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 Information Economics

Shishkin, Denis

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01w0892d86n

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 INFORMATION ECONOMICS

Denis Shishkin

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

Advisors: Stephen Morris and Pietro Ortoleva

June 2020

© Copyright by Denis Shishkin, 2020.

All rights reserved.

Abstract

Information is crucial for making decisions under uncertainty. This dissertation explores how information is designed or elicited by a principal aiming to implement a certain objective.

Chapters 1 and 2 study information design in cases when there is no commitment to reveal information. In Chapter 1, I analyze optimal evidence acquisition in a game of voluntary disclosure. A sender seeks hard evidence to persuade a receiver to take a certain action, but there is uncertainty about whether evidence has been obtained. When the probability of obtaining evidence is low, I show that the optimal evidence structure is a binary certification: all it reveals is whether the state of the world is above or below a certain threshold. When binary structures are optimal, higher uncertainty leads to less pooling at the bottom because the sender uses binary certification to commit to disclose evidence more often.

Chapter 2 (co-authored with Elliot Lipnowski and Doron Ravid) studies how credibility affects persuasion. A sender uses a weak institution to disseminate information to persuade a receiver. Specifically, the weaker is the institution, the higher is the probability that its report reflects the sender's agenda rather than the officially announced protocol. We show that increasing this probability can benefit the receiver and can lead to a discontinuous drop in the sender's payoffs. To derive our results, we geometrically characterize the sender's highest equilibrium payoff, which is based on the concave envelope of her capped value function.

Finally, Chapter 3 (co-authored with Franz Ostrizek) explores monopolistic screening with frame-dependent valuations. A principal designs an extensive-form decision problem with frames at each stage. The optimal mechanism has a simple three-stage structure and uses changes in framing (high-low-high) to induce dynamic inconsistency and thereby reduce information rents. To achieve this, the principal offers unchosen decoy contracts. Sophisticated consumers correctly anticipate that if they deviated,

they would choose a decoy, which they want to avoid in the low frame. This allows the principal to eliminate some incentive constraints. With naive consumers, the principal can perfectly screen by cognitive type and extract full surplus from naifs.

Acknowledgements

I am deeply indebted to my advisors Stephen Morris and Pietro Ortoleva for their endless wisdom and generosity. Their continuous guidance and inspiration have had a great impact on my development as an economist. Throughout all six years of my graduate studies, Stephen Morris' vision shaped my comprehension of economic theory. Ever since Pietro Ortoleva joined Princeton, his enthusiasm and encouragement have kept me motivated and passionate about economics.

I would like to thank Leeat Yariv for her invaluable advice and support, and also for sparking my interest in experimental economics. I am grateful to my co-authors and mentors Elliot Lipnowski and Doron Ravid. They were working independently on a model similar to the one in my third-year paper and graciously offered to join forces. That collaboration has proven to be a great opportunity to learn and has led to Chapter 2 of this dissertation.

I am thankful to Roland Bénabou, Faruk Gul, Sofia Moroni, Wolfgang Pesendorfer, Can Urgun, and Nikhil Vellodi, for their critical feedback on my research and advice on the job market.

I would like to thank Ellen Graf and Laura Hedden for always providing the necessary assistance. I thank Princeton University, William S. Dietrich II Economic Theory Center, and Stephen Goldfeld Fund for financial support. I am grateful to numerous conference and seminar participants for their helpful comments.

Many thanks to my colleagues and friends Lasse Mononen, Franz Ostrizek, Pellumb Reshidi, Kirill Rudov, Evgenii Safonov, Elia Sartori, and Ben Young, from whom I learned a great deal. I am especially grateful to Franz—with whom Chapter 3 of this dissertation is co-authored—for his enthusiasm that has been keeping our projects going.

Also thanks to Tasoula Bartsi, Michael Dobrew, Oleg Itskhoki, Jonas Jin, Nastya Karpova, Charis Katsiardis, Sergii and Lera Kiiashko, Alexander Kopytov, Evgeniia Lambri-

naki, Alexey Lavrov, Dima Mukhin, Oleg Muratov, Mark Razhev, Dasha Rudova, and Andrei Zeleneev, and many others who made my time at Princeton so joyful.

I am grateful to my loving and caring family, my parents Andrei and Elena, and my sister Anastasia, who have always believed in me. Finally, I thank my wife Olga for her love, encouragement, and patience throughout the years.

To my parents, Andrei and Elena.

Contents

Abstract iii

Acknowledgements v

1 Evidence Acquisition and Voluntary Disclosure **1**

1.1 Introduction 2

1.2 Model 10

1.3 Analysis 15

1.3.1 Voluntary disclosure 16

1.3.2 Value of Evidence 19

1.3.3 Optimal Evidence Acquisition 22

1.3.4 Degradation of Certification Standards 27

1.3.5 Voluntary vs Mandatory Disclosure 29

1.3.6 Welfare 30

1.4 Conclusion 34

2 Persuasion via Weak Institutions **35**

2.1 Introduction 36

2.2 A Weak Institution 44

2.3 Persuasion with Partial Credibility 45

2.4 Varying Credibility 49

2.5 Persuading the Public 52

2.6 Investing in Credibility 57
2.7 Conclusion 59

3 Screening with Frames **61**
3.1 Introduction 62
3.2 Screening with Frames and Extensive Forms 70
3.2.1 Contracts and Frames 70
3.2.2 Extensive-Form Decision Problems 72
3.2.3 The Firm's Problem 74
3.3 Optimal Screening 75
3.3.1 Optimal Structure of the Extensive Form 76
3.3.2 Necessary and Sufficient Conditions for Implementation 79
3.3.3 Optimal Contracts 83
3.3.4 Optimal Concealed Types 86
3.3.5 Discussion 88
3.4 Extensions 91
3.4.1 Naivete 91
3.4.2 Additional Participation Constraints and Cool-off Regulation . . 96
3.5 Conclusion and Discussion 99

A Supplemental Material **102**
A.1 Appendix for Chapter 1: Proofs 103
A.2 Appendix for Chapter 2: Proof Exposition 111
A.2.1 Visualizing Theorem 2.1 111
A.2.2 Exposition for Proposition 2.1 113
A.2.3 Exposition for Proposition 2.2 113
A.2.4 Exposition for Proposition 2.3 115
A.2.5 Exposition for Claim 2.1 116

A.3 Appendix for Chapter 2: Proofs 119
A.3.1 Toward the Proof of the Main Theorem 119
A.3.2 Productive Mistrust: Proofs 130
A.3.3 Collapse of Trust: Proof of Proposition 2.2 139
A.3.4 Robustness: Proof of Proposition 2.3 140
A.3.5 Persuading the Public: Proofs from Section 2.5 143
A.3.6 Proofs from Section 2.6: Investing in Credibility 154
A.4 Appendix for Chapter 3: Proofs 157
A.4.1 Preliminaries 157
A.4.2 Proofs 159

Bibliography **177**

Chapter 1

Evidence Acquisition and Voluntary Disclosure

1.1 Introduction

Hard evidence is often sought and disclosed by one party (sender) to persuade another (receiver) to take a certain action. For example, pharmaceutical companies test new drugs to get the approval from the US Food and Drug Administration, startups build and test prototypes to secure financing, sellers apply for quality certification to persuade consumers to buy products, etc. However, in many cases the receiver may be uncertain about whether the sender has obtained the evidence. In the above examples, medical test results may have been inconclusive, a prototype may have been prohibitively costly to experiment with, and quality certification may have been delayed. In many such cases, even if the sender has evidence, she may be able to pretend to be uninformed. In other words, she can conceal unfavorable evidence by claiming ignorance. This creates a trade-off for acquisition of evidence. Before evidence is obtained, the sender may prefer the receiver to learn something about the state. But after she obtains it, it might be in her best interest not to disclose it.

Consider the following example. An entrepreneur has a project of unknown quality. She can seek verifiable information on its quality to persuade an investor to provide financing. Before obtaining the evidence, she may prefer detailed information about the quality to be released to the investor, regardless of its contents. This is the case if, for example, evidence about moderately low quality allows the entrepreneur to secure at least partial funding. But suppose that the disclosure is voluntary and the investor is uncertain about whether the entrepreneur is informed. Then, if the entrepreneur learns that the quality is low, she may prefer not to disclose the information and pretend to be uninformed. This prevents the investor from learning details about low-quality projects. Therefore, the entrepreneur must decide what information to seek taking into account her future disclosure incentives. We show that this substantially affects which information is sought in the first place.

In principle, when the state of the world is rich and the set of messages that can be sent is large, one might expect to see complex communication between the agents. In reality, however, senders often rely on verifiable information that is very coarse. In many cases, it is as simple as a *binary certification*: a signal that reveals only whether the state of the world is sufficiently good. For example, often sellers apply for certifications that test whether their products have high enough quality, job candidates take professional exams with pass or fail grades, etc. This paper shows that the mere opportunity to conceal information as described above can lead in equilibrium to acquisition of simple information structures such as binary certification.

To study these interactions, we consider a communication game between a sender (she) and a receiver (he). The state of the world is continuous and unknown to both players. The sender wants the receiver to take a certain action, but the receiver takes the action only if his expectation of the state exceeds his privately known cutoff. The sender publicly chooses what information to acquire, but there is an exogenous uncertainty about whether she will obtain any evidence from this inquiry. If she obtains the evidence, then she can voluntarily disclose it or pretend to not have obtained it. Otherwise, she cannot prove that she is uninformed.

Result 1: High uncertainty leads to binary certification. Our first main result (Theorem 1.1) shows that when there is a large enough probability that no evidence is obtained, the optimal evidence structure acquired by the sender is a *binary certification*: it reveals only whether the state is above or below a certain threshold. Otherwise, the optimum is a two-sided censorship, which is similar to binary certification, but also reveals intermediate states. Fig. 1.1 illustrates these two types of optimal evidence structures.



