higher revenue. Starting from the old EDP, we now conceal type θ and set the contract $(\widehat{q_{\theta_l}}, \widehat{p_{\theta_l}} - \Delta)$. Using the surplus function ζ_θ defined in (2.A.12), note that since $q_\theta < \bar{q}_\theta$

$$\zeta_{\theta_h}(q_\theta) < \zeta_{\theta_l}(\widehat{q}_{\theta_l}) \tag{2.B.48}$$

and consequently

$$\zeta_{\theta_h}(q_\theta) - \Delta < \zeta_{\theta_l}(\widehat{q}_{\theta_l}) - \Delta \tag{2.B.49}$$

$$p_\theta - \kappa(q_\theta) < \widehat{p_{\theta_l}} - \Delta - \kappa(\widehat{q_{\theta_l}}) \tag{2.B.50}$$

and the principal receives weakly higher profit in the modified contract.

Clearly, this contract satisfies the participation constraint in frame l and delivers rent greater than Δ to type θ in the high frame, hence there is no deviation by this type. There is no downward deviation into this contract since the type is concealed. Furthermore, we don't have to worry about upward deviations. The optimal concealed contract – which is strictly better for profit – is never subject to them and we will establish that even a sub-optimal

concealed contract delivers an improvement in profits. Hence the original vector was not optimal, a contradiction. □

Proof of Proposition 2.4 on page 82: Note that there are no IC constraints into a type $\theta \in \Theta_C$. Hence we can separate the principals problem and solve for the optimal contract of θ in (RP). The contract given to type θ solves

$$\begin{aligned} \max_{(p,q)} \; & p - \kappa(q) \\ \text{s.t. } & v_{\theta_l}(q) - p \geqslant 0 \\ & v_{\theta_h}(q) - p \geqslant \Delta \end{aligned} \tag{2.B.51}$$

Dropping the second constraint, the optimal contract is $\widehat{c}_{\theta_l}$, which delivers rent $v_{\theta_h}(\widehat{q}_{\theta_l}) - v_{\theta_l}(\widehat{q}_{\theta_l})$, hence the second constraint is satisfied if

$$\Delta \leqslant \left[v_{\theta_h}(\widehat{q}_{\theta_l}) - v_{\theta_l}(\widehat{q}_{\theta_l}) \right]. \tag{2.B.52}$$

Similarly, note that the optimal contract dropping the first constraint is

$$\left(v_{\theta_h}(\widehat{q}_{\theta_h}) - \Delta, \widehat{q}_{\theta_h} \right), \tag{2.B.53}$$

which gives utility $v_{\theta_l}(\widehat{q}_{\theta_h}) - v_{\theta_h}(\widehat{q}_{\theta_h}) + \Delta$ in the low frame. Hence the first constraint is satisfied if

$$\Delta \geqslant v_{\theta_h}(\widehat{q}_{\theta_h}) - v_{\theta_l}(\widehat{q}_{\theta_h}). \tag{2.B.54}$$

In the intermediate case, both constraints are binding,

$$v_{\theta_l}(q^\star) = p \tag{2.B.55}$$

$$v_{\theta_h}(q^\star) - v_{\theta_l}(q^\star) = \Delta \tag{2.B.56}$$

and the optimal contract is $\left(v_{\theta_l}(q^\star), q^\star\right)$. Note that $q^\star \in (\widehat{q}_{\theta_l}, \widehat{q}_{\theta_h})$ by single crossing. □

Proof of Proposition 2.5 on page 84: Take any type $\theta \in \Theta$. For each μ, consider (RP) with the constraint $\theta \in \Theta_C$ ($\theta \in \Theta_R$) and denote the corresponding optimal value by $\Pi^R_{C;\mu}$ ($\Pi^R_{R;\mu}$). Next, using the surplus function ζ_{θ_f} defined in (2.A.12), we can bound those values as

$$\Pi^R_{R;\mu} \geqslant \mu_\theta \zeta_{\theta_h}(\widehat{q}_{\theta_h}) \tag{2.B.57}$$

$$\Pi^R_{C;\mu} \leqslant \mu_\theta \zeta_{\theta_l}(\widehat{q}_{\theta_l}) + \sum_{\theta' \neq \theta} \mu_{\theta'} \zeta_{\theta'_h}(\widehat{q}_{\theta'_h}) \leqslant \mu_\theta \zeta_{\theta_l}(\widehat{q}_{\theta_l}) + (1-\mu_\theta)\zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h}), \tag{2.B.58}$$

where $\tilde{\theta} := \max\{\Theta \setminus \theta\}$.

Note that Lemma 2.4 implies that

$$\zeta_{\theta_h}(\widehat{q}_{\theta_h}) > \zeta_{\theta_h}(\widehat{q}_{\theta_l}) > \zeta_{\theta_l}(\widehat{q}_{\theta_l}), \tag{2.B.59}$$

and define

$$\bar{\mu}_\theta := \frac{\zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h})}{\zeta_{\theta_h}(\widehat{q}_{\theta_h}) - \zeta_{\theta_l}(\widehat{q}_{\theta_l}) + \zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h})} \in (0,1). \tag{2.B.60}$$

Finally, combining (2.B.57), (2.B.58) and (2.B.59) yields

$$\Pi^R_{R;\mu} - \Pi^R_{C;\mu} \geqslant \mu_\theta \zeta_{\theta_h}(\widehat{q}_{\theta_h}) - \mu_\theta \zeta_{\theta_l}(\widehat{q}_{\theta_l}) - (1-\mu_\theta)\zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h}) \tag{2.B.61}$$

$$= \mu_\theta \left[\zeta_{\theta_h}(\widehat{q}_{\theta_h}) - \zeta_{\theta_l}(\widehat{q}_{\theta_l}) + \zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h})\right] - \zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h}) \tag{2.B.62}$$

$$\geqslant 0. \tag{2.B.63}$$

Therefore, for any $\mu_\theta \in [\bar{\mu}_\theta, 1]$, it is optimal to reveal θ.[31] $\square$

Proof of Proposition 2.6 on page 84: It is easy to see that the optimal contract and set of concealed types before the change of valuations is still feasible after the change. Hence $\Pi^*_\Theta \leqslant \Pi^*_{\tilde{\Theta}}$. If $\tilde{\theta}$ is concealed in the optimum, we are done. Suppose that instead it is not concealed. Then, since $\tilde{\theta}_l$ is the only difference to the initial problem and doesn't affect the constraints unless $\tilde{\theta}$ is concealed, the optimal contract under $\tilde{\Theta}$ is feasible under Θ and $\Pi^*_{\tilde{\Theta}} \leqslant \Pi^*_\Theta$. Hence, the original vector of contracts is still optimal, establishing the claim. $\square$

Proof of Proposition 2.7 on page 85: Consider the profit from concealing all types except the highest, $\Pi_C(\epsilon) := \sum_{\theta<\bar{\theta}} \mu_\theta v_{\theta_l}(\widehat{q}_{\theta_l}) + \mu_{\bar{\theta}} v_{\theta^-_h}(\widehat{q}_{\theta^-_h})$, where we consider the θ_l as a function of ϵ. It is easy to see that $\widehat{q}_{\theta_l} \to \widehat{q}_{\theta_h}$ as $\theta_l \to \theta_h$. By continuity of v, $\Pi_C(\epsilon) \to \Pi((\widehat{c}_{\theta_h})_{\theta\in\Theta})$.

Suppose that in the optimum $\theta < \bar{\theta}$ is revealed. Then $q_\theta > \underline{q_\theta} > 0$ by Proposition 2.3 and $p_\theta \leqslant v_{\theta_h}(q_\theta)$. But then, by the incentive compatibility constraint of $\bar{\theta}$, $\Pi < \Pi((\widehat{c}_{\theta_h})_{\theta\in\Theta}) - \mu_{\bar{\theta}}(v_{\bar{\theta}_h}(q_\theta) - v_{\theta_h}(q_\theta)) < \Pi((\widehat{c}_{\theta_h})_{\theta\in\Theta})$. Hence, there exists an $\epsilon_\theta > 0$ such that $\Pi_C(\epsilon) > \Pi$ for $\epsilon < \epsilon_\theta$, so it cannot have been optimal to reveal θ for sufficiently small ϵ. The result follows by taking the maximum over $\{\epsilon_\theta : \theta \in \Theta \setminus \bar{\theta}\}$. $\square$

Proof of Theorem 2.3 on page 91: Let e_0 denote an EDP constructed for sophisticated types in Theorem 2.1. Order naive types $\Theta_N = \{\theta^1, \ldots, \theta^m\}$ with $\theta^i < \theta^{i+1}$. We will construct an optimal EDP for the mixed case inductively.

Starting from $e_0 = (E_0, h)$, we add one continuation problem at the root for every naive type,

$$e_{n+1} = \left(\bigcup_{i=0}^{n+1} E_i, h \right). \tag{2.B.64}$$

[31] This bound is typically not tight, as we introduced slack in (2.B.58).

To define E_i, let the most preferred alternative in e_{i-1} for type θ^i be $x_i := \operatorname{argmax}_{\mathcal{C}(e_{i-1})} u_{\theta^i_h}$. During the construction, we ensure that

1. no sophisticated type prefers to continue to E_i,
2. no naive type θ^j with $j < i$ prefers to continue to E_i, and
3. type θ^i indeed proceeds to E_i and chooses $\widehat{c}_{\theta^i_h}$ eventually.

If we ensure this during our construction, all sophisticated types choose as in e_0 and all naive types choose their efficient contract $\widehat{c}_{\theta^i_h}$ and we establish the theorem.

Let $E_i = \left\{ \left(\left\{ \left(\{N_i, \{d^{\theta'}_{N,i}\}_{\theta' > \theta^i : \theta' \in \Theta_S}, 0\}, h \right), \left(\{\widehat{c}_{\theta^i_h}, \{d^{\theta'}_i\}_{\theta' > \theta^i : \theta' \in \Theta_S}, 0\}, h \right), 0 \right\}, l \right) \right\}$.

We now have to specify N_i and the decoys and verify 1-3 above. First, use the mapping from Lemma 2.5 to set $N_i := \mathcal{D}(x, (\theta_{i-1}, \theta^i))$ so that

$$N_i \sim_{\theta^i_h} x_i \tag{2.B.65}$$

$$q_{N_i} \geqslant q_{x_i} \tag{2.B.66}$$

$$N_i \preccurlyeq_{\theta^i_l} 0 \tag{2.B.67}$$

Second, define the decoys as

$$(d^{\theta'}_{N,i})_{\theta' > \theta^i : \theta' \in \Theta_S} := \mathcal{D}(N_i, (\theta^i, \Theta_S^{>\theta^i})), \tag{2.B.68}$$

$$(d^{\theta'}_i)_{\theta' > \theta^i : \theta' \in \Theta_S} := \mathcal{D}(\widehat{c}_{\theta^i_h}, (\theta^i, \Theta_S^{>\theta^i})), \tag{2.B.69}$$

where $\Theta_S^{>\theta^i}$ is a vector of types in Θ_S that are strictly greater than θ^i. By construction, every sophisticated type $\theta > \theta^i$ prefers the outside option to the contract chosen from the continuation problems. Hence they have no incentive to enter. Furthermore, all contracts are, by construction, worse in frame l than the outside option for all types $\theta < \theta^i$, hence lower sophisticated types have no incentive to enter. Hence, we established 1.

By construction, E_i contains a most preferred option for θ^i_h, hence continuing into E_i is part of a naive outcome for θ^i. At the subsequent decision node, the decision problem containing N_i is as attractive as the outside option: By the construction of the decoys, $N_i \succcurlyeq_{\theta^I_h} d^{\theta'}_{N,i}$ and $q_{N_i} \leqslant q_{d^{\theta'}_{N,i}}$ and hence $N_i \succcurlyeq_{\theta^i_l} d^{\theta'}_{N,i}$. But $N_i \preccurlyeq_{\theta^i_l} 0$. As the decision problem containing $\widehat{c}_{\theta^i_h}$ also contains the outside option, continuing to this menu is part of a naive outcome. From the menu $\left(\{\widehat{c}_{\theta^i_h}, \{d^{\theta'}_i\}_{\theta' > \theta^i : \theta' \in \Theta_S}, 0\}, h\right)$, the DM chooses $\widehat{c}_{\theta^i_h}$ by the construction of the decoys. This establishes 3.

To see 2, note that all decoys have higher quality than N_i and $\widehat{c}_{\theta^i_h}$, respectively, and are less preferred according to θ^i_h. Hence, they are less preferred by lower naive types θ^j_h by single crossing. Furthermore, $\widehat{c}_{\theta^i_h}$ is not attractive to lower naive types, as it is worse than the outside option. It remains to check whether N_i is attractive. But note that $N_i \sim_{\theta^i_h} x_i$ and $q_{N_i} \geqslant q_{x_i}$ imply $N_i \preccurlyeq_{\theta^j_h} x_i$ for all $j < i$. By the induction hypothesis, $N_j \succcurlyeq_{\theta^j_h} \operatorname{argmax}_{\mathcal{C}(e_{i-1})} u_{\theta^j_h} \succcurlyeq_{\theta^j_h} x_i \succcurlyeq_{\theta^j_h} N_i$. Consequently, N_i is not attractive to lower naive types, and there is a naive outcome where types $\theta^j < \theta^i$ choose E_j.