(a) Binary certification (b) Two-sided censorship

FIGURE 1.1: Two forms of optimal evidence structures.

To get some intuition why binary certification is optimal, note that it is an information structure that assigns a single message (PASS) to the states above a threshold and a single message (FAIL) to those below. In other words, the states are pooled at the top and at the bottom of the distribution. We identify two distinct forces that drive pooling of high and low states, and show that binary certification is optimal when the interaction between them is non-trivial. First, pooling at the bottom happens because the disclosure is voluntary. In our example because the entrepreneur cannot commit to always disclose, if she learns that that the project's quality is sufficiently low, she will pretend pretend to not have obtained evidence. Second, pooling at the top arises because of the sender's uncertainty about the receiver's cutoff for action. If the distribution of cutoffs is single-peaked, there are increasing returns to disclosing more (less) information about low (high) states. Therefore, in the absence of disclosure concerns, the sender ex-ante prefers to reveal low states and pool high states.

To illustrate how these two forces can interact in a non-trivial way, consider the optimal evidence structure for various values of the probability q of obtaining evidence. First, suppose q is close to 1. In this case, the optimal evidence structure is a two-sided censorship: it reveals whether the state is above an upper threshold and below a lower threshold via PASS and FAIL messages, respectively, and perfectly reveals the intermediate states. The two forces driving pooling at the top and bottom in this case do *not* interact. To see this, suppose that probability q slightly decreases. Then the receiver becomes less skeptical when the sender claims ignorance. This, in turn, incentivizes the sender to conceal more, and the lower pooling region becomes larger. But the incentive to pool the states at the top is unaffected by that. In particular, the upper threshold

stays constant at the level the sender would choose absent the voluntary disclosure problem. In other words, there is "separability" between the two forces in the case of two-sided censorship.

But now suppose that the probability q of obtaining evidence is low. In this case, the two forces interact in a non-trivial way, and we show that this leads to binary certification. Why does the sender choose to acquire so little information? Suppose that the sender instead chose fully revealing evidence structure. Then if q is low, she would often claim to be uninformed because the receiver is not too skeptical when there is no disclosure. Overall, this leads to a large concealment at the bottom, which hurts the sender's ex-ante expected payoff. To mitigate this problem, she designs the signal so that she then discloses more often. This is exactly what binary certification achieves: when the threshold is relatively low, the PASS message is assigned to the states that would otherwise be concealed. So the sender end up disclosing more often, albeit only a single message.

Fig. 1.2 illustrates the evidence structure acquired by the sender in equilibrium. For each value of q on the vertical axes, it shows the optimal partition of the state space. If the probability of obtaining evidence is low ($q < \overline{q}$), there is a binary certification threshold, such that states are pooled above and below this threshold. If the probability of obtaining evidence is high ($q > \overline{q}$), the states are pooled above the upper threshold, pooled below the lower threshold, and fully revealed otherwise. As discussed above, the interaction between the two forces driving pooling at the top and bottom is trivial under two-sided censorship: upper threshold stays constant as the size of lower pooling region changes. But at $q = \overline{q}$ the interaction becomes non-trivial and the sender switches to binary certification. Since she uses binary certification to commit to disclose more often, the threshold for upper pooling region may drop discontinuously as q declines below $\overline{q}$.

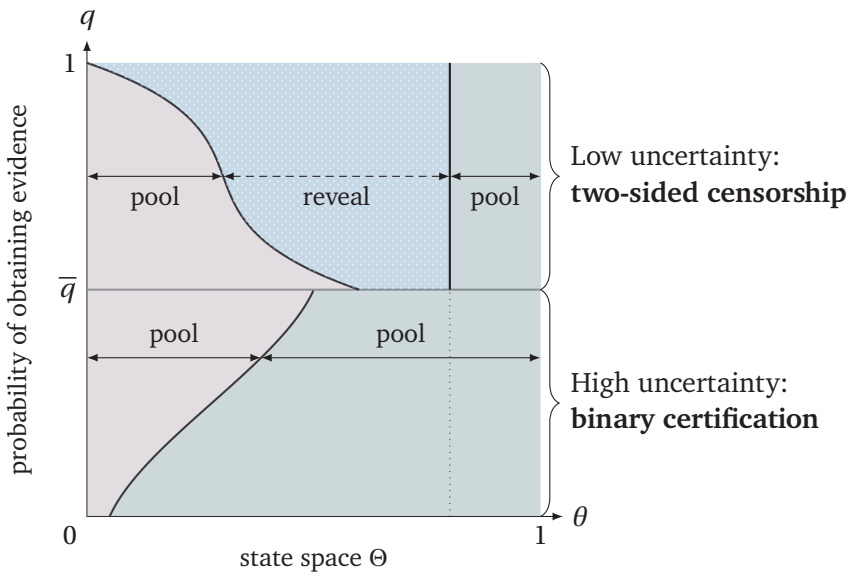


FIGURE 1.2: Optimal evidence structure for various levels of uncertainty.

Result 2: Pooling at the bottom is non-monotone. The second main result (Theorem 1.2) shows that when there is binary certification, the certification standards degrade as uncertainty increases. That is, the lower the probability q, the lower is the threshold. This implies that the sender facing less skeptical receiver will choose a signal with less pooling at the bottom. At first, it might sound surprising as, for a fixed evidence structure, lower skepticism incentivizes the sender to conceal more. Indeed, due to a trivial interaction between the design and disclosure forces, it leads to more pooling at the bottom. In contrast, when binary certification is optimal, this effect is reversed. This further highlights the interaction between the design and disclosure forces. The intuition is the following. The sender switches to a relatively low binary certification threshold because it allows the sender to commit to disclose more often. This allows to mitigate the problem of limited commitment due to voluntary disclosure. As uncertainty increases, this problem becomes more severe and lower thresholds become more effective. The non-monotonicity of the pooling at the bottom is evident in Fig. 1.2: as q decreases, the pooling at the bottom grows, but then begins to shrink once q declines below $\overline{q}$.

Welfare. We also study the implications for players' welfare. Notice that the higher q is, the more skeptical the receiver is when the sender claims ignorance. This disciplines the sender to disclose more and, therefore, the conflict between the sender's ex-ante and interim preferences for disclosure is lower. Unsurprisingly, this implies that the sender's equilibrium value is increasing in her probability q of obtaining evidence. As for the receiver, it follows from our equilibrium characterization that the informativeness of two-sided censorship is increasing in q in the Blackwell sense. However, optimal binary-certification signals are not Blackwell-comparable for different values of q, since higher q makes PASS more informative and FAIL less informative. But as higher q means there is a smaller chance the sender is uninformed, we show that the overall disclosed signal is Blackwell more informative. Thus, the receiver also benefits from a higher probability that evidence is obtained.

The above analyses compare environments with different upper bounds on the information that the receiver can get. If q is very small, then the receiver learns very little, regardless of the sender's strategy. Therefore, the players' payoffs are increasing in q partly because higher q allows the sender to communicate more often. To isolate this effect, we normalize the players' payoffs by q and show that the normalized equilibrium payoffs are also increasing. This means that there are two channels through which the equilibrium payoffs are affected: higher probability of obtaining evidence allows the sender to communicate not only more often, but also more efficiently.

Related literature. This paper is related to the literature on disclosure of verifiable information (for a survey, see Milgrom, 2008).[1] The seminal works of Grossman (1981), Milgrom (1981), and Milgrom and Roberts (1986) study disclosure under complete provability, that is when the sender can prove any true claim. The key insight of those

[1]Coarseness of information is also a common feature in cheap-talk models (Crawford and Sobel, 1982) which study a different environment: the information is soft, it is given to the sender exogenously, and the coarseness follows from partially aligned preferences of the players. In our model, information is hard, acquired endogenously, and the sender has state-independent preferences. See Pei (2015) and Argenziano, Severinov, and Squintani (2016) on information acquisition in a cheap-talk model.

papers is that complete provability implies "unraveling", which leads to full information revelation in equilibrium (for a recent generalization, see Hagenbach, Koessler, and Perez-Richet, 2014).[2]

Our model is based on the approach of Dye (1985) and Jung and Kwon (1988), in which evidence is obtained with some probability and there is partial provability: if the sender is uninformed, she cannot prove it.[3] The main innovation compared to this literature is that the evidence the sender obtains is chosen endogenously. Some recent papers (Kartik, Lee, and Suen, 2017; Bertomeu, Cheynel, and Cianciaruso, 2018; DeMarzo, Kremer, and Skrzypacz, 2019) endogenize the sender's endowment of evidence in Dye (1985) framework.[4] Kartik, Lee, and Suen (2017) study a multi-sender disclosure game, where senders can invest in higher probability of obtaining evidence, while taking the evidence structure as given.

Bertomeu, Cheynel, and Cianciaruso (2018) study a closely related problem, in which the firm is maximizing its expected valuation by choosing an asset measurement system, subject to strategic withholding and disclosure costs. The firm makes an additional interim investment decision with a convex cost, which leads to its objective being convex in the market's posterior mean. Their model with zero disclosure costs can be mapped into a special case of our model, where the PDF of the receiver's type is increasing. In this case, it is optimal to acquire a fully-informative evidence structure

[2]Another common point of inquiry in this literature is informational efficiency of voluntary disclosure compared to the receiver's commitment outcome, see e.g. Glazer and Rubinstein (2004, 2006); Sher (2011); Hart, Kremer, and Perry (2017); Ben-Porath, Dekel, and Lipman (2019).