Clearly, the contract implemented for naive types is optimal given the participation constraint in the high frame any implemented contract needs to satisfy. Furthermore, suppose there is an EDP implementing contracts for sophisticated types that are not implemented by an optimal EDP in Theorem 2.1. Then the contracts don't solve (RP), so we can find a strictly better set of contracts and use the above construction. Hence every optimal EDP in (2.4.5) satisfies Theorem 2.3. From that, the decomposition theorem is immediate.

If there is an ex-post participation constraint, any naive outcome needs to satisfy $v(\theta) \succcurlyeq_{\theta_n} 0$. The revenue maximal vector of contracts satisfying these contraints is $(\widehat{c}_{\theta_n})_{\theta \in \Theta_N}$. It is immediate from the proof of Theorem 2.3 that this set of contracts can be implemented using an analogous construction.

The outside option trivially satisfies the participation constraint in every frame, so continuing at the root is always part of a valid naive strategy in the interim modification of any extensive-form decision problem. Hence, $N^e \subseteq N^{\bar{e}}$, which establishes the observation. □

Proof of Remark 2.3 on page 92: Let us denote the contract for type θ in the sophisticated problem as c^s_θ and note that the contract in the naive problem is $\widehat{c}_{\theta_h}$. Note that $c^s_\theta \succcurlyeq_{\theta_h} 0 \sim \widehat{c}_{\theta_h}$ and $q^s_\theta \leqslant \widehat{q}_{\theta_h}$. Hence by single crossing $c^s_\theta \succcurlyeq_{\theta_f} \widehat{c}_{\theta_h}$, strictly for $f \neq h$ if $c^s_\theta \neq \widehat{c}_{\theta_h}$. □

Proof of Remark 2.4 on page 94: To implement the vector of contracts $(\widehat{c}_{\theta_n})_{\theta\in\Theta}$, the principal can simply conceal *all* types using neutral frame n.

Notice that $(\widehat{c}_{\theta_n})_{\theta\in\Theta}$ satisfies all the constraints of (RP) for $\Theta_R = \varnothing, \Theta_C = \Theta$. Therefore, by Theorem 2.2, there exists a standard EDP e^* that implements it. Notice that since the contract $\widehat{c}_{\theta_n}$ for type θ satisfy P^n_θ, the interim and ex-post modifications $\bar{e}^*$ and $\underline{e}^*$ also implement $(\widehat{c}_{\theta_n})_{\theta\in\Theta}$. □

Proof of Remark 2.5 on page 95: First, consider the ex-post modification. To implement $(\widehat{c}_{\theta_n})_{\theta\in\Theta}$, the principal can the same construction as in Theorem 2.3, but with n. Notice that because any contract that is implemented with ex-post participation constraints must satisfy those constraints, the principal cannot do better.

Second, consider the interim modification. Suppose that the optimal EDP without the modification is e^* and notice that e^* implements $(\widehat{c}_{\theta_h})_{\theta\in\Theta}$. Now consider its interim modification $\bar{e}^*$. Since naive consumers think they would get a better option than 0, they would proceed to e^*. Therefore, there exists a naive solution $\bar{v}$ to $\bar{e}^*$, such that $\bar{v}_\theta = \widehat{c}_{\theta_h}$ for all θ. □

Chapter 3

Hidden Costs: Anticompetitive Effects of Transparency[1]

3.1 Introduction

In many markets consumers are faced with the choice of a base good which then leads to follow-up fees or purchases from the same firm. Printers require ink cartridges, hotel guest buy WiFi and drinks from the minibar, bank accounts have surcharges for overdrafts and some transactions, and phone plans come with overage charges and roaming fees. Casual empiricism and a growing literature[2] show that some consumers fail to anticipate these additional, hidden costs. Firms, in turn, actively hide the true price of their products and exploit consumers mistakes.

At the same time, some firms increase price transparency by voluntarily disclosing information about potentially hidden prices and even reminding consumers of them in

[1]I am grateful for valuable comments and suggestions to Roland Bénabou, Botond Kőszegi, Denis Shishkin and seminar audiences at Princeton University.

[2]In the credit card market, Ausubel (1991) notes excessive profits despite a high number of nearly identical firms and attributes them partially to consumer naïveté. Stango and Zinman (2009) find that many consumers pay large easily avoidable fees for banking services. Consumers are inattentive to their past usage of their cell phone plan and misjudge the variance of future use, leading to excessive overage (Grubb and Osborne, 2015). Mortgage advertising focuses on teaser rates and is associated with higher prices (Gurun et al., 2016).

advertising campaigns. Instead of hiding ink costs, manufacturers of inkjet printers started selling ink subscriptions and heavily advertise these services (including price information) in their information material or sell printers bundled with large amounts of ink.[3] Phone plans are often 'unlimited', without overage charges or roaming fees. Firms also sometimes support consumer protection legislation and industry agreements that make it harder to exploit consumers unawareness of fees and costs.[4] Why do firms in some industries enter into legal agreements to be transparent while they invent new ways to trick consumers in others?

To resolve this transparency puzzle, we provide a general explanation of profit and loss from the exploitation of consumer naïveté. Two firms compete in a horizontally differentiated market. Firms set transparent prices for a base good and receive a follow-up hidden price b (receive a surprise fee or sell an add-on) from some of their consumers. A share α of consumers is naïve. Naïve consumers do not anticipate the fee while sophisticated consumers avoid it. A key feature of our model is that the share of naïve consumers on average can be different from the share among those at the margin between the two firms. This is the case if naïve and sophisticated consumer differ in their cross-price elasticity of demand or, equivalently, the variance of the idiosyncratic taste shocks. Both the positive predictions and the impact of transparency on consumer surplus depend on this difference in responsiveness to price changes as a sufficient statistic. We show that exploiting naïveté

[3]HP launched an ink subscription service 'Instant Ink' for businesses in 2014 and quickly expanded this service to its range of consumer printers (HP, 2014, 2015). At the time of writing, all but two consumer printers are compatible with this service. Banners advertising 'Instant Ink' are also present on the overview page introducing the line of printers that is not fully compatible. The service is marketed both as an increase in convenience and a decrease in ink costs. Clearly, it makes future inks costs salient for any consumer looking at information about HP printers. Epson's "Eco Range" is the first range of printers advertised on the companies website. The main selling point of this line are lower ink costs. This feature is further emphasized as these printers are bundled with large amounts of ink. The first two out of four bullet points in the description of the cheapest printer of this range read "Ink equivalent to about 20 ink cartridge sets" and "Save up to 80% on ink with low-cost replacement bottles".

[4]Almost all wireless service providers are signatories of the Consumer Code for Wireless Service sponsored by the industry trade group CTIA (CTIA, 2011). This code requires providing consumers with free notifications for plan usage and roaming fees. Timely warning messages can substitute for consumer sophistication. For a detailed analysis of bill-shock regulation see (Grubb and Osborne, 2015; Grubb, 2015). See also FN 18.

is profitable if and only if naïve consumers have more dispersed firm-specific taste shocks than sophisticated consumers.

To understand this result, consider the two effects of naïveté on profits. On the one hand, naïve consumers pay an additional fee $\mathfrak{b}$ to the firm, so the firm receives additional revenue proportional to the average share of naïfs among its consumers. On the other hand, attracting a consumer is also more profitable proportionally to the share of naïfs at the margin between the two firms and competition induces firms to hand these gains back to consumers through a lower transparent price. The first effect dominates and exploitation is profitable if naïfs are underrepresented at the margin.

We then allow firms to increase price transparency (Section 3.4). In applications, increased transparency can take many forms, such as advertisement campaigns containing information about formerly hidden fees, general awareness campaigns, labeling products or lobbying for consumer protection legislation. Incentives to increase transparency depend on the timing of transparency relative to price setting and on which groups of consumers firms can reach. Increasing transparency can be profitable if prices react. This *competitive channel* is crucial: Transparency is profitable only insofar it acts a tool to soften competition in the industry. In line with the literature, increasing transparency is never profitable when prices cannot react to the new market situation.

The targeting of transparency matters. Transparency that affects all consumers equally, e.g. awareness campaigns or a coordinated commitment to transparency (through regulation, industry standards or tacit collusion), is profitable as long as exploiting consumer naïveté is not. While in the short run, both gains and losses in profits due to naïveté are equally plausible, exploitation of consumer naïveté cannot persist unless it is profitable. Transparency targeted at consumers groups that are unlikely to switch brands in response to small price changes is typically not profitable. Equilibrium prices stay unchanged and revenues from hidden prices decrease as fewer consumers are naïve. Transparency targeted at groups that are very responsive to price changes, on the other hand, is typically profitable.

Reducing the share of naïfs among marginal consumers reduces the revenue from cutting transparent prices and hence softens competition sufficiently to outweigh the direct loss of fee revenue. Hence, firms desire to be transparent towards such undecided consumers, while keeping captured consumers in the dark.

The welfare considerations in this paper focus on the the distribution of surplus and the consumer participation margin. In the main sections, we consider a fully covered market in order to focus on the competitive margin, where our mechanism operates. In this benchmark case, transparency has no impact on overall welfare by construction, but it crucially affects the distribution of surplus. In particular, we show that we cannot rely on firms to introduce transparency when it is good for consumers. Firms' incentives to be transparent are strongest exactly when transparency reduces consumer surplus. We introduce an outside option for consumers in Section 3.5.1. Then, the visible component of the price is typically below marginal cost, inducing over-participation. Transparency often increases welfare, but firm-induced transparency still reduces consumer surplus unless it has a sufficient (negative) effect on market size.

In Subsection 3.5.2, we solve the model when transparency changes the demand elasticity of affected naïve consumers and show that our main results carry over under fairly general conditions. We discuss the implications of our model for different industries, several assumptions and the measurement of the key statistics identified by our analysis in Section 3.6. All proofs are in Appendix 3.A.

Related Literature

Consumer unawareness of fees and add-on prices has received considerable attention in the literature. For a recent survey, see Heidhues and Kőszegi (2018).

Our treatment of profits and losses from exploiting consumer naïveté reconciles Ellison (2005) and Gabaix and Laibson (2006). In Ellison (2005), hidden add-on prices increase profits because they create an adverse selection problem in the market for the base good:

Most consumers attracted by a price reduction are unprofitable "cheap skates", softening competition in the market. If firms can unilaterally advertise their add-on price, however, they choose to do so. Hidden add-on prices are profitable, but not a chosen equilibrium outcome. Gabaix and Laibson (2006) study naïve consumers who fail to anticipate add-on prices. Firms don't desire to be transparent, but all gains from naïveté are competed away in the market for the base good. Hidden add-on prices are an equilibrium outcome, but not more profitable than transparency. We combine the two insights. In line with Gabaix and Laibson (2006) firms don't desire to price transparently in the short run. There are gains from hidden fees when marginal consumers are less profitable than inframarginal consumers as in Ellison (2005). We use the model to study the long run level and welfare impact of price transparency.

Following Gabaix and Laibson (2006), in most models of markets with naïveté about follow-up fees industry profits are unaffected by consumer naïveté, but there are several exceptions. Heidhues et al. (2016) find positive profits because of a binding lower bound on the transparent price. This is very natural in the case of bank and credit card fees, as interest rates and annual fees are bounded below to avoid arbitrage gains by savvy consumers and we often observe free base goods in such industries. In Miao (2010) a similar bound arises endogenously in a durable goods model. Johnen (forthcoming) considers the case of repeated firm consumer interactions where past usage data allows firms to infer whether a return consumer is naïve or sophisticated. This allows firms to make targeted offers to return naïfs. As a result, price cutting is more likely to attract less profitable sophisticated consumers, softening competition. The reason we see profits from exploiting naïveté in our model is similar, in that profits depend on the share of naïve consumers attracted on the margin. By contrast to these contributions, our mechanism doesn't rely on a constrained transparent price or firm information and targeted discount offers towards naïve return customers.

The modal result in the literature with naïve consumers is that increasing price transparency is not profitable (Gabaix and Laibson, 2006; Grubb, 2015; Johnen, forthcoming; Miao, 2010).[5] Most of these results are based on *short run* "inform and undercut" type deviations. This result also holds in our model: If we fix the competitor's price, informing consumers is not profitable. For the main results of our paper, we take a *long run* perspective, treating consumer naïveté as a state variable of the market that firms affect in order to improve their competitive position. With this timing, Dahremöller (2013) shows that transparency is profitable for one of the firms if they differ in their costs of providing the add-on good. Li et al. (2014) analyze the effect of the timing of disclosure relative to pricing on firms' incentives to disclose adverse product attributes such as health hazards, also with firm heterogeneity. In contrast to our model, allowing prices to respond to a reduction of naïveté reduces incentives to inform consumers. This is because in our framework per-consumer revenue is reduced ceteris paribus as naïveté decreases, creating incentives to increase prices as a response, while in their setting, increased consumer sophistication decreases demand, creating incentives to lower prices. Our mechanism relies on differences in demand elasticities between naïve and sophisticated consumers instead of heterogeneity in the efficiency of nontransparent business practices between otherwise symmetric firms.

We find that regulation requiring transparency can increase profits and reduce consumer surplus. This is typically the case if firms individually have an incentive to support the regulation. We document a new reason to lobby for regulation and an associated anticompetitive effect of consumer protection regulation.[6] Grubb (2015) finds that bill-shock regulation

[5] Again, a notable exception are Heidhues et al. (2016, 2017). They combine Bertrand competition and a binding lower bound on the transparent component of the price. Informing consumers enables firms to undercut using the hidden fee, which is unconstrained. As a result, informing consumers and thereby unlocking an opportunity to undercut is profitable as long as the market is not completely eliminated as a result of informing (because the good cannot be sold above cost to sophisticated consumers as it is socially wasteful) and the number of firms is sufficiently high.