[3]Other approaches in which unraveling fails include costly disclosure models of Jovanovic (1982) and Verrecchia (1983) and multidimensional disclosure models of Shin (1994) and Dziuda (2011). Okuno-Fujiwara, Postlewaite, and Suzumura (1990) provide sufficient conditions for unraveling in two-stage games, where in the first stage players can disclose private information, and give examples in which unraveling does not happen.

[4]In Matthews and Postlewaite (1985), the sender makes a binary evidence acquisition decision before playing a voluntary disclosure game under complete provability. Gentzkow and Kamenica (2017) study overt costly acquisition of evidence in a disclosure model where each type can perfectly self-certify and show that one or more sender(s) disclose everything they acquire. Escudé (2019) provides an analogous result in a single-sender setting with covert costless acquisition and partial verifiability.

for any probability of obtaining evidence, and, therefore, the interaction between the design and disclosure incentives plays no role.

In DeMarzo, Kremer, and Skrzypacz (2019), evidence acquisition is covert, that is, the sender's signal choice is observed only if she discloses its realization.[5] They characterize the ex-ante incentive compatibility with a "minimum principle" and show that it is sufficient for the sender to choose simple tests, equivalent to binary certification. Interestingly, their result is driven by forces that are very different from ours. More precisely, as their sender's objective is linear, she is ex-ante indifferent between all information structures and might as well choose a simple test that satisfies a "minimum principle". In contrast, we provide conditions for binary certification to be the unique optimum (up to outcome equivalence) in environments with the convex-concave sender's objective and acquisition is overt. Although some of our results will continue to hold even if the choice of a signal was unobserved, in general, it is not clear what would happen in the case of covert acquisition and non-trivial incentives for evidence design.

This paper also contributes to the literature on Bayesian persuasion and information design (for a survey, see Kamenica, 2019). In the special case of our model when the sender is known to possess the evidence ($q = 1$), the unraveling argument applies, and the optimal evidence acquisition problem becomes equivalent to the one of Bayesian persuasion (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011). This problem in similar environments was studied by Alonso and Câmara (2016b), Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), Kolotilin (2018), and Dworczak and Martini (2019). In particular, it follows from their analyses that upper censorship is optimal if the receiver's type distribution is unimodal. Information structures equivalent to our binary certification and two-sided censorship also appear in Kolotilin (2018) in cases

[5]Ben-Porath, Dekel, and Lipman (2018) study a related voluntary disclosure problem, in which there is an ex-ante covert choice between risky projects, which, in our setting, corresponds to a choice between priors.

when the distribution of the receiver's type is not unimodal. There, binary certification can be optimal because of a particular shape of the the receiver's type distribution (e.g. bimodal), rather than the interaction between the design and disclosure incentives.

A standard assumption in this literature is that the sender commits to a signal, whose realization is directly observed by the receiver, while in our model it is voluntarily disclosed by S. Some recent works (Felgenhauer, 2019; Nguyen and Tan, 2019) also relax the assumption that the receiver observes signal realizations. In Felgenhauer (2019), the sender designs experiments sequentially at a cost and can choose when to stop experimenting and which outcomes to disclose. Nguyen and Tan (2019) study a model of Bayesian persuasion with costly messages, where a special case of the cost function corresponds to verifiable disclosure of hard evidence studied in this paper. The difference is that their sender can choose not only a signal about the state, but also the probability of obtaining evidence. In contrast, q is exogenous in our model. If it could be chosen by the sender, she would set $q = 1$ and obtain her full commitment payoff.

1.2 Model

Setup. There are two players: a sender (S, she) and a receiver (R, he). The state of the world is $\theta \in \Theta = [0,1]$, unknown by both players, who share a prior $\mu_0 \in \Delta\Theta$, which admits a full-support density and has a mean $\theta_0 := \mathbb{E}\mu_0$.[6] R has a private payoff type $\omega \in \Omega = [0,1]$, which is independent of θ and distributed according to a continuous distribution with CDF H and strictly quasi-concave PDF h with a peak at $\hat{\omega} \geqslant \theta_0$.[7] R either acts ($a = 1$) or not ($a = 0$) and has a utility $u_R(a,\theta,\omega) = a(\theta - \omega)$. That is, R

[6]Throughout the paper, $\Delta\Theta$ denotes the set of all Borel probability measures on Θ and, for any $\mu \in \Delta\Theta$, $\mathbb{E}\mu$ denotes the expectation $\int \theta \, \mathrm{d}\mu(\theta)$.

[7]The assumption $\hat{\omega} \geqslant \theta_0$ can be interpreted as the conflict between the players' preferences being moderately large for a given H. If conflict is small ($\hat{\omega} \leqslant \theta_0$) and H is close enough to be degenerate at $\hat{\omega}$, then the uninformative signal is optimal. For a fixed H, if the conflict is small, the uninformative signal may not necessarily optimal.

prefers to act if and only if his expectation of the state is at least as high as his type. The sender always wants R to act and has a utility $u_S(a,\theta,\omega) = a$.

The timing of the game is as follows. First, S publicly chooses what evidence to acquire at no cost. Formally, she commits to a signal $\pi : \Theta \to \Delta M$, where M is a rich enough set of messages.[8] Then, the nature draws the state θ from μ_0, the message m from $\pi(\theta)$, and the set of available messages $\hat{M}$ as follows. With probability $q \in (0,1]$, $\hat{M} = \{m, \varnothing\}$, which means S obtains a proof that the realized message is m and chooses a message $\hat{m} \in \hat{M}$, i.e. whether to disclose it or claim to not have obtained it. With probability $1-q$, $\hat{M} = \{\varnothing\}$, which means that she has not obtained any proof and must send $m = \varnothing$.[9] Finally, R's type ω realizes, he observes $\hat{m}$ and π, updates his belief, and chooses an action.

There exist a number of interpretations of this setting. First, as described above, ω can be interpreted as R's private type. Second, the set of R's private types Ω can be viewed as a population of receivers. In this interpretation, S persuades the public to maximize the mass of those who choose to act. Third, one can consider a setting, in which R does not have a private type, but the action space is continuous. For example, suppose that R is matching the state ($u_R(a,\theta) = -(a-\theta)^2$) by taking a continuous action ($A = [0,1]$), and S has a state-independent utility function that is convex-concave in the action ($u_S(a,\theta) = H(a)$).[10] Then such a model is strategically equivalent to the one we study.

We analyze Perfect Bayesian Equilibria of the game. Without loss of generality, messages can be labeled so that they represent the corresponding posterior means. For example, in equilibrium, a message $m \in [0,1]$ induces a posterior mean that equals m.

[8] In particular, the cardinality of M is assumed to be at least that of $\operatorname{supp}\mu_0 = [0,1]$.

[9] The restriction to a single 'cheap-talk' (i.e. always available) message $\hat{m} = \varnothing$ is without loss of generality here.

[10] Dworczak and Martini (2019) provide an example of a continuous-action game in which the sender's objective is convex-concave.

Belief-based approach. Below we describe a framework that will be convenient for analyzing the equilibria of the game. It relies on the representation of information structures with convex functions, which has proven to be useful in information design literature (Gentzkow and Kamenica, 2016; Kolotilin, 2018). Although it might not seem as the most intuitive way of representing information, the investment into this framework will pay off. In particular, we will show that the voluntary disclosure game can be analyzed using the same approach. A unified treatment of all aspects of the model will then allow to solve the optimal evidence acquisition problem.

To characterize the equilibria of the game, we adopt the so-called belief-based approach. First, we solve for R's best response for a given posterior belief; then, we write S's payoff as an indirect utility function of R's posterior. This allows to treat R as a passive player who forms beliefs and express equilibrium conditions in terms of S's indirect utility function.

Moreover, R's best response depends on a posterior belief $\beta \in \Delta\Theta$ only through the mean $\mathbb{E}\beta$: [11]

$$a^*(\beta) := \mathbb{I}(\mathbb{E}\beta \geqslant \omega).$$

Therefore, it suffices to look at the **posterior mean** $\mathbb{E}\beta \in \Theta$.

We can now express S's interim payoff as an indirect utility function of the induced posterior mean. If S induces a posterior belief β with mean $\theta := \mathbb{E}\beta$, her interim (expected) payoff is

$$\int_0^1 u_S(a^*(\beta), \theta, \omega)\,\mathrm{d}H(\omega) = \int_0^1 \mathbb{I}(\mathbb{E}\beta \geqslant \omega)\,\mathrm{d}H(\omega) = H(\mathbb{E}\beta) = H(\theta).$$

So S's indirect utility function is exactly H, which measures the mass of R's types below the induced posterior mean.

[11] The tie-breaking rule here is without loss of generality.

Information structures as integral CDFs. Because only the posterior mean matters, each signal π can be associated with the corresponding **distribution over posterior means** $\mu_\pi \in \Delta\Theta$. We will identify a distribution over posterior means $\mu \in \Delta\Theta$ with its **integral CDF** (ICDF), which is an increasing convex function I_μ defined as the antiderivate of the CDF F_μ[12]

$$\begin{aligned} I_\mu : \mathbb{R}_+ &\to \mathbb{R}_+, \\ \theta &\mapsto \int_0^\theta F_\mu. \end{aligned}$$

Clearly, knowing I_μ, one can recover the CDF as the right derivative $(I_\mu)'_+ = F_\mu$.

To illustrate the approach, consider two extreme information structures: full information $\overline{\pi}$ and no information $\underline{\pi}$. Since $\overline{\pi}$ fully reveals the state, all posteriors are degenerate at the corresponding states, and the distribution over posterior means then coincides with the prior

$$\mu_{\overline{\pi}} = \mu_0.$$

Since $\underline{\pi}$ reveals no information, there is a unique posterior that is equal to the prior μ_0. This means that the corresponding distribution over posterior means is degenerate at the prior mean $\theta_0 = \mathbb{E}\mu_0$

$$\mu_{\underline{\pi}} = \delta_{\theta_0}.$$

Denote the integral CDFs of $\mu_{\overline{\pi}}$ and $\mu_{\underline{\pi}}$ as $\overline{I}$ and $\underline{I}$, respectively.