[6] The literature identified anticompetitive effects of regulation through increased barriers to entry (See for example Ryan (2012) for an empirical analysis of the impact of amendments of the US Clean Air Act on entry, profits and consumer surplus in the cement industry.) or an increased severity of search frictions (Armstrong et al. (2009) show that price regulation decreases consumer incentives to become informed in a consumer search framework and can backfire, increasing average prices.).

requiring the disclosure of past usage to consumers can decrease consumer welfare if firms price discriminate between high and low demand consumers. Kosfeld and Schüwer (2017) show that partial transparency can decrease consumer surplus and welfare, because prices increase and sophisticated consumers still bear an avoidance costs. We show consumer surplus can decrease even if there are no avoidance costs and transparency is complete, and that this is the typical case if firms choose transparency. While transparency often reduces consumer surplus, it is typically welfare improving in our model, as we focus on the participation distortion.[7]

3.2 The Model

We consider a market with two competing firms indexed A and B, respectively. Firm i offers a base good (bank account) at transparent price p_i. We normalize production costs to zero.

There is a unit mass of consumers with two types $\theta \in \{N, S\}$. A share α of consumers is naïve ($\theta = N$), a share $1-\alpha$ is sophisticated ($\theta = S$). Naïve consumers pay an unanticipated (overdraft) fee $b > 0$ to the firm providing them with the base good. Sophisticated consumers avoid paying this fee. (They are mindful of their account balance.) Consumers have unit demand and purchase the product that gives them the higher surplus $v_i - p_i$ where v_i is a consumer specific taste shock from an atomless distribution with full support. For now, we assume that there is no outside option, every consumer needs to purchase from one of the two firms. We set $v_A \equiv 0$ without loss of generality and drop the subscript on v_B. We allow the distribution of the taste shocks F_θ to differ between naïve and sophisticated

[7] Other distortions from the exploitation of naïveté can arise, depending on the precise mistake naïve consumers are making, e.g. through wasteful avoidance of the fee (Gabaix and Laibson, 2006), incentives for socially wasteful innovation in improvements in the extent of exploitation (Heidhues et al., 2016) and overborrowing **?**. Heidhues and Kőszegi (2017) discuss a wide range of examples and study the impact of naïveté-based discrimination on welfare.

consumers. Let $F = \alpha F_N + (1-\alpha)F_S$ denote the distribution of the average consumer and $\alpha(v) = \frac{\alpha f_N(v)}{f(v)}$ the share of naïve consumer among consumers with taste shock v.

Assumption 3.1. *The median consumer is indifferent between the two firms,* $F(0) = \frac{1}{2}$*, the average distribution* F *(and* $1-F$*) is log-concave, and the share of naïve consumers at a given taste shock doesn't change too rapidly,* $|\alpha'(v)b| < 1$.[8]

We say that tastes of naïve consumers are more dispersed if $f_N(0) < f_S(0)$.[9] Rewriting the expression for $\alpha(0)$ shows that $\alpha(0) < \alpha$ if and only if the tastes of naïve consumers are more dispersed.

$$\alpha(0) = \alpha + \alpha(1-\alpha)\frac{f_N(0) - f_S(0)}{f(0)} \tag{3.2.1}$$

Both more and less dispersed tastes can be the result of naïveté. In the case of costly avoidance of fees, idiosyncratic differences in avoidance cost result in sophisticated consumers having more dispersed tastes. Take the example of ATM fees. For a sophisticated consumer taking these fees into account when picking a bank the distance to the nearest ATM usable without paying a fee is more relevant than for a naïve consumers ignoring ATM fees. This creates additional taste variation. If sophisticated consumers avoid the fee by using a firm's service less, as in the case of overage charges, the effect of idiosyncratic tastes, say the quality of the local cell network, is less important, creating lower variation in tastes. Differences in the variability of tastes can also be the results of correlates of sophistication, such as experience. Sophisticated consumers of exchangeable lens cameras both foresee future lens purchases and are likely to have an installed base of lenses compatible with one manufacturer's system, making their tastes more dispersed. In other situations, sophisticated consumers are more likely to see through product differentiation in attributes that are irrelevant in actual use, making their tastes less dispersed.

[8] We assume equal market shares for concreteness, as our mechanism is driven by differences between naïve and sophisticated consumers, not firm asymmetries. The other two conditions ensure that the problem has an interior equilibrium.

[9] Note that Assumption 3.1 implies that conditions for $v = 0$ hold in a neighborhood of 0 that is large relative to the hidden fee b.

Firms simultaneously pick a price for their base good. Profits are given by

$$\begin{aligned}\pi_A(p_A,p_B) &= \alpha(p_A+b)\int_{-\infty}^{p_B-p_A} dF_N(v) + (1-\alpha)p_i\int_{-\infty}^{p_B-p_A} dF_S(v)\\ &= \int_{-\infty}^{p_B-p_A} (p_A+\alpha(v)b)\, dF(v), \end{aligned} \tag{3.2.2}$$

and analogously for firm B. Let the average share of naïve consumers with taste shock smaller/larger than zero be denoted by $\alpha_A = 2\int_{-\infty}^{0}\alpha(v)\,dF(v)$ and $\alpha_B = 2\int_{0}^{\infty}\alpha(v)\,dF(v)$ and let $\alpha_0 = \alpha(0)$. We say that we are in the symmetric case if $\alpha_A = \alpha_B = \alpha$ (but note that this does not imply that the game is symmetric).

Several situations with naïve consumers map into the above model. The simplest that will be used throughout in the discussion is that naïve consumers pay a "surprise" fee, while sophisticated consumers foresee the fee and avoid it altogether.[10] Alternatively, both naïve and sophisticated consumers may foresee the fee and expect to be able to avoid it, but naïve consumers fail to avoid it ex-post due to inattention, temptation or unexpected taste shocks.[11]

3.3 Static Equilibrium

We now characterize the equilibrium. Denote the inverse semi-elasticity of demand when the two firms share the market equally by $\mu = \frac{F(0)}{f(0)} = \frac{1}{2}\frac{1}{\alpha f_N(0)+(1-\alpha)f_S(0)}$. To ensure the existence of an equilibrium in which both firms are active, we assume that $\mu + (\alpha_i - \alpha_0)\,b \geqslant 0$. This assumption implies that equilibrium profits are positive as it puts a lower bound on the losses from naïveté relative to market power μ. In the symmetric case, this condition is satisfied if and only if tastes of naïfs are not much less dispersed than tastes of sophisticated

[10] As avoidance is costless, sophisticates and naïfs have identical tastes over bundles and firms cannot screen by consumer sophistication. We briefly discuss costly avoidance and screening in Section 1.5. For a model of screening based on naïveté, see Eliaz and Spiegler (2008, 2006).

[11] See e.g. Grubb (2015) who analyzes bill shock regulation when naïve consumers are inattentive of their past consumption and incur overdraft fees or overage charges, but fail to anticipate their inattention.

consumers:

$$f_N(0) \leqslant f_S(0) + \frac{1}{2\alpha(1-\alpha)b} \tag{3.3.1}$$

Even though firms can differ in their share of naïve consumers α_i, the equilibrium is symmetric as the price is set at the margin between the two firms.

Proposition 3.1. *The unique equilibrium is given by*

$$p_i^\star = \mu - \alpha_0 b \tag{3.3.2}$$

$$\pi_i^\star = \frac{\mu + (\alpha_i - \alpha_0)\, b}{2}, \tag{3.3.3}$$

where α_i is the share of naïfs among consumers choosing firm i.

Firm i's profits π_i^* are increasing in hidden fees charged b if and only if

$$\alpha_i - \alpha_0 \geqslant 0. \tag{3.3.4}$$

In the symmetric case, this holds if and only if the tastes of naïve consumers are more dispersed than the tastes of sophisticated consumers ($f_N(0) < f_S(0)$).

The exploitation of consumer naïveté has two impacts on profits. First, profits increase through the direct effect of fee revenue from naïve consumers. This effect is proportional to the share of naïve consumers among the customers of the firm. Second, consumer naïveté makes it more attractive to attract consumers, which affects the equilibrium price. Prices are set to equate per consumer profit at the margin to the inverse semi-elasticity of demand. When attracting an additional consumer, a firm earns not only the base good price, but also hidden fees in proportion to the share of naïve consumers among marginal consumers, α_0. In equilibrium, prices have to adjust for this ex-post transfer from consumers to firms. This indirect effect depresses the base good price and firms set low margins (or engage in loss-leading for sufficiently small μ). Whether firms gain or lose profits because of the presence of naïve consumers overall depends on the relationship between the share

of naïve consumers inframarginally and on the margin. In the benchmark case with an equal elasticity for naïve and sophisticated consumers (and no firm preferences of either type, $F_\theta(0) = \frac{1}{2}$), the average and marginal share of naïfs is the same ($\alpha_0 = \alpha$). The base good is fully discounted for gains from hidden fees and profits are unaffected by consumer naïveté.[12] If naïfs have less elastic demand, the share of naïve consumers is lower at the margin than on average ($\alpha_0 < \alpha$), gains from naïveté are competed away only partially and firm profits are higher in the presence of hidden fees. If naïfs have more elastic demand, the share of naïve consumers is higher at the margin then on average ($\alpha_0 > \alpha$), the increased intensity of competition outweighs inframarginal gains from naïveté and firm profits are lower in the presence of hidden fees.

To see these forces directly as a result of the different taste distributions, we can rewrite profits in the symmetric case as

$$\pi_i^\star = \frac{1}{2}\frac{\frac{1}{2} + \alpha(1-\alpha)(f_S(0) - f_N(0))b}{\alpha f_N(0) + (1-\alpha)f_S(0)} \tag{3.3.5}$$

The revenue from hidden fees remaining after "reimbursement" through lower ex-ante prices is proportional to $f_S - f_N$. This statistic and its equivalent form (in the symmetric case $\alpha_A = \alpha_B$)

$$\alpha - \alpha_0 = \frac{\alpha(1-\alpha)}{f(0)}(f_S - f_N)$$

will appear throughout our discussions. In terms of observables, it is closely related to the cross-price elasticities of demand between the goods of the two firms. Denote the absolute value of the demand elasticity of type θ for the good of firm A with respect to the price

[12] This case corresponds to the shrouding equilibrium in Gabaix and Laibson (2006).

charged by firm B by ε_θ, then:

$$\varepsilon_\theta = \frac{p}{F_\theta(0)} f_\theta(0) \tag{3.3.6}$$

$$f_S - f_N = \frac{\varepsilon_S F_S(0) - \varepsilon_N F_N(0)}{p} \tag{3.3.7}$$

Hence, firms profit from the exploitation of consumer naïveté if the demand of naïve consumers is less elastic than the demand of sophisticated consumers. If naïve consumers are more elastic, the marginal consumer attracted by a price change is overproportionally likely to be naïve, which depresses prices and profits.

3.3.1 The Distributional Impact of Consumer Naïveté

In the baseline model, the whole market is covered, so there can be no over- or under-consumption. Furthermore, there are no avoidance costs or costs of exploitation. Hence, there are no sources of inefficiency. Still, the exploitation of consumer naïveté impacts the distribution of surplus between firms and consumers.

With hidden fees, the base price is discounted. Firms may engage in loss-leading, pricing below marginal cost, to attract more consumers from which a fee can be extracted. This leads to a transfer from naïve consumers who pay a hidden price to sophisticated consumers who benefit from lower transparent prices while avoiding the hidden price. Consumer surplus is given by

$$CS = (1-\alpha)\underbrace{\left(\mathbb{E}_{F_S}[v\chi_{\{v\geqslant 0\}}] - (\mu - \alpha_0 b)\right)}_{\text{CS sophisticated}} + \alpha\underbrace{\left(\mathbb{E}_{F_N}[v\chi_{\{v\geqslant 0\}}] - (\mu - \alpha_0 b) - b\right)}_{\text{CS naive}} \tag{3.3.8}$$

This transfer averages out, i.e. sophisticates gain as much as naïfs lose if and only if $\alpha_0 - \alpha = 0$, i.e. if the taste shocks of naïfs are equally dispersed as those of sophisticated consumers. If the taste shocks of naïfs are more dispersed, there is only partial reimbursement of revenue from the hidden price as competition is soft, the remainder going into firm

profits. Conversely, if taste shocks of naïfs are less dispersed, consumers on average are reimbursed for more than hidden price revenue, the additional funds drawing from reduced firm profits, increasing consumer surplus.

3.4 Price Transparency

In this section we analyze firms' incentive to label hidden fees transparently, inform or debias naïve consumers and keep consumers informed about past usage and account balances. Translated into our model, all these actions effectively decrease the share of naïve consumers in the population.

There are two effects of changes in naïveté on profits. On the one hand, a lower share of naïve consumers decreases fees collected from consumers, reducing profits if we hold prices fixed. On the other hand, a reduced share of naïve consumers softens competition, increasing equilibrium prices and profits. The sign of the combined effect depends on how strongly marginal consumers are affected relative to inframarginal consumers. The goal of firms is to soften competitors' price setting by reducing the share of naïfs among marginal consumers while keeping inframarginal consumers naïve and hence profitable.

First, we analyze firms' incentives for price transparency in the short run, as the competitor's price remains fixed. This represents situations where price adjustment is slow relative to the introduction of more transparent business practices and is considered mainly as a benchmark. As in Gabaix and Laibson (2006)[13], transparency is never profitable in this case. We then allow prices to react to the share of naïve consumers. Transparency changes the competitive environment and can be profitable through its effect on equilibrium prices.