Fig. 1.3 below illustrates $\overline{I}$ and $\underline{I}$ for $\mu_0 \sim \mathcal{U}[0,1]$. Since $\mu_{\underline{\pi}} = \delta_{\frac{1}{2}}$ is degenerate at θ_0, the ICDF is piece-wise linear $I_{\mu_{\underline{\pi}}}(\theta) = (\theta - \frac{1}{2})^+ := \max(\theta - \frac{1}{2}, 0)$, where the kink at $\theta_0 = \frac{1}{2}$ with slope 1 corresponds to the point mass. The ICDF of $\mu_{\overline{\pi}} = \mathcal{U}[0,1]$ is the integral of a piece-wise linear function and is, therefore, quadratic: $I_{\mathcal{U}[0,1]}(\theta) = \frac{\theta^2}{2}$ on $[0,1]$.

[12] We omit the variable of integration whenever it is unambiguous, adopting the following notation: $\int_a^b f := \int_a^b f(x)\,\mathrm{d}x$, $\int_a^b f\,\mathrm{d}h := \int_a^b f(x)\,\mathrm{d}h(x)$.

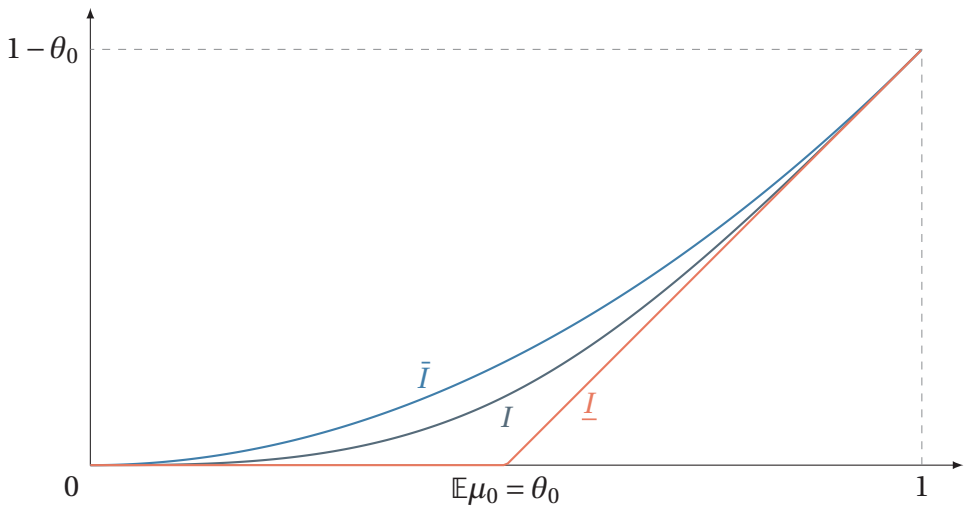


FIGURE 1.3: Integral CDFs of evidence structures, corresponding to full information $\bar{I}$, no information $\underline{I}$, and partial information I, for $\mu_0 \sim \mathcal{U}[0,1]$.

To describe the space of all information structures using this approach, define the informativeness order as follows. As is well known,[13] Blackwell informativeness order over information structures translates into **mean-preserving spreads** over distributions of posterior means. Formally, the partial order $\succcurlyeq_{\text{MPS}}$ is defined as

$$\mu' \succcurlyeq_{\text{MPS}} \mu'' \iff \left(I_{\mu'} \geqslant I_{\mu''} \text{ and } \mathbb{E}\mu' = \mathbb{E}\mu''\right).$$

Now since any information structure π is more informative than $\underline{\pi}$ and less informative than $\bar{\pi}$, it follows that $\bar{I} \geqslant I_{\mu_\pi} \geqslant \underline{I}$. Gentzkow and Kamenica (2016) and Kolotilin (2018) show that the converse also holds: for any convex function I, such that $\bar{I} \geqslant I \geqslant \underline{I}$, there exists an information structure π and a unique distribution over posterior means μ such that $I = I_\mu, \mu = \mu_\pi$. Define the set of ICDFs of all distributions over posterior means bounded between $\underline{I}$ and $\bar{I}$ as

$$\mathcal{I} := \{I : \mathbb{R}_+ \to \mathbb{R}_+ \mid I \text{ convex and } \bar{I} \geqslant I \geqslant \underline{I}\}.$$

[13] Rothschild and Stiglitz (1970) prove equivalence in the context of a risk averter's preferences over monetary lotteries and Leshno, Levy, and Spector (1997) provide a corrected proof of their result. Blackwell and Girshick (1954) prove a decision-theoretic equivalence result in the finite case.

Note that the requirement that the mean must be preserved is satisfied for any $I \in \mathcal{I}$, since $\overline{I}(1) = \underline{I}(1)$ and for any $\mu \in \Delta\Theta$

$$\mathbb{E}\mu = \int_0^1 \theta \, \mathrm{d}F_\mu(\theta) = 1 - \int_0^1 F_\mu(\theta) \, \mathrm{d}\theta = 1 - I_\mu(1).$$

Therefore, the informativeness ranking in $\mathcal{I}$ is represented with a simple point-wise inequality, i.e. partial order $\geqslant$.

This approach allows us to treat all information structures in a unified way. In particular, the signal chosen ex-ante by S and the distribution of R's posterior means (equivalently, evidence that is disclosed by S) can be both viewed as information structures and, therefore, can be represented with elements of $\mathcal{I}$. The approach of representing distributions over posterior means allows us to treat all information structures in a unified way. First, S's ex-ante choice of a signal corresponds to some distribution over posterior means and, therefore, can be represented with an element of $\mathcal{I}$. Second, what S's discloses, in equilibrium, corresponds exactly to the distribution of R's posteriors, which is then also an element $\mathcal{I}$.

1.3 Analysis

We analyze the model by backward induction. First, we fix an arbitrary evidence structure and solve the voluntary disclosure subgame. Next, we compute S's subgame (sequential) equilibrium value for a given evidence structure. Finally, we solve the optimal evidence acquisition problem and discuss properties of the optima.

1.3.1 Voluntary disclosure

In this section, we characterize equilibria of the voluntary disclosure subgame. That is, we derive the equilibrium disclosure strategy and the distribution of R's posteriors for an arbitrary evidence structure I.

Recall that S's indirect utility function coincides with the CDF H of R's cutoffs distribution and is, therefore, strictly increasing. It then follows that the equilibrium disclosure strategy is a threshold rule: S discloses the evidence if and only if it is sufficiently good.[14]

Lemma 1.1. *For any acquired evidence structure I, in any sequential equilibrium of the corresponding subgame, evidence is (not) disclosed if it induces a posterior mean above (below) the* ***disclosure threshold*** $\bar{\theta}_{q,I}$*, defined as*

$$qI(\bar{\theta}_{q,I}) = (1-q)(\theta_0 - \bar{\theta}_{q,I}),$$

In addition, this threshold $\bar{\theta}_{q,I}$ is decreasing in q and I (with respect to the informativeness order $\geqslant$) and unique if and only if $q \neq 1$ or $I(\theta) > 0$ for $\theta > 0$.

Lemma 1.1 tells us that whatever evidence structure S chooses ex-ante, she discloses only realizations that are "good enough". Intuitively, if $q = 1$, then R is certain that S has evidence and the standard unraveling argument of Grossman (1981) and Milgrom (1981) applies. Since R knows S has evidence, R's skepticism makes the highest type want to separate from all types, and so on for lower types. This means that $\bar{\theta}_{0,I} = 0$ for any I.

But when $q < 1$, R's skepticism is 'muted', which allows S to credibly conceal evidence. To understand how the threshold $\bar{\theta}_{q,I}$ is constructed, suppose, first, that S

[14] An equivalent model of voluntary disclosure was analyzed in Dye (1985) and Jung and Kwon (1988) for continuous distributions. Lemma 1.1 provides a unified treatment of general distributions, including distributions with atoms, e.g. discrete. Such a generalization will be useful in our context, since I is chosen endogenously at the ex-ante stage. Indeed, as can be seen from the equilibrium characterization below (Theorem 1.1), the optimal evidence structure might be discrete.

discloses any evidence she obtains. Then R's posterior mean after seeing message $\varnothing$ is θ_0. If I is not uninformative, then the worst evidence S might obtain is below θ_0, which means S prefers to conceal it. By iterating this argument, we arrive at a fixed point: if S uses the threshold strategy with $\bar{\theta}_{q,I}$, then the corresponding distribution of R's posteriors is such that the evidence inducing $\bar{\theta}_{q,I}$ makes S indifferent between disclosure and concealment. Note that $\bar{\theta}_{q,I}$ is decreasing in q, which means that as uncertainty grows, R's skepticism weakens and leads to less disclosure, for a fixed I.

Transformation of Information. It will be useful to think about S's strategic disclosure of information as a garbling of the acquired information structure. In particular, one can represent this garbling in terms of a mapping from S's chosen evidence structure into the induced R's distribution over posterior means. Since we identify evidence structures with distributions over posterior means, both objects can be represented as ICDFs. The following corollary characterizes the transformation of evidence structure due to voluntary disclosure.

Corollary 1.1. *For any acquired evidence structure I, there exists a unique subgame equilibrium* ***disclosed evidence structure****. Moreover, it is given by the following* ***voluntary disclosure transformation***

$$\begin{aligned} \mathscr{D}_q^V &: \mathcal{I} \to \mathcal{I}, \\ & I \mapsto \left[qI + (1-q)(\mathbf{id} - \theta_0) \right]^+, \end{aligned}$$

where $(\cdot)^+ := \max(\cdot, 0)$ *and* $\mathbf{id}$ *denotes the identity function* $\theta \mapsto \theta$.