Throughout, we focus on profits from exploitation of naïveté and incentives to inform naïve consumers resulting from the exploitation of naïveté. Hence, we assume that there

[13]There are cases where shrouding is not an equilibrium in Gabaix and Laibson (2006), but they rely on strictly positive avoidance costs for sophisticated consumers. In our parameter specification, there is always a shrouded equilibrium and Proposition 3.2 generalizes this no-unshrouding result.

are no changes to the average density of consumers at zero, i.e. the average elasticity of demand.

Assumption 3.2. *The average distributions of taste shocks* F *is unchanged by increases in transparency.*

This assumption is satisfied if the link between demand elasticities and naïveté is not causal but merely statistical. Suppose inexperienced consumers are both more likely to be naïve and differ in their taste shock distribution. In this case, making the price structure of the product more transparent reduces the share of consumers who act naïve about hidden prices, but doesn't affect the underlying tastes. The assumption is violated if transparency changes the very basis of the difference in taste shock distributions, e.g. when the difference stems from avoidance of the fee, as in the case of out-of-system ATM charges. We analyze the model when an increase in transparency changes the taste shock distribution in Subsection 3.5.2. The main results identified in this section carry over to that case under weak conditions.

3.4.1 Short-run Incentives for Transparency

First, we analyze incentives to increase transparency when the competitor's price cannot react to changes in the extend of consumer naïveté. Consider the above price setting game augmented with a transparency technology that allows firms to inform naïve consumers, thus effectively turning them into sophisticated consumers. In addition, firms can choose to charge lower hidden fees or no hidden fees at all.

Formally, firms can choose a level of price transparency $t_i \in [0,1]$ and a fee $b_i \in [0,b]$ in addition to prices. Given transparency choices t_A, t_B, the share of consumers with taste shock ν acting as naïfs is given by $\alpha(\nu, t_A, t_B)$.

Assumption 3.3. *The share of naïve consumers* $\alpha(\nu, t_A, t_B)$ *is* $\mathcal{C}^2$ *and satisfies*

1. *Normalization:* $\alpha(\nu, 0, 0) = \alpha(\nu)$,

2. *Decreasing Naïveté:* $\frac{\partial \alpha}{\partial t_A}, \frac{\partial \alpha}{\partial t_B} \leqslant 0$,

3. *Nontriviality: For all* $t_i \in [0,1]$ *and* $\nu \geqslant 0$,

$$\frac{\partial \alpha}{\partial t_B}(\nu, t_A, t_B) < 0 \text{ and } \frac{\partial \alpha}{\partial t_A}(-\nu, t_A, t_B) < 0.$$

4. *Concavity:* $\alpha_i(t_i, t_{-i}) - \alpha_0(t_i, t_{-i})$ *is strictly concave in* t_i *for all* t_{-i}.

The first condition ensures that we nest the static model if there is no increase in transparency. The second condition implies that more transparency by either firm cannot increase the share of consumers acting naïvely. The third condition ensures that each firm can always reach some of its own consumers by its increase in transparency. This assumption rules out trivially transparent equilibria in weakly dominated strategies, where firms are transparent only because their transparency has no impact on their own profit. The final condition is assumed for ease of exposition and allows us to focus on pure strategy equilibria.

Transparency is costless and profits are given by

$$\pi_i(p_i, p_{-i}) = \left[p_A \int_{-\infty}^{p_B - p_A} dF(\nu) + b_A \int_{-\infty}^{p_B - p_A} \alpha(\nu, t_A, t_B)\, dF(\nu) \right], \tag{3.4.1}$$

and similarly for firm B. Consumers observe prices before they make their choice, so there is no scope to use transparency to signal.[14]

Proposition 3.2. *In the game with simultaneous price setting, fee choice and informing, the unique equilibrium is* $(p_A^\star, 0, b)$, $(p_B^\star, 0, b)$.

In line with the literature, keeping naïve consumers uninformed about hidden prices and exploiting naïveté as much as possible is the unique equilibrium. Consumers are not informed and both firms charge the maximal fee. The intuition for this result is simple.

[14] In particular, remaining naïve consumers are by definition those that are not made aware of the hidden fee by the increase in transparency.

For a given market share, reducing naïveté or fees strictly decreases profits. Furthermore, a firm can never profitably increase its market share by informing consumers, as the market share only depends on the transparent price. Since both per consumer profit as well as market share cannot increase as a consequence of informing consumers, it cannot be a profitable deviation from the equilibrium in 3.1. Note that we made no assumptions about the difference in elasticities between types or the targeting of transparency. Short-run transparency isn't profitable for arbitrary elasticities and even if it can be arbitrarily tailored to a consumers taste shock ν.

3.4.2 Long-run Transparency: Consumer Naïveté as a State Variable

In this section, we treat transparency and hence consumer naïveté as a state variable of the competitive environment, allowing firm prices to respond to its current level. Recall the comparative statics of the equilibrium in Proposition 3.1: Prices fully adjust for fees gained from the marginal consumer. Profits are decreasing in the share of naïve consumers at the margin but are increasing in the average share of naïve consumers. This creates a role for transparency: Even though informing consumers decreases inframarginal gains from naïveté, a firm will have an incentive to inform consumers if doing so sufficiently softens competition by making it less profitable to attract the marginal consumer.

To formalize this intuition, consider the following two stage game. In the first stage, both firms can increase transparency, thereby effectively decreasing the share of naïve consumers as described above. Having observed the resulting level of naïveté, the two firms compete in the market as described in Section 3.2. Consumers make demand decisions after observing prices as in the static model. As before, they have complete information about the variables they deem decision relevant, so there is no inference based on the level of transparency.

Firms can influence the state variable $\alpha(\nu)$ using a transparency technology. Under Assumption 3.2, the impact of transparency on profit is characterized by its effect on

marginal naïveté α_0 and average naïveté of consumers of each firm α_i. We denote the transparency choice of firm i by $t_i \in [0,1]$ and these outcomes by

$$\alpha_0(t_A, t_B) = \alpha(0, t_A, t_B) \tag{3.4.2}$$

$$\alpha_A(t_A, t_B) = \int_{-\infty}^{0} \alpha(v, t_A, t_B)\, dF(v) \tag{3.4.3}$$

$$\alpha_B(t_A, t_B) = \int_{0}^{\infty} \alpha(v, t_A, t_B)\, dF(v) \tag{3.4.4}$$

where $\alpha(v, t_A, t_B)$ satisfies Assumption 3.3. To ensure the existence of an interior solution in the second stage, we assume that $\mu + (\alpha_i(t_A, t_B) - \alpha_0(t_A, t_B))\, b \geqslant 0$ for all $t_A, t_B \in [0,1]^2$. We allow for costs of increasing transparency $\frac{1}{2}c(t_i)$ with $c' \geqslant 0$, $c'' > 0$, where the factor $\frac{1}{2}$ renormalizes cost to the market share.

Proposition 3.3. *The set of SPE is characterized by*

$$t_i^\star \in \arg\max_{t \in [0,1]} \frac{\mu + \left(\alpha_i(t_i^\star, t_{-i}^\star) - \alpha_0(t_i^\star, t_{-i}^\star)\right) b - c(t^\star)}{2} \tag{3.4.5}$$

$$p_i^\star(t) = \mu - \alpha_0(t) b \tag{3.4.6}$$

The Competitive Effect of Transparency

In the previous section, we saw that transparency is never an equilibrium outcome in the absence of an effect on the competitive environment. We now isolate the impact of transparency purely through this channel. Suppose increased transparency only targets marginal consumers and set costs of transparency to zero.[15] In this case, increased transparency has no direct effect on profits, it only affects profits through the competitive channel. It is easy to see that

$$\frac{\partial \pi_A}{\partial t_A} = -\left(\frac{\partial \pi_A}{\partial p_A} + \frac{\partial \pi_A}{\partial p_B}\right) \frac{\partial \alpha_0}{\partial t_A} b \tag{3.4.7}$$

[15] Clearly, this doesn't fit the fully specified model in Section 3.2, as we restricted attention to continuous distributions, but it is a useful corner case to consider.

Increased transparency acts on second stage profits as a concerted price increase by both firms.[16] In the simple benchmark model, consumers do not have an outside option and such a price increase is always profitable. Transparency (effectively) commits not only the firm instituting it to higher prices (as could be replicated more directly with advertisements, for instance), but it allows a single firm to influence the strategic environment and thereby commit both firm to softer competition. This competitive channel is crucial: A monopolist would never increase transparency.

3.4.3 Long-run Transparency: Broad Campaigns

We discussed the effects of transparency in two polar cases: A pure direct effect when firms engage in short-run transparency (Section 3.4.1) which is never profitable and a pure indirect effect through the equilibrium price if transparency only affects the marginal consumer, which is always profitable. All feasible transparency increases will combine these two channels to varying degrees. A simple benchmark case is an increase in transparency that is directed equally at the whole population of consumers. Examples are advertisement campaigns containing information on hidden fees, lobbying effort towards regulation limiting exploitation of consumer naïveté and self-regulation through industry standards an trade associations. A joint decrease in $\mathfrak{b}$ for both firms (say, through tighter regulation) has the same effect as this form of transparency increase. We show that when such a transparency technology is available, firms either gain from naïveté or they remove it by increasing transparency. We then show that increases in price transparency are not as benevolent as they may seem. Firms' incentives for price transparency and its impact on consumer surplus

[16] Gomes and Tirole (2018) show that firms may want to price transparently to avoid a hold up problem if sophisticated consumers differ in their cost of avoiding the hidden fee. In our model, there is no hold-up problem as sophisticated consumers avoid the hidden fee costlessly and naïfs are not aware of the fee even at the final stage of their purchase decision. Instead, transparency effectively serves as a commitment between firms to softer competition and higher industry prices by making the marginal consumer less profitable to attract.

are opposed: Firms choose higher transparency whenever it reduces consumer surplus and refrain from transparency whenever it consumer surplus would increase.

To model that all consumers are reached equally, we assume that the relative reduction in the share of effectively naïve consumers is the same in all three categories.

Assumption 3.4 (Equal Proportions)**.** *The transparency technology satisfies*

$$\frac{\alpha_0(t_A,t_B)}{\alpha_0} = \frac{\alpha_A(t_A,t_B)}{\alpha_A} = \frac{\alpha_B(t_A,t_B)}{\alpha_B} \tag{3.4.8}$$

for all $t_A, t_B \in [0,1]^2$.

We maintain the basic conditions on transparency in Assumption 3.3 and that $\mu + (\alpha_i(t_A,t_B) - \alpha_0(t_A,t_B))\, b \geqslant 0$ for all $t_A, t_B \in [0,1]^2$. To rule out corner cases, we assume that the initial increase in transparency is free locally.

Assumption 3.5 (Locally Free Transparency)**.** *The transparency technology satisfies* $c'(0) = 0$.

Let us first discuss the symmetric case ($\alpha_A = \alpha_B$). We can rewrite profits taking into account the pricing game equilibrium as

$$\pi_i = \left[\mu + \alpha_0(t_i, t_{-i}) \left(\frac{\alpha}{\alpha_0} - 1\right) b\right] - c(t_i) \tag{3.4.9}$$

Note that except for the cost of transparency the game is a common interest game in the transparency stage. Setting $c \equiv 0$, increasing transparency increases profits if

$$\alpha_A - \alpha_0 < 0, \tag{3.4.10}$$

i.e. if and only if the tastes of naïve consumers are less dispersed. Firms have an incentive to increase transparency (i.e. to jointly limit their exploitation of consumer naïveté) if and only if exploiting consumer naïveté is not profitable.

This result generalizes when we relax symmetry: In the general case, the first order condition for an interior solution is given by

$$\frac{\partial \alpha_0(t_A^\star, t_B^\star)}{\partial t_i}\left(\frac{\alpha_i}{\alpha_0} - 1\right) b = c'(t_i^\star) \tag{3.4.11}$$

It is easy to see that transparency is only attractive if the firm loses from naïveté. The equilibrium amount of transparency is the smallest amount (possibly 0) ensuring that no firm loses from exploiting naïveté in the pricing stage sufficiently to outweigh costs from increasing transparency (or every firm losing from naïveté engages in maximal transparency, $t_i^\star = 1$).

Proposition 3.4. *Firm i does not engage in transparency ($t_i^\star = 0$) in any SPE if and only if exploiting consumer naïveté is profitable for it ($\alpha_i > \alpha_0$). Conversely, suppose exploiting consumer naïveté is not profitable for firm i ($\alpha_i < \alpha_0$). Then, in every SPE, it engages in transparency ($t_i^\star > 0$). Suppose there are no direct costs of transparency ($c \equiv 0$). Then firm i does engage in full transparency ($t_i^\star = 1$) in any SPE.*

In any stable SPE, transparency by firm i is decreasing in its share of captured naïve consumers ($\frac{\partial t_i^\star}{\partial \alpha_i} < 0$).

In the static setting, both gains and losses in profits due to naïveté were equally plausible, depending on the specifics of the market. When firms can increase price transparency, either through awareness campaigns or coordinated commitment (through regulation, industry standards and agreements or tacit collusion) not to exploit consumers, exploitation of consumer naïveté cannot persist unless it is profitable or costly to remove and self-regulate against.[17]

[17] Equivalently, inappropriable innovation in exploitation technologies (increasing b) by an individual firm is profitable only if transparency is unprofitable and vice versa.

The Distribution of Surplus

Recall that we can decompose the impact of the exploitation of consumer naïveté on consumer surplus into the loss of naïve and the gain of sophisticated consumers. Taking the derivative of CS with respect to transparency yields

$$\frac{\partial CS}{\partial t_i} = \frac{\partial \alpha_0}{\partial t_i}\left(1 - \frac{\alpha}{\alpha_0}\right) \tag{3.4.12}$$

Firms only increase transparency if $\alpha_i < \alpha_0$, that is when increased transparency reduces consumer surplus obtained by consumers of firm i. The very reason firms increase transparency is to reduce pass-through more than fee revenue, i.e. reduce consumer surplus.

Proposition 3.5. *Consider the symmetric case. Firms engage in transparency if and only if higher transparency lowers consumer surplus.*

Increased transparency has a heterogeneous impact across consumer types. Sophisticated consumers always lose as do naïve consumers who still buy the product after the transparency increase. Both act the same way irrespective of transparency but pay a higher transparent price for the same service. Naïve consumers who are turned sophisticated gain. They still buy the product, but are now able to avoid the hidden fee thanks to higher price transparency.

3.4.4 Targeting Transparency

The degree to which transparency has the effect desired by firms (i.e. softens competition), depends on who it targets. Transparency can be directed mainly at a firm's own consumers, such as labeling. Other forms of transparency apply across all consumers, such as industry agreements precluding the exploitation of naïveté. Firms can also try and disabuse their competitor's likely consumers, for example through comparative advertisements pointing out the competitor's nontransparent business practices.

Firms' transparency choice trades off the loss of revenue from exploiting inframarginal naïve consumers with the gain associated with the increase in price through the competitive channel. The loss in fee revenue is proportional to the change in α_i, the gain is proportional to the change in α_0. Increasing transparency is individually profitable as long as

$$\left|\frac{\partial \alpha_i(t,t)}{\partial t_i}\right| + \frac{c'(t_i)}{b} \leqslant \left|\frac{\partial \alpha_0(t,t)}{\partial t_i}\right| \tag{3.4.13}$$

Higher transparency is only desirable as long as the direct loss of fee revenue (the first term on the left) and direct costs of transparency (the second term on the right) are smaller than the gain in profits through pushing up the equilibrium price . This is the case as long as the impact on inframarginal consumers $\left|\frac{\partial \alpha_i(t,t)}{\partial t_A}\right|$ is sufficiently small compared to the impact on marginal consumers $\left|\frac{\partial \alpha_0(t,t)}{\partial t_A}\right|$.

Many realistic options for informing and debiasing consumers reach mainly consumers with a preference for the informing firm. Labeling, advertisement, employee advice and reminder messages tend to target consumers who look at the product, watch a firm's ads, visit the store or at the most extreme have already purchased the product. Firms engage in these forms of transparency as long as undecided consumers are affected to a sufficient degree. On the one extreme, providing information about hidden fees to comparison websites used by undecided consumers is typically profitable. On the other extreme, providing fee warnings before retained customers incur fees is not profitable.[18] When transparency reduces naïveté equally across the distribution, we are back in the previous section and it is only desirable for firms if they loose from the exploitation of naïve consumers via hidden fees.

[18]If firms can credibly commit to such warnings, they act as a debiasing tool and can be profitable by softening competition if firms jointly commit to such a policy. Such a joint commitment is in place in the US mobile phone market. Note that reminders can also serve other purposes and might then be designed as to help naïve consumers as little as possible and this may indeed be the case in the specific example. They can come too late for naïve consumers, as naïve consumers may overindulge in low value usage of their plan early in the month and then unexpectedly but rationally incur overage. At some phone companies "[t]he alert email sent out when you hit 100% of your data usage can push you over your limit and trigger the $15 charge, even if you turn off your data immediately upon notification," (Wolff-Mann, 2016) suggesting a less charitable interpretation.

Notably, firms ignore their effect on the inframarginal consumers of their competitor. This suggests that advertisement campaigns directed at a rival's consumer base pointing out their nontransparent business practices are often profitable. Sufficiently strong spillovers of transparency to the competitor's consumer base can create a prisoner's dilemma situation between firms, with both firms increasing transparency to soften competition but jointly eroding naïveté in the market, even when firms profit from its presence. Such transparency increases aimed at a competitors' consumer base may be an effective way of reaching marginal consumers without spoiling one's own. There are limits to this strategy, however, as firms within one market tend to exploit consumer naïveté in similar ways.

The Distribution of Surplus

An increase in transparency increases consumer surplus if and only if

$$\frac{1}{2}\left(\left|\frac{\partial \alpha_i(t_A,t_B)}{\partial t_i}\right|+\left|\frac{\partial \alpha_{-i}(t_A,t_B)}{\partial t_i}\right|\right) \geqslant \left|\frac{\partial \alpha_0(t_A,t_B)}{\partial t_i}\right|, \tag{3.4.14}$$

i.e. when the impact on average naïveté is larger (in absolute value) than at the margin. Note that firm's incentives to increase transparency are greatest when $\left|\frac{\partial \alpha_0(t_A,t_B)}{\partial t_i}\right|$ is large, whereas transparency is desirable for consumers when $\left|\frac{\partial \alpha_0(t_A,t_B)}{\partial t_i}\right|$ is small. This is the transfer channel creating the stark conflict of interest in the previous section. Now, there is an additional channel: By increasing transparency firms inadvertently also affect their competitor ($\left|\frac{\partial \alpha_{-i}(t_A,t_B)}{\partial t_i}\right|$). This effect doesn't enter the firm's own decision problem, but the resulting decrease in inframarginal naïveté is beneficial for consumers. Hence, transparency chosen by firm i may be beneficial to consumers, as long as there are considerable spillovers of increased transparency at firm i to effective naïveté at firm $-i$. Consequently, it is regrettable from a consumer surplus perspective that many realistic options for informing and debiasing consumers reach mainly consumers with a preference for the informing firm.

3.5 Extensions

3.5.1 Extensive Margin

An important inefficiency in markets with hidden costs is consumer overparticipation.[19] To analyze the welfare consequences of transparency, we add a participation decision to the consumer problem. We show that, while transparency is often welfare improving, still firms tend to engage in it when it reduces consumer surplus. Our positive predictions are modified by a participation term, but qualitatively unchanged as long as participation is not highly responsive to an increase in prices.

The Model and Static Analysis

There is an outside option available to consumers before they know their firm specific taste shock ν and firm prices. Consumers draw a random utility of participation $u \sim G_\theta$, dependent on the consumer type, and enter the market if

$$u + \mathbb{E}_{F_\theta}[\max\{-p_A, \nu - p_B\}] \geqslant 0 \tag{3.5.1}$$

where prices are the expected equilibrium prices. We assume that G_θ is continuously differentiable with increasing hazard rate. In a symmetric equilibrium with price p, the share of consumers of type θ entering is given by

$$1 - G_\theta\left(p - \mathbb{E}_{F_\theta}[\nu \chi_{\{\nu \geqslant 0\}}]\right), \tag{3.5.2}$$

where χ denotes the indicator function. Denote by $G(p) = \alpha G_N(p - \mathbb{E}_{F_N}[\nu\chi_{\{\nu\geqslant 0\}}]) + (1-\alpha) G_S(p - \mathbb{E}_{F_S}[\nu\chi_{\{\nu\geqslant 0\}}])$ the average participation shock distribution, assuming a symmetric

[19] See e.g. Heidhues and Kőszegi (2015) who suggest that the US credit card market has a welfare cost due to overparticipation of up to 61% of revenue.

pricing equilibrium. We assume that the share of naïve consumers who enter the market is independent of the price p.

Assumption 3.6. *The distributions of taste shocks* F_N, F_S, G_N, G_S *satisfy*

$$G_N\left(p - \mathbb{E}_{F_N}[v\chi_{\{v\geqslant 0\}}]\right) = G_S\left(p - \mathbb{E}_{F_S}[v\chi_{\{v\geqslant 0\}}]\right) \qquad \forall p \in \mathbb{R}. \tag{3.5.3}$$

Whether exploiting consumer naïveté is profitable or not is determined by the relationship of naïveté at the margin and among captured consumers as before. If naïveté is higher among captured consumers, it is always profitable. There is an additional effect through the participation channel. It is easiest to see the new effect for $\alpha_i = \alpha_0$. Then, an increase in hidden fees decreases the transparent price but is neutral for profits for a fixed market size. But, since the size of the market depends on the transparent price, the market expands. Hence exploiting consumer naïveté is profitable even if the fees are competed away fully in the transparent price, as it inflates consumer participation.

Proposition 3.6. *Firm* i*'s profits* π_i *are increasing in hidden fees* b *if and only if*

$$\frac{\alpha_i - \alpha_0}{\alpha_0} \geqslant -\frac{g(\mu - \alpha_0 b)}{1 - G(\mu - \alpha_0 b)}\left[\mu + (\alpha_i - \alpha_0)\, b\right]. \tag{3.5.4}$$

This condition is always satisfied if $\alpha_i > \alpha_0$ *or, equivalently in the symmetric case, if the tastes of naïve consumers are more dispersed than the tastes of sophisticated consumers (*$f_N(0) < f_S(0)$*).*

Increasing Transparency

As before, short-run transparency (Prop. 3.2) is unprofitable in this setting. Transparency is only attractive because it changes the competitive environment. By lowering naïveté at the margin, both firms effectively commit to a higher price. When consumers can opt-out of the market, this is only profitable as long as the initial price is below the cartel price.

Proposition 3.7. *Let $\bar{p}$ denote the price chosen by a cartel with commitment.*[20] *With symmetric firms ($\alpha_A = \alpha_B$), profits are decreasing in α_0 if and only if*

$$\bar{p} = \frac{1 - G(\bar{p})}{g(\bar{p})} - \alpha b > p^\star = \mu - \alpha_0 b. \tag{3.5.5}$$

Our results about the broad campaigns transparency technology generalize. Firms keep exploiting consumer naïveté essentially if and only if it is profitable.

Proposition 3.8. *Firm i does not engage in transparency ($t_i^\star = 0$) in any SPE if exploiting consumer naïveté is profitable for it at $t = (0,0)$ (in the sense of Prop. 3.6: $\frac{\partial \pi_i^\star}{\partial b} \geqslant 0$).*

Conversely, suppose exploiting consumer naïveté is not profitable for at least one of the firms at $t = (0,0)$ (in the sense of Prop. 3.6: $\frac{\partial \pi_i^\star}{\partial b} < 0$). Then, in every equilibrium, at least one of the firms engages in transparency.

In any stable SPE, transparency by firm i is decreasing in its share of captured naïve consumers ($\frac{\partial t_i^\star}{\partial \alpha_i} < 0$).

As before, firms desire to target their transparency at marginal consumers and there is a conflict of interest in terms of consumer surplus.

Welfare

In the efficient allocation, consumers enter the market if and only if their expected valuation is greater than marginal cost (which we normalized to zero). With hidden fees, the price is discounted as firms may engage in loss-leading to attract more consumers from which a fee can be extracted. Thus there may be overparticipation by consumers.[21]

[20] Furthermore, we restrict the cartel to charge the same price at both firms. This is to ensure that even for asymmetric prices, the share of naïve consumers in the market in independent of prices. A sufficiently strong generalization of 3.6 to get this as a result is inconsistent with $F_N \neq F_S$. But note that we use the cartel solution only as an illustration of the incentives for the firms in oligopoly.

[21] The exploitation of consumer naïveté can result in additional distortions, e.g. through avoidance cost or inefficient contract design. Depending on the specifics of the market, these distortions are ameliorated by transparency, e.g. when contracts are more efficient in the presence of fewer naïve consumers, or made worse by transparency, e.g. when a larger share of sophisticated consumers implies greater total avoidance costs. For a general discussion of the welfare consequences of such distortions and how changes in the share of naïve consumers and degree of price discrimination affects welfare, see Heidhues and Kőszegi (2017).

The shortfall in total welfare compared to the efficient benchmark is given by

$$\Delta W = -\left| \int_{\mu - \alpha_0 b}^{0} s \, dG(s) \right|. \tag{3.5.6}$$

Consumers overparticipate whenever $\alpha_0 > \frac{\mu}{b}$, i.e. if the transparent price is below cost. Even though a small share of naïve consumers can counteract underparticipation due to market power, casual empiricism suggests that overparticipation is the usual case in many industries with hidden fees.[22] This implied by our model for industries in which surprise fees are a large share of revenue (e.g. credit cards), as overparticipation is required to square high fee revenue with reasonable variable profits.[23]

Let us rewrite consumer surplus as

$$CS = \underbrace{\int_{\mu}^{\infty} u - \mu \, dG(u)}_{\text{CS with all } S} + \underbrace{(1 - G(\mu - \alpha_0 b))(\alpha_0 - \alpha) b}_{\text{transfer term}} + \underbrace{\int_{\mu - \alpha_0 b}^{\mu} u - \mu \, dG(u)}_{\text{participation term}} \tag{3.5.7}$$

In the benchmark model, transparency affected consumer surplus only through the transfer term. Now we also have a participation term which lowers consumer surplus. To see this, note that at the margin, both types are indifferent, but naïve consumers will pay a surprise fee. Hence, entering the market causes negative expected consumer surplus averaged over both types. This overparticipation from a consumer surplus perspective is always present, even when participation is efficient from a welfare point of view ($\mu = \alpha_0 b$). This is because consumers underestimate the price they will pay on average. The gain in efficiency happens not because firms charge the efficient price on average, but because they can hide part of

[22] A market with exploitation of naïve consumer can replicate the socially efficient allocation if $\alpha_0 = \frac{\mu}{b}$. This result crucially depends on the absence of distortions from exploiting naïveté except overparticipation, e.g. the assumption of free avoidance of the add-on by sophisticated consumers. Furthermore, there still is a transfer from naïve to sophisticated consumers, which may negatively affect utilitarian social welfare if naïve consumers are poorer on average when quasi-linearity is relaxed.