Notice that the subgame equilibrium disclosed evidence $\mathscr{D}_q^V I$ is unique, even though the subgame equilibrium disclosure strategy may be non-unique. This is true for any $I \in \mathcal{I}$, even if there is an atom at $\bar{\theta}_{q,I}$, the point of indifference between disclosure and

concealment. The reason is that when S obtains evidence $\bar{\theta}_{q,I}$, both disclosure and non-disclosure lead to the same posterior mean and, consequently, the same $\mathscr{D}_q^V I$.

Benchmark: Mandatory disclosure. To understand the logic behind the voluntary disclosure transformation, it will be useful to compare it to the case of mandatory disclosure. That is, when S must reveal any evidence she obtains. In this case, with probability q, she obtains and discloses evidence and, with the remaining probability, she is uninformed and sends message $\varnothing$. Thus, the ICDF of R's posterior means is a convex combination of the chosen evidence structure I and the uninformative structure $\underline{I}$, given by

$$\begin{aligned}\mathscr{D}_q^M : \mathcal{I} &\to \mathcal{I},\\ I &\mapsto qI + (1-q)\underline{I}\\ &= qI + (1-q)(\mathbf{id} - \theta_0)^+.\end{aligned}$$

Fig. 1.4 illustrates the difference between the two transformations for a fixed I. First, notice that $\mathscr{D}_q^V I$ lies below $\mathscr{D}_q^M I$. Since $\geqslant$ represents the Blackwell order on $\mathcal{I}$, it means that $\mathscr{D}_q^M I$ is more informative than $\mathscr{D}_q^V I$. This is because $\mathscr{D}_q^M I$ represents the most information S can possibly disclose.

Moreover, the transformation of information from the acquired evidence I into the disclosed evidence $\mathscr{D}_q^V I$ can be seen as a two-stage garbling. First, the acquired evidence structure I is exogenously garbled into the available evidence structure $\mathscr{D}_q^M I$ because S obtains evidence only with probability q. Second, it is garbled again into the disclosed evidence structure $\mathscr{D}_q^V I$, due to the strategic concealment of unfavorable evidence.

Note that for $q < 1$ the mandatory disclosure transformation $\mathscr{D}_q^M I$ has a kink at θ_0, corresponding to the mass $1-q$ of uninformed sender types. This kink is due to the fact that none of the informed type pools with the uninformed type under mandatory

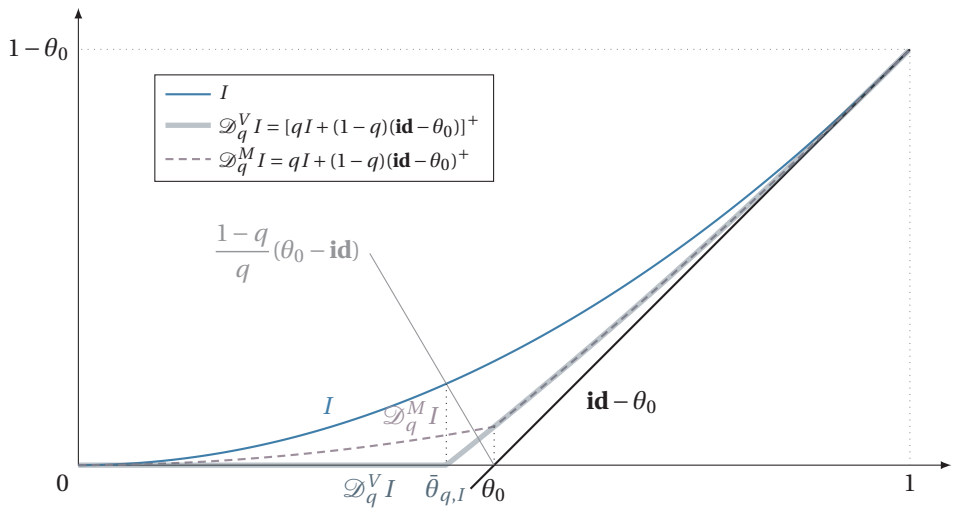


FIGURE 1.4: Construction of the disclosure threshold $\bar{\theta}_{q,I}$ and the voluntary and mandatory disclosure transformations $\mathscr{D}_q^V I$ and $\mathscr{D}_q^M I$.

disclosure. Compare this to the voluntary disclosure transformation $\mathscr{D}_q^V$. All types with evidence above θ_0 disclose it, which is why $\mathscr{D}_q^V$ coincides with $\mathscr{D}_q^M$ on $[\theta_0, 1]$. In contrast to mandatory disclosure, there is no mass point at θ_0 anymore, since the uninformed will be pooled with the low types and the corresponding posterior mean will be lower. This implies that $\mathscr{D}_q^V I$ continues below $\mathscr{D}_q^M I$ as a convex combination of I and $\mathbf{id} - \theta_0$. This convex combination reaches zero exactly at $\bar{\theta}_{q,I}$, which is where $\mathscr{D}_q^V I$ has a kink, corresponding to the combined mass of uninformed and low evidence types.

1.3.2 Value of Evidence

Before turning to the optimal evidence acquisition problem, we characterize S's value from evidence structures.

Suppose that S induces an ICDF of R's posterior means I. Given S's interim value function H, her S's ex-ante payoff simply the expectation of H with respect to the distribution corresponding to I. Equivalently, it can be written as $\int_0^1 H \, \mathrm{d}I'_+$, since the right derivative I'_+ corresponds to the CDF of R's posterior means. It will be convenient

to normalize the S's payoff from no information to zero and define the value as

$$\begin{aligned} v : \mathcal{I} &\to \mathbb{R}, \\ I &\mapsto \int_0^1 H \,\mathrm{d}(I'_+ - \underline{I}'_+). \end{aligned}$$

Integrating by parts twice, one can rewrite it as[15]

$$v(I) = \int_0^1 (I - \underline{I}) \,\mathrm{d}h.$$

Such a (Riemann–Stieltjes) integral representation implies that the S's value can be visualized as the "area" between I and $\underline{I}$ weighted by the measure induced by h. Fig. 1.5 illustrates this idea. Since h is increasing (decreasing) on $[0, \hat{\omega}]$ ($[\hat{\omega}, 1]$), it induces a positive (negative) measure on the corresponding interval. Thus, S's value is composed of the positive part $\int_0^{\hat{\omega}} (I - \underline{I}) \,\mathrm{d}h$ and the negative part $\int_{\hat{\omega}}^1 (I - \underline{I}) \,\mathrm{d}h$. This implies, in particular, that S benefits from more information at the bottom and less information at the top.

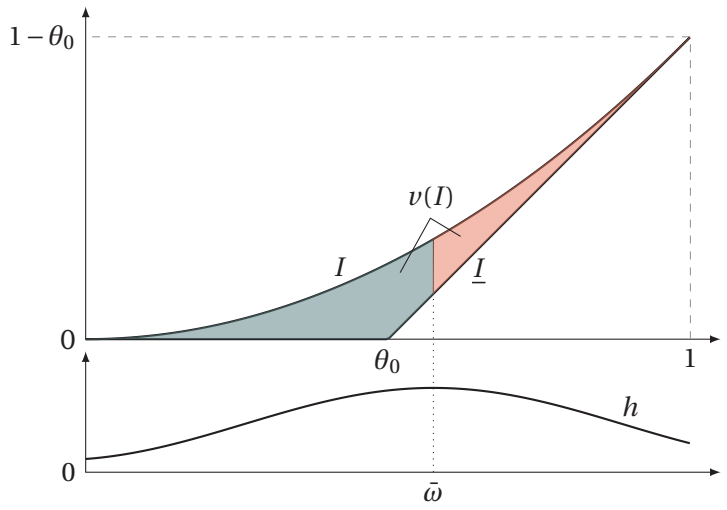


FIGURE 1.5: $v(I)$ as a sum of a positive and a negative part.

[15]All integrals are Riemann–Stieltjes. For any continuous g, we define $\int_0^1 g \,\mathrm{d}h$ as the difference $\int_0^{\hat{\omega}} g \,\mathrm{d}h - \int_{\hat{\omega}}^1 g \,\mathrm{d}(-h)$ of two Riemann–Stieltjes integrals with respect to strictly increasing functions.

We can now characterize S's expected payoff for any acquired evidence structure I.

Lemma 1.2. *The sender's value from acquisition of evidence structure I under mandatory disclosure is given by*

$$v(\mathscr{D}_q^M I) = q\, v(I),$$

and that under voluntary disclosure is given by

$$v(\mathscr{D}_q^V I) = q(v(I) - L_q(I)),$$

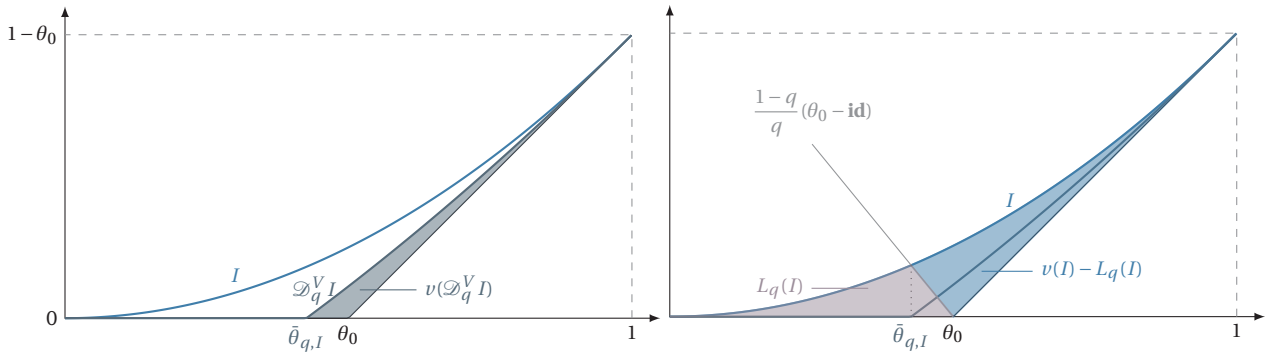
where L_q is the ***concealment loss*** *defined as*

$$L_q(I) := \int_0^{\bar{\theta}_{q,I}} I \,\mathrm{d}h + \int_{\bar{\theta}_{q,I}}^{\theta_0} \frac{1-q}{q}(\theta_0 - \mathbf{id})\,\mathrm{d}h.$$

This result highlights the difference between voluntary and mandatory disclosure in terms of the effect of uncertainty on the value from acquisition of evidence. For a fixed I, higher uncertainty shifts the available evidence $\mathscr{D}_q^M I$ down towards the uninformative structure $\underline{I}$. Since the value from $\underline{I}$ is normalized to zero and all available evidence is disclosed under mandatory disclosure, q enters as a multiplier in the expression for $v(\mathscr{D}_q^M I)$.

The same effect is retained under voluntary disclosure, but there is an additional term L_q due to strategic disclosure, which we call the concealment loss. As the uncertainty increases, the disclosure threshold $\bar{\theta}_{q,I}$ increases as well. Since ex-ante S dislikes less information at the bottom, she incurs the loss.

Fig. 1.6 illustrates the decomposition of S's acquisition value. As Lemma 1.2 shows, the shaded "area" $v(\mathscr{D}_q^V I)$ in Fig. 1.6a must be equal to q times the shaded "area" $v(I) - L_q(I)$ in Fig. 1.6b.



(a) $v(\mathscr{D}_q^V I)$ (b) $v(I) - L_q(I)$

FIGURE 1.6: Decomposition of the value of evidence

1.3.3 Optimal Evidence Acquisition

In this section, we endogenize the evidence structure as S's ex-ante choice. She designs the evidence structure strategically to mitigate the effect of voluntary disclosure.

Before we characterize the equilibrium evidence structure, it will be instructive to look at the extreme case $q = 1$, in which S always obtains evidence. Recall that under $q = 1$ there is unraveling: the receiver's skepticism makes the sender always disclose. In this case, the voluntary disclosure transformation leaves I unchanged, the acquisition problem becomes equivalent to the problem of Bayesian persuasion. Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017) and Kolotilin (2018) study a similar model of Bayesian persuasion with R's private payoff type and show, in particular, that if the distribution of R types is unimodal, the optimal signal is an *upper censorship*: it perfectly reveals all states below and pools all states above some threshold.[16] The intuition behind this result is the following. As discussed in the previous section, when the state is low (high), more information benefits (hurts) S because S's indirect utility function H is convex-concave. It turns out that the optimal signal simply reveals (pools) all states below (above) some threshold $\theta_1^* \in [0, \hat{\omega}]$.

[16]Optimality of upper censorship in slightly different settings also appears in Alonso and Câmara (2016b) and Dworczak and Martini (2019).

Now consider the case of $q < 1$. It will be useful to define the following class of information structures.[17]

Definition 1.1. An evidence structure $I \in \mathcal{I}$ is a (θ_l, θ_h) **two-sided censorship** if and only if there exist $0 \leqslant \theta_l \leqslant \theta_h \leqslant 1$, such that

$$I(\theta) = \begin{cases} \max(\overline{I}(\theta_l) + \overline{I}'(\theta_l)(\theta - \theta_l), \underline{I}), & \theta < \theta_l, \\ \overline{I}(\theta), & \theta \in [\theta_l, \theta_h], \\ \max(\overline{I}(\theta_h) + \overline{I}'(\theta_h)(\theta - \theta_h), \underline{I}), & \theta > \theta_h. \end{cases}$$

In addition, call I

- a θ_h **upper censorship**, if $\theta_l = 0$,
- a θ_l **lower censorship**, if $\theta_h = 1$, and
- a θ **binary certification**, if $\theta_l = \theta_h = \theta \in (0, 1)$.

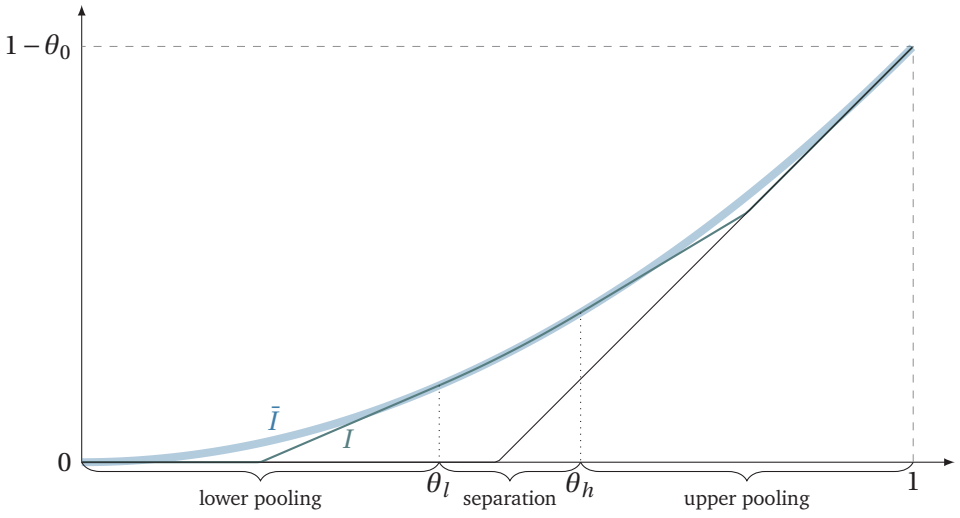


FIGURE 1.7: Two-sided censorship.

In words, an evidence structure I is a two-sided censorship if it perfectly reveals all states in $[\theta_l, \theta_h]$, pools states above θ_h and pools states below θ_l. It can be interpreted

[17] Kolotilin (2018) introduces an equivalent class of information structures called interval revelation mechanisms.

as a grading system that assigns the PASS grade to the states above the upper cutoff, the FAIL grade to the states below the lower cutoff, and has a number of intermediate grades.

Consider some special cases. First, note that if $\theta_l = 0$ and $\theta_h = 1$, both pooling intervals are empty. This corresponds to the case of the fully informative structure $\overline{I}$. Second, if $\theta_l = \theta_h \in \{0, 1\}$, then all states are pooled, which corresponds to the uninformative structure $\underline{I}$. Next, if the lower pooling intervals is empty ($\theta_l = 0$), then all states below θ_h are revealed, which corresponds to an upper censorship. Finally, if $\theta_l = \theta_h$, then the evidence structure reveals only whether the state is above or below $\theta_l = \theta_h$ and produces exactly two messages (with probability 1). We call such an evidence structure binary certification.

Before stating the main result, we discuss the multiplicity of equilibria that arises in the model. Call two evidence structures disclosure-equivalent if they induce the same disclosed evidence structure. Clearly, the corresponding equivalence classes are given by pre-images of $\mathscr{D}_q^V$. Fig. 1.8 illustrates the set of all evidence structures that induce a given disclosed evidence J. Note that the set $\left(\mathscr{D}_q^V\right)^{-1} J$ is an "interval" $[I_*, I^*] := \{I \in \mathcal{I}, I_* \leqslant I \leqslant I^*\}$ of evidence structures that coincide on $[\bar{\theta}_{q,I}, 1]$ and have the same disclosure threshold. As the S's value depends only on $\mathscr{D}_q^V I$, it follows that the set of equilibrium evidence structures consists of such "intervals". In other words, the sender can always acquire more or less evidence about states below the disclosure threshold $\bar{\theta}_{q,I}$, without changing the outcome of the game.

Note that this implies the following "revelation principle": for every equilibrium of the game, there exists a "canonical" outcome-equivalent equilibrium, in which there is a unique realization of a signal that is concealed by S. To ease the exposition of the results, we will focus on equilibria of the latter type and define the notion of optimal evidence structures as follows.

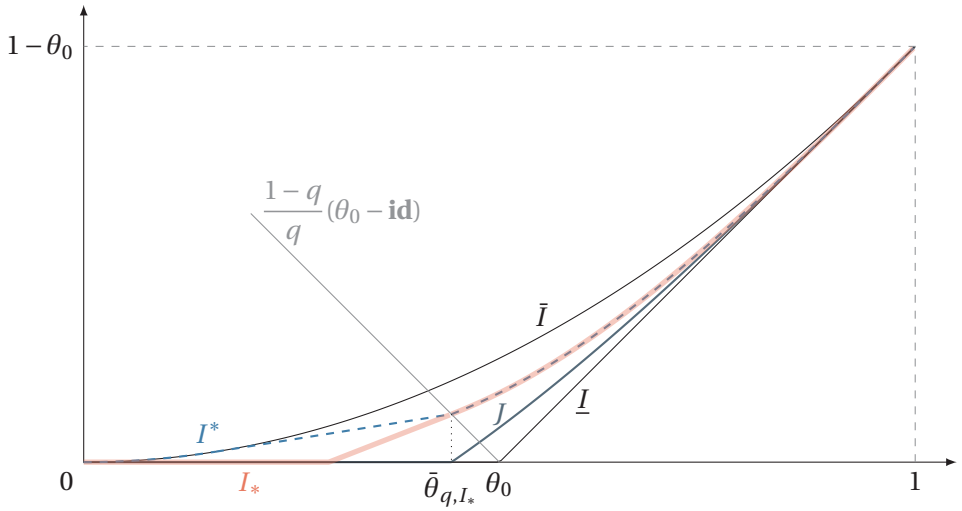


FIGURE 1.8: An interval $\left(\mathscr{D}_q^V\right)^{-1} J = [I_*, I^*]$ of disclosure-equivalent evidence structures corresponding to a disclosed evidence J.