[23] Per-consumer variable profit is given by $p + \alpha_i b$. It is lower than fee revenue ($p + \alpha_i b < \alpha_i b$) only if there is overparticipation ($p < 0$).

the price. In the efficient case, the hidden part is exactly the oligopoly markup, restoring efficiency at a cost of consumer surplus.

Proposition 3.9. *Suppose the market pre-transparency had lower social welfare than the benchmark case without consumer naïveté (or, equivalently, without exploitation of consumer naïveté,* $b = 0$*). Then, welfare (excluding transparency costs) is weakly higher under any SPE of the transparency game than in the static game.*

Transparency tends to increase total welfare. The impact on consumer surplus remains mixed. Consider broad campaigns transparency with symmetric firms. Taking the derivative of CS with respect to transparency yields

$$\frac{\partial CS}{\partial t_i} \propto \left(\frac{\alpha}{\alpha_0} - 1\right) + \frac{g}{1-G}\alpha b \tag{3.5.8}$$

The second term is the change due to overparticipation. This term is always positive, the loss in consumer surplus due to overparticipation decreases with increased transparency. Note that firms only increase transparency if $\alpha_i < \alpha_0$, hence the first term is negative whenever firms increase transparency in the symmetric case. The very reason firms increase transparency is to reduce pass-through more than fee revenue, i.e. reduce this term contributing to consumer surplus. Recall that without direct costs of transparency firms increase transparency if (rewriting 3.5.4)

$$\left(\frac{\alpha}{\alpha_0} - 1\right) + \frac{g}{1-G}\alpha b + \frac{g}{1-G}p < 0 \tag{3.5.9}$$

Theoretically, transparency can be both profitable and increase consumer surplus if it increases total welfare $(p < 0)$. This requires that the semi-elasticity of industry demand $\frac{g}{1-G}$ is sufficiently large. This is doubtful for bank accounts and phone plans, two prime

examples of markets with hidden costs, where few consumers are willing to stop demanding this type of service altogether.[24]

3.5.2 Transformative Transparency

In the main sections, we assumed than an increase in transparency (i.e. a decrease in the share of consumers acting as naïve consumers) nevertheless kept the average taste shock distribution F unchanged. This assumption is warranted if the difference in taste shock distributions is caused by third factors (e.g. inexperienced consumers are both more likely to be naïve and to have a more dispersed taste shock as they fall for spurious product differentiation). Another motivation for the difference in taste shock distributions, however, is the avoidance of the hidden price itself, e.g. in the case of out-of-system ATM chargers: Sophisticated consumers rely more on in-system ATMs in convenient locations, hence increasing the variance of idiosyncratic taste shocks. In this case, the assumption that increasing the share of consumers acting as sophisticates leaves F unchanged is more questionable. In this subsection, we solve the model when increased transparency also changes the taste shock distribution of affected consumers. We will only consider broad campaigns transparency and fully symmetric firms (f_N, f_S both symmetric around 0).[25]

Now, the taste shock distribution f is no longer invariant to transparency. We assume instead that

$$f_{t_A,t_B} = \alpha(t_A,t_B) f_N + [\alpha(0,0) - \alpha(t_A,t_B)] f_{NS} + [1-\alpha(0,0)] f_S$$

[24] An increase in transparency that is desired by firms and consumers also relies on our assumption that the market is viable under full transparency (as $\mu + (\alpha_i(1,1) - \alpha_0(1,1))\, b \geqslant 0$). There are strong firm incentives against transparency when it eliminates the market altogether, because the product is socially wasteful (either as such, or because there is a dominating alternative, see Heidhues et al., 2017).

[25] With other, targeted, forms of transparency, only part of the distribution of naïfs is affected. Hence, a more general transformation of f_N is required. This is beyond the scope of this section.

where f_{NS} denotes the taste shock distribution of "newly sophisticated" consumers, i.e. those that act sophisticated as a result of the increase in transparency. We assume that f_{NS} is also symmetric around 0 and $f_{NS}(0) \in [f_S(0), f_N(0)]$.

A firm increases transparency as long as

$$\frac{\partial}{\partial t_i}\pi_i = \frac{\partial}{\partial t_i}\left[\frac{1}{2f_{t_A,t_B}} + \left(\alpha(t_A,t_B) - \frac{\alpha(t_A,t_B)f_N}{f_{t_A,t_B}}\right)b\right] \geqslant 0$$

Assume for simplicity that transparency is costless ($c \equiv 0$). We showed in the main section that full transparency is the unique equilibrium if $f_N(0) > f_S(0)$ assuming f is unchanged by an increase in transparency ($f_{NS} = f_N$). This proposition generalizes under a weak condition on parameters and initial naïveté. In particular, the proposition generalizes independently of b if the share of naïve consumers is not too large ($1 \geqslant \alpha + \alpha_0$).

Proposition 3.10. *Assume* $1 + 2\frac{1-\alpha_0}{(1-\alpha)bf_S} \geqslant \alpha + \alpha_0$. *Both firms engage in full transparency if the exploitation of consumer naïveté is unprofitable* ($\alpha < \alpha_0$).

If exploiting naïveté is profitable, there is an additional force that may make transparency attractive to the firm. Sophisticated consumers have less elastic demand. Hence education naïfs decreases the average demand elasticity, increasing profits.

3.6 Discussion

In this section, we first apply our qualitative insights to explain firm-led transparency increases and their lack in our three example industries: banking, printers and phone plans. We then discuss several of our assumptions and how the model can be used to estimate the share of naive consumers on average and on the margin.

3.6.1 Transparency in Shrouded Markets

For consumer credit cards, there is no trend towards higher price transparency trough industry initiatives.[26] We add another explanation for this pattern to those in the literature.[27] Based on our model, we predict that consumer exploitation is stable in the long run, if naïve consumers are less elastic than sophisticates. There is evidence that this is the case in this market. Nelson (2018) shows that behaviors associated with naïveté (late payments less than 30 days and over-limit transactions) are a signal of inelastic demand. In banking services more generally, we also see very little transparency and debiasing. Even though we are not aware of direct evidence outside credit cards, this suggests that naïveté is associated with inelastic demand in these markets as well. One possible explanation is a limited awareness even of the transparent price by naïfs. Based on our welfare results, the very reluctance of firms to engage in transparency suggests that additional regulation mandating it would increase consumer surplus.

The market for printers and ink shows a more nuanced development illustrating our results on the targeting of transparency. Print costs are shrouded on in-store displays where they reach all consumers. Information material provided by the manufacturer contains an intermediate level of information: many general remarks about print costs and cartridge capacity but a suspicious lack of any concrete information on per page print costs. This information can be computed using standardized data readily provided in the fine print on manufacturer websites, but is only directly displayed and emphasized on third party websites targeted towards undecided consumers.[28] Transparency through such comparison

[26] According to the CFPB (2017) report on the consumer credit card market, "The cost of card credit remains largely stable [...] Cost structures—the allocation of costs to interest charges and fees—are also unchanged." There has been a marked decline in the number of debt suspension fees assessed, but other fees became more prevalent. The 2009 Credit CARD act regulated interest rates and fees and mandated additional consumer information, but was generally opposed by the industry. For an estimation of the effects of the CARD act, see Agarwal et al. (2015); Nelson (2018).

[27] E,g, Heidhues et al. (2016, 2017), who show that exploitation of consumer naïveté and even investment in inappropriable exploitative innovation is profitable, as the transparent price is bounded below and the product is dominated by nonexploitative alternatives, as is arguably the case in this market.

[28] HP, for example, advertises 'Instant Ink' heavily on their website, but doesn't mention it on the boxes of low-end printers that are displayed in many stores. The ink subscription service seems targeted towards

websites has a large impact on naïveté at the margin while keeping captured consumers in the dark. Undecided consumers are more likely to use these sites to compare products and prices, where they will find information about hidden prices and attributes and take them into consideration. Naïve consumers that have a firm opinion of their tastes for the competing products are unlikely to use these websites, remain unspoiled and fall into the trap of hidden fees.

Among phone plan providers, there has been a push towards more transparent pricing, including "unlimited" plans and industry self-regulation on overage. Such consumer protection legislation – whether by legislators or by the industry itself – is the clearest example for an increase in price transparency that affects all consumer equally: Mandating information that is easier to understand effectively decreases the share of naïve consumers. Equivalently, capping the hidden fee also maps to a decrease in the share of naïve consumers in our model. For phone plans, almost all wireless service providers committed to providing consumers with free notifications for plan usage and roaming fees. Our analysis suggests that we should be skeptical about the desirability of such initiatives.

3.6.2 Discussion of Assumptions

Several of our assumptions deserve additional discussion.

First, we assumed that sophisticated consumers can avoid the hidden fee at no cost. If there are avoidance costs for sophisticated consumers, there will be an unshrouding equilibrium, where firms set the fee to the avoidance cost and all consumers pay the fee to the firm, as long as sufficiently many consumers are sophisticated. This equilibrium is equivalent to the outcome without consumer naïveté in terms of profits. An additional issue raised by avoidance costs is screening. If firms can offer separate contracts and commit to a fee without educating naïve consumers through this very offer, they have an incentive to

undecided consumers who are arguable more likely to browse manufacturer websites, while avoiding to alarm decided consumers at the point of purchase of follow-up ink costs.

do so. A full analysis of this model is beyond the scope of this paper, but we conjecture that the difference in demand elasticities between sophisticated and naïve consumers continues to be an important determinant of the resulting contracts.[29]

Second, related to the first point, we assumed that there are no inefficiencies directly resulting from to the exploitation of consumer naïveté. Such inefficiencies can arise for sophisticated consumers, e.g. when they stem from avoiding the fee, or from the naïve side, e.g. when the contract is designed to induce naïve consumers to overborrow. Depending on the specifics of the market, these effects need to be taken into account for a complete welfare analysis of increased transparency.

Third, when analyzing the model with a consumer participation margin, we assumed that the share of naïve consumers is constant along this margin. In other words, the prices charged by firms leave the share of naïve consumers in the market overall constant. In general, prices still depend on the share of naïve consumers at the margin, but this marginal naïveté $\alpha_{0,i}$ is firm dependent and needs to take into account the participation margin. Our basic insights generalize to this setting, but there is an additional targeting possibility: Firms prefer to inform consumers on the between-firm margin over informing consumers at their participation margin (though this is still profitable).

Finally, our discussion depends on transparency being feasible. In some applications, it can be difficult and prohibitively costly. Consumers forget about hidden fees even after they were directly and painfully confronted with them by paying (Agarwal et al., 2013). Similarly, Alan et al. (2018) find that reminder messages containing information about overdraft fees reduce overdraft as long as they are sent to consumers, with no long

[29] In particular, if naïfs are more dispersed than sophisticates ($\alpha > \alpha_0 \iff f_N(0) < f_S(0)$ in the symmetric case), there is a separating equilibrium with profit μ. To see this, let $e < b$ denote the avoidance cost. Suppose there is a separating equilibrium. Then, the equilibrium contract for naïfs is $(\mu_N - b, b)$ and for sophisticates (note that there is multiplicity, as only the bundle price matters as long as the fee is below e) $(\mu_S, 0)$. This is incentive compatible for sophisticates as long as $\mu_N - b + e \geqslant \mu_S$. Naïfs never find the sophisticates contract optimal, as we can increase the transparent price while reimbursing it through a negative fee. The IC condition is satisfied since $\mu_\theta = \frac{1}{2f_\theta(0)}$ in the symmetric case.

If the separation condition is violated, existence is problematic because of cream skimming. In the pooling contract, the fee needs to be b, since otherwise naïve consumers can be skimmed. With this fee, sophisticated consumers can be skimmed by offering a contract in which they don't have to avoid the fee.

lasting effects, while Stango and Zinman (2014) find that repeated surveys about overdraft fees had persistent impact on overdraft. When transparency is costly, regulation can be doubly counterproductive: It may decrease aggregate welfare because of direct costs of transparency and further transfer surplus from consumers to firms as the transparent price increases.

From a regulators perspective, there can be additional costs of higher transparency. Arguably, the factor limiting consumer sophistication in many applications is not knowledge but attention and increasing the salience of one source of surprise fees may divert attention from others.

3.6.3 Identification of the Summary Statistic $\alpha_0 - \alpha$

To directly assess the prevalence and impact of price transparency, we require estimates of our summary statistic $\alpha_0 - \alpha$, the (reweighted) difference in cross-price elasticities. Several articles document consumer naïveté about hidden costs (e.g. Stango and Zinman (2009); Grubb and Osborne (2015)), but to our knowledge there is little direct evidence on differences in base price responsiveness between naïve and sophisticated consumers. Using micro data on fees paid by consumers, Nelson (2018) shows that realized profits from otherwise identical consumers are higher after late payments as firms reacted to these events with higher prices, without causing consumers to change credit cards. Hence, such events are a signal of low elasticity. Our model implies an estimation strategy based on aggregate data if there is plausibly exogenous variation in the hidden fee. Then

$$\frac{\partial p^\star}{\partial b} = \alpha_0 \tag{3.6.1}$$

$$\frac{\partial \pi_i^\star}{\partial b} = \frac{1}{2}(\alpha_i - \alpha_0) \tag{3.6.2}$$

and all three key quantities of our model ($\alpha_A, \alpha_B, \alpha_0$) are identified.[30] This variation could come from changes in laws across time or locations., but note that the laws are endogenous if they are affected by lobbying.

3.7 Conclusion

We show that the exploitation of consumer naïveté affects profits by changing the intensity of competition in the market. If naïve consumers are more elastic than sophisticates and hence concentrated at the margin, the marginal consumer is overproportionally attractive and profits are depressed. If, by contrast, naïve consumers are less elastic and hence underrepresented at the margin, exploiting naïveté is profitable. Firms increase transparency towards all consumers, e.g. via self-regulation, if and only if exploiting consumer naïveté depresses profits. The interests of firms and consumers are opposed: The very fact that firms refrain from increasing price transparency on their own initiative suggests that regulation imposing price transparency would increase consumer surplus. The targeting of transparency matters. Firms desire to be transparent to marginal consumers in order to increase the industry price, while keeping captured consumers in the dark in order to to continue to extract surprise fees.

There is evidence that for credit cards, naïve consumers are less elastic and – in line with our predictions – we see little industry-led transparency in this market. The prevalence of industry-led transparency in phone plans suggests that the opposite is true in this market and that transparency decreases consumer surplus. Our results about the targeting of transparency help to understand the strategies in the printer market, where print costs are shrouded in widely available advertisements but addressed aggressively in information targeted to undecided (and sophisticated) consumers.

[30] These equations hold in the simple case of a fully covered market. The model with a consumer participation margin in addition requires an estimate of the average elasticity along this margin.

References

_ , _ , _ , and _ , "Learning in the Credit Card Market," SSRN Scholarly Paper April 2013.

_ , Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel, "Regulating Consumer Financial Products: Evidence from Credit Cards," *The Quarterly Journal of Economics*, February 2015, *130* (1), 111–164.

Alan, Sule, Mehmet Cemalcilar, Dean Karlan, and Jonathan Zinman, "Unshrouding: Evidence from Bank Overdrafts in Turkey," *The Journal of Finance*, 2018, *73* (2), 481–522.

_ , _ , and Jidong Zhou, "Consumer Protection and the Incentive to Become Informed," *Journal of the European Economic Association*, 2009, *7* (2-3), 399–410.

Ausubel, Lawrence M., "The Failure of Competition in the Credit Card Market," *The American Economic Review*, 1991, pp. 50–81.

_ , "The Consumer Credit Card Market 2017," *Report*, December 2017, p. 352.

CTIA, "CTIA, Federal Communications Commission and Consumers Union Announce Free Alerts to Help Consumers Avoid Unexpected Overage Charges," October 2011.

Dahremöller, Carsten, "Unshrouding for Competitive Advantage: Unshrouding for Competitive Advantage," *Journal of Economics & Management Strategy*, September 2013, *22* (3), 551–568.

Eliaz, Kfir and Ran Spiegler, "Contracting with Diversely Naive Agents," *The Review of Economic Studies*, July 2006, *73* (3), 689–714.

_ **and** _ , "Consumer Optimism and Price Discrimination," *Theoretical Economics*, 2008, *3* (4), 459–497.

Ellison, Glenn, "A Model of Add-On Pricing," *The Quarterly Journal of Economics*, May 2005, *120* (2), 585–637.

Gabaix, Xavier and David Laibson, "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *The Quarterly Journal of Economics*, May 2006, *121* (2), 505–540.

Gomes, Renato and Jean Tirole, "Missed Sales and the Pricing of Ancillary Goods*," *The Quarterly Journal of Economics*, November 2018, *133* (4), 2097–2169.

Grubb, Michael D., "Consumer Inattention and Bill-Shock Regulation," *The Review of Economic Studies*, January 2015, *82* (1), 219–257.

_ **and Matthew Osborne**, "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock," *American Economic Review*, January 2015, *105* (1), 234–271.

Gurun, Umit G., Gregor Matvos, and Amit Seru, "Advertising Expensive Mortgages," *The Journal of Finance*, October 2016, *71* (5), 2371–2416.

_ **and** _ , "On the Welfare Costs of Naiveté in the US Credit-Card Market," *Review of Industrial Organization*, November 2015, *47* (3), 341–354.

_ **and** _ , "Naivete-Based Discrimination," *The Quarterly Journal of Economics*, May 2017, pp. 1019–1054.

_ **and** _ , "Behavioral Industrial Organization," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics: Applications and Foundations 1*, Vol. 1, North-Holland, January 2018, pp. 517–612.

_ **,** _ **, and Takeshi Murooka**, "Exploitative Innovation," *American Economic Journal: Microeconomics*, 2016, pp. 1–23.

_ **,** _ **, and** _ , "Inferior Products and Profitable Deception," *The Review of Economic Studies*, January 2017, *84* (1), 323–356.

HP, "Just In Time for the Holidays, Staples Delivers Customers Huge Savings on Printing with HP Instant Ink-Eligible Printers," December 2014.

_ , "HP Eliminates Ink Anxiety, Slashes Ink Costs up to 50 Percent with New Instant Ink Printers," September 2015.

Johnen, Johannes, "Dynamic Competition in Deceptive Markets," *The RAND Journal of Economics*, forthcoming.

Kosfeld, Michael and Ulrich Schüwer, "Add-on Pricing in Retail Financial Markets and the Fallacies of Consumer Education*," *Review of Finance*, May 2017, *21* (3), 1189–1216.

Li, Sanxi, Martin Peitz, and Xiaojian Zhao, "Vertically Differentiated Duopoly with Unaware Consumers," *Mathematical Social Sciences*, July 2014, *70*, 59–67.

Miao, Chun-Hui, "Consumer Myopia, Standardization and Aftermarket Monopolization," *European Economic Review*, October 2010, *54* (7), 931–946.

Nelson, Scott T, "Private Information and Price Regulation in the US Credit Card Market," *mimeo*, July 2018, p. 79.

Ryan, Stephen, "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, 2012, *80* (3), 1019–1061.

Stango, Victor and Jonathan Zinman, "What Do Consumers Really Pay on Their Checking and Credit Card Accounts? Explicit, Implicit, and Avoidable Costs," *American Economic Review*, April 2009, *99* (2), 424–429.

_ **and** _ , "Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees," *Review of Financial Studies*, April 2014, *27* (4), 990–1030.

Wolff-Mann, Ethan, "Verizon Overage Warning Can Give You an Overage and $15 Fee," *Time*, March 2016.

Appendix

3.A Proofs

For ease of notation we drop the argument of functions where no confusion can arise. To avoid duplication of arguments, we prove many results with a participation margin as in Subsection 3.5.1. The results in the main sections follow by setting $G \equiv 0, g \equiv 0$. Note that Assumption 3.6 assures that the share of naïve consumers in the market

$$\frac{\alpha(1-G_N(p-\mathbb{E}_{F_N}[v\chi_{\{v\geqslant 0\}}]))}{\alpha(1-G_N(p-\mathbb{E}_{F_N}[v\chi_{\{v\geqslant 0\}}]))+(1-\alpha)(1-G_S(p-\mathbb{E}_{F_S}[v\chi_{\{v\geqslant 0\}}]))}=\alpha \tag{3.A.1}$$

independent of p. We can hence work with α as a parameter independent of prices also in this case.

Proof of Proposition 3.1 and Proposition 3.6: Fix a share of participating consumers.

An interior equilibrium requires $p_A = p_B$: From the FOC

$$\begin{aligned}\int_{-\infty}^{p_B-p_A} dF(v)-(p_A+\alpha(p_B-p_A)b)\,f(p_B-p_A) &= 0 &(3.A.2)\\ \int_{p_B-p_A}^{\infty} dF(v)-(p_B+\alpha(p_B-p_A)b)\,f(p_B-p_A) &= 0 &(3.A.3)\end{aligned}$$

and subtract one from the other. Then

$$2F(p_B-p_A)-1=(p_A-p_B)\,f(p_B-p_A) \tag{3.A.4}$$

Since $F(0) = \frac{1}{2}$, the LHS is greater than zero iff $p_A \leqslant p_B$. The RHS is greater than zero iff $p_A \geqslant p_B$, which establishes $p_A = p_B$ in any equilibrium.

To get the equilibrium price, we compute

$$F(0) - (p_i + \alpha(0)b)\, f(0) = 0 \tag{3.A.5}$$

$$p_i = \frac{F(0)}{f(0)} - \alpha(0)b \tag{3.A.6}$$

As consumers have correct expectations about prices in equilibrium, our restriction of notation to the symmetric case was justified and we can solve for profits

$$\pi_A = (1 - G(p)) \int_{-\infty}^{0} (\mu - \alpha(0)b + \alpha(v)b)\, dF(v) \tag{3.A.7}$$

$$= \frac{1 - G(p)}{2} [\mu + (\alpha_i - \alpha(0))\, b] \tag{3.A.8}$$

and analogously for firm B. The unique interior solution has positive profits (Condition (3.3.1)). The second order conditions are satisfied if

$$\left(Ff' - f^2\right) - (1 + \alpha' b) f^2 < 0, \tag{3.A.9}$$

$$-\left((1 - F)\, f' + f^2\right) - (1 - \alpha' b) f^2 < 0, \tag{3.A.10}$$

which is the case globally, since the first term is negative by log-concavity and the second by the condition on α'. To rule out global deviations, we need to check that firms achieve positive profits in the equilibrium, but this is true by assumption. Furthermore, note that the best response is unique also to mixed strategies, so there can be no mixed EQ. To check the expression as densities in the symmetric case, note that

$$\pi^\star = \frac{1 - G(p^\star)}{2} \left[\frac{\frac{1}{2}}{f} + \left(\alpha - \frac{\alpha f_N}{f}\right) b\right] \tag{3.A.11}$$

$$= \frac{1 - G(p^\star)}{2} \frac{\frac{1}{2} + \alpha(1 - \alpha)(f_S - f_N) b}{f} \tag{3.A.12}$$

This is positive whenever

$$\frac{\frac{1}{2}+\alpha(1-\alpha)(f_S-f_N)b}{f} \geqslant 0 \tag{3.A.13}$$

$$\frac{1}{2\alpha(1-\alpha)b}+(f_S-f_N) \geqslant 0 \tag{3.A.14}$$

$$f_N \leqslant f_S+\frac{1}{2\alpha(1-\alpha)b} \tag{3.A.15}$$

as we assumed.

By computation

$$\frac{\partial}{\partial b}\frac{1-G(\mu-\alpha(0)b)}{2}\left[\mu+(\alpha_i-\alpha(0))\,b\right] \tag{3.A.16}$$

$$=\frac{1-G(\mu-\alpha(0)b)}{2}(\alpha_i-\alpha(0))+\alpha(0)\frac{g(\mu-\alpha(0)b)}{2}\left[\mu+(\alpha_i-\alpha(0))\,b\right] \tag{3.A.17}$$

Rearranging the inequality yields

$$\left(\frac{\alpha_i-\alpha(0)}{\alpha(0)}\right) \qquad \geqslant \qquad -\frac{g(\mu-\alpha(0)b)}{1-G(\mu-\alpha(0)b)}\left[\mu+(\alpha_i-\alpha(0))\,b\right] \tag{3.A.18}$$

The claim about taste dispersion follows since the RHS is always negative and

$$\frac{\alpha_i-\alpha(0)}{\alpha(0)}=\frac{f_S-f_N}{f_N} \tag{3.A.19}$$

□

Proof of Proposition 3.2: The strategies $(p_A^\star, p_B^\star)$, $t_A^\star = t_B^\star = 0$ and $b_A^\star = b_B^\star = b$ are an EQ: Suppose firm A deviates to (t', b', p'). Then,

$$p' \int_{-\infty}^{p_B^\star - p'} dF(v_B) + b' \int_{-\infty}^{p_B^\star - p'} \alpha(v, t', 0)\, dF(v) \leqslant \tag{3.A.20}$$

$$p' \int_{-\infty}^{p_B^\star - p'} dF(v) + b \int_{-\infty}^{p_B^\star - p'} \alpha(v, 0, 0)\, dF(v) \leqslant \tag{3.A.21}$$

$$p_A^\star \int_{-\infty}^{p_B^\star - p_A^\star} dF(v) + b \int_{-\infty}^{p_B^\star - p_A^\star} \alpha(v, 0, 0)\, dF(v) = \pi_A^\star \tag{3.A.22}$$

where the first inequality follows from monotonicity in b_i and α and the second from the optimality of $p_A^\star$. The argument for firm B is identical.