Definition 1.2. An evidence structure I^* is called *optimal*, if it solves

$$v_q^* := \max_{I \in \mathcal{I}} v(\mathscr{D}_q^V I), \tag{*}$$

and there is no $\tilde{I} \in \mathcal{I}$, such that $I^* \geqslant \tilde{I}, I^* \neq \tilde{I}$ and $\mathscr{D}_q^V I^* = \mathscr{D}_q^V \tilde{I}$.

The following theorem provides characterization of optimal evidence structures.

Theorem 1.1. *An optimal evidence structure exists. There exists $\overline{q} \in [0,1)$, such that if $q < \overline{q}$, then any optimum is a binary certification. Moreover, if $q > \overline{q}$, then the unique optimum is the $(\bar{\theta}_{q,\bar{I}}, \theta_1^*)$ two-sided censorship.*

This result shows that the optimal evidence structure depends on the probability that evidence is obtained. Moreover, it shows that the interaction between the forces that drive pooling at the top and bottom of the state distribution can take different forms. When the uncertainty is low ($q > \overline{q}$), the interaction between the two forces is trivial and optimal evidence structure is a two-sided censorship of the state. The lower threshold $\bar{\theta}_{q,\bar{I}}$ is not affected by the design of the evidence structure and coincides with the disclosure threshold under fully-revealing evidence structure. Moreover, the upper

threshold θ_1^* is unaffected by voluntary disclosure: it stays constant and coincides with the optimal upper threshold that the sender would use under $q = 1$.

However, when uncertainty is high ($q < \overline{q}$), the interaction between the two forces becomes non-trivial and the sender adopts binary certification. Notice that voluntary disclosure leads to pooling of low states. From the ex-ante perspective, this hurts the sender because her interim payoff function is convex at the bottom. Therefore, she would commit to reveal low states, but cannot because disclosure is voluntary. When the q drops below $\overline{q}$, it becomes optimal to design evidence structure in order to reduce the ex-ante loss from non-disclosure of low states. This is achieved by binary certification, as it allows to reduce the lower pooling interval by enlarging the upper pooling interval.

The proof is given in Appendix A.1 and based on constructing a one-dimensional optimization problem that is equivalent to (*). We present the main idea below. First, Lemma 1.2 implies that the sender's ex-ante problem can be written as

$$\max_{I \in \mathcal{I}} v(\mathscr{D}_q^V I) = q \max_{I \in \mathcal{I}} \left(v(I) - L_q(I) \right).$$

Second, we show that any maximizer of $v - L_q$ must coincide with some upper censorship on $[\bar{\theta}_{q,I}, 1]$, generalizing a standard argument used in the extreme case of $q = 1$. Intuitively, if $q < 1$, there is pooling at the bottom due to strategic disclosure. Even though the pooling interval is determined endogenously, it follows from the geometrical characterization of S's value that any upper censorship that is an improvement under $q = 1$ will still be an improvement under $q < 1$. This allows to formulate the evidence acquisition problem as a one-dimensional optimization problem

$$\max_{\theta \in [0,1]} v(I_\theta) - L_q(I_\theta), \qquad (**)$$

where I_θ is the θ upper censorship. Then, the definition of an optimum implies that it must be the $\bar{\theta}_{q,I}$ lower censorship of I_θ. If $\theta > \bar{\theta}_{q,I}$, it is the $(\bar{\theta}_{q,I}, \theta)$ two-sided censorship, otherwise, it is the θ binary certification.

Next, we show that $\theta \mapsto v(I_\theta)$ has a unique maximum and that $\theta \mapsto L_q(I_\theta)$ is constant on $[\bar{\theta}_{q,I}, 1]$. The threshold $\overline{q}$ is identified as the lowest value of q, such that the loss $L_q(I_\theta)$ does not affect the maximizer and thereby obtain the second part of the result. Finally, we show that the marginal concealment loss is decreasing in q. This implies that, for $q < \overline{q}$, we have $\theta < \bar{\theta}_{q,I}$, which implies that the optimum, given by the $\bar{\theta}_{q,\overline{I}}$ lower censorship of I_θ is a binary certification.

1.3.4 Degradation of Certification Standards

In this section, we study how optimal binary certification threshold depends on the probability of obtaining evidence q. The role of a binary certification threshold is twofold. First, it serves as a certification standard because only the states above it get a passing grade. Second, it bounds the lower pooling region, determining the states that are going to be concealed. This is in contrast to a two-sided censorship, when the two pooling regions are controlled by different thresholds.

Note that when the optimum is a two-sided censorship, the lower pooling threshold coincides with the disclosure threshold of the fully-revealing information structure $\overline{I}$. This, together with Lemma 1.1, implies that as q decreases, the pooling interval becomes larger. As follows from a standard argument, as uncertainty increases, R becomes less skeptical when S claims to not have obtained any evidence. This incentivizes S to conceal evidence, and, in equilibrium, leads to more pooling at the bottom.

But this argument valid for a fixed evidence structure no longer applies in the case of binary certification because the two forces shaping the optimal evidence structure have non-trivial interaction. This leads to the reversal of the effect of higher uncertainty. S is strategically choosing a binary threshold that is below the disclosure threshold to mit-

igate the effect of voluntary disclosure problem. With lower q this problem becomes more severe and lower thresholds become more effective. Note that our decomposition of S's value implies that q affects the optimal choice of information only through the concealment loss. But lower thresholds reduce the concealment loss more when uncertainty is higher. This implies that the optimal binary certification threshold will is increasing in q, as summarized by the following result.

Theorem 1.2. *Let $\theta^*_{q_1}$ and $\theta^*_{q_2}$ be optimal binary certification thresholds for q_1 and q_2, respectively. Then $q_1 < q_2$ implies $\theta^*_{q_1} < \theta^*_{q_2}$.*

Theorem 1.2 highlights that the interaction of the two forces that lead to pooling at the top and bottom of state distribution becomes non-trivial when q drops below $\overline{q}$. Because S reduces the lower pooling interval to be able to credibly disclose more good states, the effect of uncertainty on the lower pooling interval is reversed, compared to the case of two-sided censorship. As can be seen in Fig. 1.2, the lower threshold is non-monotone in q.

The main idea of the proof is the following. Recall that the sender's objective function is the difference between the value function v and the concealment loss L_q, where only L_q depends on q. Since S's problem can be represented as a one-dimensional program (**), it is sufficient to show that the concealment loss satisfies strictly decreasing marginal differences property (Edlin and Shannon, 1998). That is, we show that the marginal increase in the concealment loss from an increasing the threshold is decreasing in q by using the integral representation of L_q given in Lemma 1.2. As the uncertainty decreases, the sender discloses more evidence at the bottom, so the marginal concealment loss is lower.

Uniqueness. Note that neither of the above results establishes the uniqueness of the optimum for $q \in (0, \overline{q})$. However, the strict comparative statics of Theorem 1.2 implies uniqueness for almost all q. To see this, note that any selection from the optimal binary

certification threshold correspondence must be strictly decreasing on $q \in (0, \overline{q})$. But then this selection can have at most a countable set of points of discontinuity. Therefore, the optima correspondence is single-valued almost everywhere. One can interpret this result as establishing that uniqueness of the solution holds generically across q.

Corollary 1.2. *The optimal evidence structure is unique for almost all* $q \in (0, 1]$.

More precisely, there exists a subset $\mathcal{Q} \subseteq (0, 1]$ with a countable complement, such that $\mathcal{Q} \supset (\overline{q}, 1]$ and for any $q \in \mathcal{Q}$ there is a unique optimal evidence structure. Henceforth, denote the unique optimum as I_q^* for $q \in \mathcal{Q}$ and the unique optimal binary certification threshold as θ_q^* for $q \in \mathcal{Q} \cap (0, \overline{q}]$. Note that Theorem 1.2 then implies that θ_q^* is strictly increasing in q on $\mathcal{Q} \cap (0, \overline{q}]$.

1.3.5 Voluntary vs Mandatory Disclosure

In this section, we compare optimal evidence acquisition under voluntary and mandatory disclosure. How does inability of S to commit to full disclosure affect optimal evidence acquisition?

To answer this question, consider S's problem under mandatory disclosure. Lemma 1.2 allows to write it as

$$\max_{I \in \mathcal{I}} v(\mathcal{D}_q^M I) = \max_{I \in \mathcal{I}} q\, v(I) = q\, v_1^*.$$

But this implies that the optimum under mandatory disclosure is the same as under no uncertainty, equivalently, when $q = 1$. Thus, the following proposition holds.

Proposition 1.1. *For any q, the optimum under mandatory disclosure coincides with the optimum under voluntary disclosure with* $q = 1$.

The intuition is the following. Under mandatory disclosure, S does not always obtain evidence. But when she does, it is necessarily fully revealed. Therefore, she can

simply maximize her value conditional on obtaining evidence, which is equivalent to solving the evidence acquisition problem under $q = 1$.

Note that Proposition 1.1 implies that (i) any optimal binary certification threshold is strictly lower than the optimal upper censorship certification under mandatory disclosure and (ii) the mandatory disclosure optimum is strictly more informative than any voluntary disclosure optimum under $q < 1$. To see this, note that Theorem 1.1 implies that the optimum under mandatory disclosure is the θ_1^* upper censorship I_1^*. Now consider any optimal θ-binary certification I_θ. First, by Corollary 1.2, θ must necessarily be below θ_1^*. To see why I_θ is a garbling of I_1^*, consider the θ upper censorship J_θ and note that $I_1^* > J_\theta \geqslant I_\theta$. In other words, J_θ provides less information than I_1^* because it pools more states at the top, and more information than I_θ because it doesn't pool the states below θ. Finally, any optimal two-sided censorship is a garbling of I_1^* because it has the same upper threshold, but also has pooling at the bottom.