There can be no other equilibrium (t_A, b_A, p_A), (t_B, b_B, p_B) : Consider profits of one firm with $(t_A, b_A) \neq (0,0)$, WLOG firm A, with resulting share of naïve consumers $\alpha(v, t_A, t_B)$. We know that

$$\int_{-\infty}^{p_B - p_A} \alpha(v, t_A, t_B)\, dF(v) < \int_{-\infty}^{p_B - p_A} \alpha(v, 0, 0)\, dF(v) \tag{3.A.23}$$

by the nontriviality condition. Consider a deviation keeping the price fixed and setting $t_A = 0$ and $b_A = b$. It is profitable:

$$p_A \int_{-\infty}^{p_B - p_A} dF(v_B) + b_A \int_{-\infty}^{p_B - p_A} \alpha(v, t_A, t_B)\, dF(v_B) < p_A \int_{-\infty}^{p_B - p_A} dF(v) + b \int_{-\infty}^{p_B - p_A} \alpha(v, 0, 0)\, dF(v) \tag{3.A.24}$$

□

Proof of Proposition 3.3: SPE implies that price setting follows Proposition 3.1. Hence firms maximize pricing game profits (Proposition 3.1). All equilibria are in pure strategies as the BR is always unique by strict concavity. □

Proof of Proposition 3.4: SPE implies that price setting follows Proposition 3.1. The FOC for a solution in the first stage is (for notational compactness, we focus on pure strategies. Everything generalizes by linearity.)

$$\frac{\partial \alpha_0(t_A, t_B)}{\partial t_i}\left(\frac{\alpha_i}{\alpha_0} - 1\right) \gtreqless c'(t_i)$$

for firm i, with equality if $t_i^\star \in (0,1)$, $t_A^\star = 0$ if the LHS is always lesser and $t_A^\star = 1$ if LHS is always greater than the RHS. Note that the LHS is always negative if $(\alpha_i - \alpha_0) > 0$. Hence no firm with a larger share of naïfs among captured consumers compared to marginal consumers sets $t_i > 0$ in any SPE.

Suppose conversely that $\alpha_i < \alpha_0$ and $g \equiv 0$. Then the FOC cannot be satisfied at $t_i = 0$ by our assumption on costs. Furthermore, it is immediate that $t_i = 1$ if $c \equiv 0$.

To see the final result, note that the BR relationship of firm j is shifted upwards relative to that of firm i if $\alpha_j > \alpha_i$. The result about stable EQ follows immediately. □

Proof of Proposition 3.5: The conditions for transparency and CS reduce to

$$\frac{\partial \alpha_0(t_A^\star, t_B^\star)}{\partial t_i}\left(\frac{\alpha}{\alpha_0} - 1\right) \geqslant c'(t_i)$$

$$\frac{\partial CS}{\partial t_i} = \frac{\partial \alpha_0(t_A^\star, t_B^\star)}{\partial t_i}(1 - \frac{\alpha}{\alpha_0})$$

By Assumption (3.5), there is nonzero transparency iff $\left(\frac{\alpha}{\alpha_0} - 1\right) < 0$, i.e. iff transparency reduces consumer surplus. (Recall that $\frac{\partial \alpha_0(t_A^\star, t_B^\star)}{\partial t_i} < 0$.) □

Proof of Proposition 3.7: First, we need to solve the cartel problem. Industry profits are given by

$$\Pi = \left(1 - [\alpha G_N + (1-\alpha) G_S]\,(\bar{p} - \int_0^\infty \nu \, dF(\nu))\right)\{\bar{p} + \alpha b\}$$

First, note that it is never optimal for a cartel to prescribe different prices for the two firms. Taking FOC

$$\frac{\partial \Pi}{\partial \bar{p}} = (1 - [\alpha G_N + (1-\alpha) G_S](\cdot)) - [\alpha g_N + (1-\alpha) g_S](\cdot)\{\bar{p} + \alpha b\}$$

The price is pinned down by a familiar inverse hazard rate formula.

$$\bar{p} = \frac{1 - G(\bar{p})}{g(\bar{p})} - \alpha b \tag{3.A.25}$$

A unique solution exists as we assume that G has a strictly increasing hazard rate, i.e. strictly decreasing inverse hazard rate. Note that the cartel also discounts the transparent price for gains from naïveté. It does so based on average gains, however, not on the margin between firms.

To see the statement of the proposition, note that

$$\begin{aligned}
&\frac{\partial \pi^\star}{\partial \alpha_0} = -(1 - G(p^\star))\, b \frac{\partial \alpha_0}{\partial t_A} + g(p^\star) \frac{\partial \alpha_0}{\partial t_A} b\, [\mu + (\alpha_A - \alpha_0)\, b] \geqslant 0 \\
\iff &(1 - G(p^\star)) - g(p^\star)\, [\mu + (\alpha_A - \alpha_0)\, b] \leqslant 0 \\
&\frac{(1 - G(p^\star))}{g(p^\star)} - \alpha_A b \geqslant p^\star
\end{aligned}$$

As we argued above, 3.A.25 has a unique solution with LHS<RHS iff $p < \bar{p}$, hence the above condition is satisfied iff $p^\star \leqslant \bar{p}$. □

Proof of Proposition 3.4: SPE implies that price setting follows Proposition 3.1. The FOC for a solution in the first stage is

$$\frac{\partial \alpha_0(t_A^\star, t_B^\star)}{\partial t_i} \left\{ \left(\frac{\alpha_i}{\alpha_0} - 1 \right) + \frac{g(\mu^\star)}{1 - G(\mu^\star)} \left[\mu + \alpha_0(t_A^\star, t_B^\star) \left(\frac{\alpha_i}{\alpha_0} - 1 \right) b \right] \right\} b \gtreqless \frac{c'(t_i^\star)}{1 - G(\cdot)}$$

for firm i, where $\mu^\star := \mu - \alpha_0(t_A^\star, t_B^\star)b$, with equality if $t_i^\star \in (0,1)$, $t_i^\star = 0$ if the LHS is always less and $t_i^\star = 1$ if LHS is always greater than the RHS.

If the condition of Prop. 3.6 is satisfied at $t = (0,0)$, the expression in curly brackets is weakly positive, hence the LHS weakly negative. It remains to show that the expression is increasing in transparency. First, suppose $\left(\frac{\alpha_i}{\alpha_0} - 1\right) > 0$: then it is always positive, independent of transparency. Second, suppose $\left(\frac{\alpha_i}{\alpha_0} - 1\right) \leqslant 0$: then the bracket term is increasing in transparency as is the hazard rate (as the price is increasing in transparency), so the whole expression is increasing.

The conversely statement is immediate. The statement about comparative statics follows as in Prop. 3.4. □

Proof of Proposition 3.9: Immediate, since welfare is single peaked. □

Proof of Proposition 3.10: To prove this result, we need to study the FOC.

$$\frac{\partial}{\partial t}\left[\frac{1}{2f_{t_A,t_B}} + \left(\alpha(t_A,t_B) - \frac{\alpha(t_A,t_B)f_N}{f_{t_A,t_B}}\right)b\right] = \frac{\frac{\partial\alpha(t_A,t_B)}{\partial t_i}(f_{NS}-f_N)}{2f_{t_A,t_B}^2} +$$
$$\left(\frac{\partial\alpha(t_A,t_B)}{\partial t_i} - \frac{\frac{\partial\alpha(t_A,t_B)}{\partial t_i}f_N f_{t_A,t_B} + \alpha(t_A,t_B)f_N\frac{\partial\alpha(t_A,t_B)}{\partial t_i}(f_{NS}-f_N)}{f_{t_A,t_B}^2}\right)b \propto$$
$$\left(\frac{f_N f_{t_A,t_B} + \alpha(t_A,t_B)f_N(f_{NS}-f_N)}{f_{t_A,t_B}^2} - 1\right)b - \frac{(f_{NS}-f_N)}{2f_{t_A,t_B}^2} \propto$$
$$\left(f_N f_{t_A,t_B} + \alpha(t_A,t_B)f_N(f_{NS}-f_N) - f_{t_A,t_B}^2\right)b - \frac{1}{2}(f_{NS}-f_N) =$$
$$\left(\underbrace{f_{t_A,t_B}(f_N - f_{t_A,t_B})}_{\text{pass-through effect}} - \underbrace{\alpha(t_A,t_B)f_N(f_N-f_{NS})}_{\text{relative population effect}}\right)b + \underbrace{\frac{1}{2}(f_N-f_{NS})}_{\text{dispersion effect}}$$

Let us interpret this equation. The first term is the direct effect of higher transparency on profits via pass-through as treated in the main sections of the article. If $f_N > f_{t_A,t_B}$, the "reimbursement" of revenue from hidden fees is larger than the revenue itself and firms gain from increased transparency. The second term dubbed relative population effect is

the impact of transparency on marginal naïveté through it's effect on the average elasticity in the market. Recall that $\alpha_0 = \frac{\alpha f_N}{f_{t_A,t_B}}$. If $f_N > f_{NS}$, transparency decreases f_{t_A,t_B} hence increasing the share of naïve consumers at the margin α_0 indirectly, decreasing the transparent price. The third term is the effect of transparency on the inverse elasticity μ. If $f_N > f_{NS}$, transparency makes demand less elastic, resulting in higher prices and profits.

Let us now assume for now that $f_{NS} = f_S$, we check that this is WLOG later.

$$
\begin{aligned}
&(f_{t_A,t_B}(f_N - f_{t_A,t_B}) - \alpha(t_A,t_B) f_N (f_N - f_{NS}))\, b + \frac{1}{2}(f_N - f_{NS}) = \\
&[f_{t_A,t_B}((1-\alpha(t_A,t_B))f_N - [\alpha(0,0) - \alpha(t_A,t_B)]\, f_{NS} - (1-\alpha(0,0))f_S) - \\
&\alpha(t_A,t_B) f_N (f_N - f_{NS})]\, b + \frac{1}{2}(f_N - f_{NS}) = \\
&[f_{t_A,t_B}((1-\alpha(t_A,t_B))(f_N - f_S)) - \alpha(t_A,t_B) f_N (f_N - f_S)]\, b + \frac{1}{2}(f_N - f_S) = \\
&\left[((1-\alpha(t_A,t_B))f_{t_A,t_B} - \alpha(t_A,t_B) f_N)\, b + \frac{1}{2}\right](f_N - f_S) = \\
&\left[((1-\alpha(t_A,t_B))(\alpha(t_A,t_B) f_N + (1-\alpha(t_A,t_B))f_S) - \alpha(t_A,t_B) f_N)\, b + \frac{1}{2}\right](f_N - f_S) = \\
&\left[\left((1-\alpha(t_A,t_B))^2 f_S - [\alpha(t_A,t_B)]^2 f_N\right) b + \frac{1}{2}\right](f_N - f_S)
\end{aligned}
$$

If

$$\left((1-\alpha(t_A,t_B))^2 f_S - [\alpha(t_A,t_B)]^2 f_N\right) \geqslant 0$$

the usual condition is satisfied independently of b. Note that the LHS is increasing in t, hence it is necessary that

$$\begin{aligned}
(1-\alpha)^2 f_S - \alpha^2 f_N &\geqslant 0 \\
(1-\alpha)^2 f_S &\geqslant \alpha^2 f_N \\
\frac{(1-\alpha)^2}{\alpha^2} &\geqslant \frac{f_N}{f_S} = \frac{1-\alpha}{\alpha}\frac{\alpha f_N}{f_0}\frac{f_0}{(1-\alpha)f_S} \\
\frac{1-\alpha}{\alpha} &\geqslant \frac{\alpha f_N}{f_0}\frac{f_0}{(1-\alpha)f_S} = \frac{\alpha_0}{1-\alpha_0} \\
1 &\geqslant \alpha + \alpha_0
\end{aligned}$$

Going back to the general condition

$$\begin{aligned}
\left((1-\alpha)^2 f_S - \alpha^2 f_N\right) b + \frac{1}{2} &\geqslant 0 \\
1 + \frac{1}{2}\frac{1-\alpha_0}{(1-\alpha)b f_S} &\geqslant \alpha + \alpha_0
\end{aligned}$$

It remains to show that checking the condition for $f_{NS} = f_S$ is sufficient. In the general case, we can rewrite (dropping zero arguments)

$$\begin{aligned}
&[f_{t_A,t_B}((1-\alpha(t_A,t_B))f_N - [\alpha - \alpha(t_A,t_B)]f_{NS} - (1-\alpha)f_S) - \\
&\qquad \alpha(t_A,t_B)f_N(f_N - f_{NS})]b + \frac{1}{2}(f_N - f_{NS}) = \\
&[f_{t_A,t_B}((1-\alpha(t_A,t_B))f_N - [\alpha - \alpha(t_A,t_B)]f_S - (1-\alpha)f_S) - \alpha(t_A,t_B)f_N(f_N - f_S)]b + \\
&\frac{1}{2}(f_N - f_S) + [f_{t_A,t_B}[\alpha - \alpha(t_A,t_B)](f_S - f_{NS}) - \alpha(t_A,t_B)f_N(f_S - f_{NS})]b + \frac{1}{2}(f_S - f_{NS}) = \\
&\qquad \left[\left((1-\alpha(t_A,t_B))^2 f_S - [\alpha(t_A,t_B)]^2 f_N\right)b + \frac{1}{2}\right](f_N - f_S) + \\
&\qquad \underbrace{\left[(f_{t_A,t_B}[\alpha - \alpha(t_A,t_B)] - \alpha(t_A,t_B)f_N)b + \frac{1}{2}\right](f_S - f_{NS})}_{\Delta(f_{NS})}
\end{aligned}$$

To show that our condition is sufficient, it is sufficient to show that $\Delta(f_{NS}) \geqslant 0$. Note that Δ is quadratic in f_{NS} and that

$$\Delta''(f_{NS}) = -[\alpha - \alpha(t_A, t_B)]^2 \, b$$

Hence it is sufficient to show that $\Delta(f_{NS}) \geqslant 0$ at the boundary of the range of f_{NS}. The first corner case $f_{NS} = f_N$ is treated in the main sections and $\Delta(f_S) = 0$. □