1.3.6 Welfare

How does the level of uncertainty affect the players welfare? In this section we show that both players' ex-ante expected equilibrium payoffs are strictly increasing in q. This comparative statics result holds for the two players for distinct reasons. The monotonicity of S's payoff follows directly from the properties of the objective function in her optimal acquisition problem. However, the monotonicity of R's payoff follows from the characterization of the optimal evidence structures.

Such welfare analyses compare environments with different probabilities of obtaining evidence. This means that, for example, S's equilibrium value increases in q partly because she gets an opportunity to persuade R more often. Therefore, a sensible way to compare welfare under different levels of uncertainty in the model is to compare payoffs normalized by the probability of obtaining evidence, which we call *normalized value*. We then strengthen the result by showing that the normalized payoffs of

both players are also strictly increasing in q. In other words, there are two channels through which higher q improves players' welfare: communication happens more often and more efficiently.

The normalized value can also be interpreted as the fraction of the value that is achieved under mandatory disclosure. To see this, recall that S's equilibrium value under mandatory disclosure is given by qv_1^*. Therefore, the normalized value is proportional to the ratio

$$\frac{v_q^*}{qv_1^*} = \frac{\max_{I\in\mathcal{I}} v(\mathscr{D}_q^V I)}{\max_{I\in\mathcal{I}} v(\mathscr{D}_q^M I)}.$$

Next we provide detailed analysis of both players' welfare.

Sender. First, consider S's value v_q^*. An immediate observation is that whatever distribution of posterior beliefs S can induce in equilibrium under lower q, she can also implement under higher q. Equivalently, the set $\mathscr{D}_q^V\mathcal{I} = \{\mathscr{D}_q^V I : I \in \mathcal{I}\}$ of all evidence structures that can be voluntary disclosed is monotone in q with respect to set inclusion. Thus, S's equilibrium value v_q^* is increasing in q.

The following proposition shows that not only S's ex-ante value, but also S's normalized value is strictly increasing in q.

Proposition 1.2. *Both S's value v_q^* and normalized value $\frac{v_q^*}{q}$ are strictly increasing in q.*

The proof is by inspection of the derivative. We apply Lemma 1.2 to rewrite S's normalized value as

$$\frac{v_q^*}{q} = \max_{I\in\mathcal{I}} v(I) - L_q(I).$$

The Envelope Theorem implies that the sign of the derivative of the normalized value is determined by the sign of the derivative of the concealment loss L_q. To see why L_q is decreasing in q, note that, as q increases, R becomes more skeptical if S claims to be uninformed as he is more certain that S obtains evidence. In equilibrium,

this leads to a lower disclosure $\bar{\theta}_{q,I}$. But this benefits S in expectation, since ex-ante she prefers to disclose more information at the bottom.

Receiver. To define R's ex-ante value function, note that the payoff of type ω with posterior mean θ is given by $(\theta-\omega)^+$. Therefore, the aggregate interim payoff is $\int(\theta-\omega)^+\,\mathrm{d}H(\omega)$. Now define R's ex-ante value function of induced distributions of posterior means as

$$\begin{aligned} &w:\mathcal{I}\to\mathbb{R},\\ &I:\mathcal{I}\mapsto\int_\Theta\int_\Omega(\theta-\omega)^+\,\mathrm{d}H(\omega)\,\mathrm{d}(I'_+-\underline{I}'_+)(\theta). \end{aligned}$$

Note that the derivative of the inner integral with respect to θ is given by

$$\int\frac{\mathrm{d}}{\mathrm{d}\theta}(\theta-\omega)^+\,\mathrm{d}H(\omega)=\int\mathbb{I}(\theta\geqslant\omega)\,\mathrm{d}H(\omega)=\int_0^\theta\mathrm{d}H(\omega)=H(\theta).$$

Integrating by parts twice, rewrite R's value function as

$$w(I)=\int(I-\underline{I})\,\mathrm{d}H.$$

Clearly, Blackwell Theorem implies that w is weakly increasing in I with respect to $\geqslant$. But notice that w is also strictly increasing with respect to our strict informativeness order $>$. Applying Lemma 1.2 to w, we obtain the following representation of R's value from an acquired evidence structure I:

$$w(\mathscr{D}_q^V I)=q(w(I)-L_q(I)).$$

Finally, define R's equilibrium value as

$$w_q^*=w(\mathscr{D}_q^V I_q^*)=q(w(I_q^*)-L_q(I_q^*)),$$

for $q \in \mathcal{Q}$, so that it is uniquely defined (by Corollary 1.2).

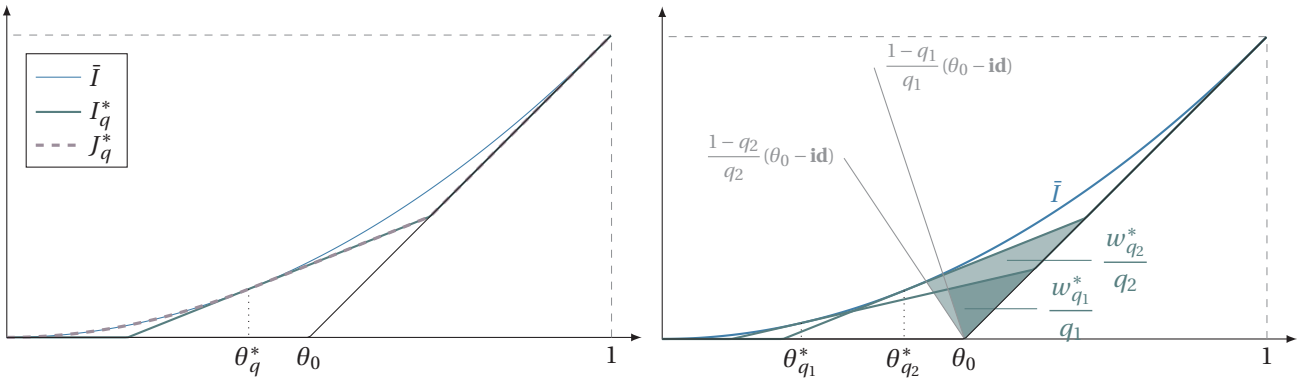
How does q affect w_q^* and $\frac{w_q^*}{q}$? In contrast to the S's value v_q^*, the properties of w_q^* do not follow from properties of the objective function in an optimization problem. We, therefore, need to analyze how the solution $\mathscr{D}_q^V I_q^*$ depends on q. What is, perhaps, surprising is that the comparative statics of R's welfare is similar to that of S, as the following proposition shows.

Proposition 1.3. *Both R's value w_q^* and conditional value $\frac{w_q^*}{q}$ are strictly increasing in q.*

To get some intuition, consider, first, the case of low uncertainty ($q > \overline{q}$). As we know from Theorem 1.1, the unique optimal evidence structure I_q^* is the $(\bar{\theta}_{q,\bar{I}}, \theta_1^*)$ two-sided censorship. As q increases, less states are pooled at the bottom, which means that I_q^* is $\geqslant$-increasing in q. In addition, notice that $\mathscr{D}_q^V$ is $\geqslant$-increasing in q and in I (with respect to $\geqslant$). That is more acquired evidence and less uncertainty leads to more disclosed evidence. Thus, $\mathscr{D}_q^V I_q^*$ is $\geqslant$-increasing in q, and, therefore, so is w_q^* on $(\overline{q}, 1]$. Moreover, it is straightforward to check that $\mathscr{D}_q^V I_q^*$ is in fact strictly $>$-increasing in q.

Now consider the case of high uncertainty ($q \in \mathcal{Q} \cap (0, \overline{q}]$). Theorem 1.2 implies that the optimal binary certification threshold θ_q^* strictly increases in q. Note that any two different binary certification evidence structures are incomparable in the sense of Blackwell, since a lower threshold provides more information about low states and less information about high states. Moreover, if we consider two binary certifications with relatively high thresholds, then even their disclosure transformations will be incomparable. However, on $\mathcal{Q} \cap (0, \overline{q}]$, the disclosure transformation of the optimal binary certifications is $>$-increasing in q, as can be clearly seen from Fig. 1.9a. This is for any binary certification I_q^*, there is a disclosure-equivalent θ_q^* upper censorship J_q^* that is $>$-increasing in q.

As we discussed above, ex-ante value increases in q in part because lower uncertainty provides more means for mutually beneficial information transmission between S and R. Thus, one can be interested in the relative efficiency of information transmis-



(a) The construction of a disclosure-equivalent upper censorship J^*_q for a given I^*_q.

(b) The effect of higher uncertainty on R's conditional value.

FIGURE 1.9: Decomposition of the value of evidence.

sion, which we quantify with the conditional value $\frac{w^*_q}{q}$. Fig. 1.9b illustrates the effect on R's conditional value from as uncertainty increases ($q_2 \to q_1$). First, the middle straight part of $I^*_{q^*}$ rotates, which might potentially provide more value for the receiver about low states. But since those states among those the sender conceals, the receiver suffers from higher concealment loss, which erases all potential benefits.

1.4 Conclusion

This paper endogenizes evidence structures in a game of voluntary disclosure. The main contribution is twofold. First, we show that a combination of design and disclosure incentives can lead to verifiable information taking a simple form of binary certification. Second, we show that the non-trivial interaction between these two incentives leads to a reversal of the effect of uncertainty on the set of concealed states. We also show that higher probability of obtaining evidence benefits both players, not just because it allows the sender to communicate more often, but also because she does so more efficiently.

Chapter 2

Persuasion via Weak Institutions

joint with Elliot Lipnowski and Doron Ravid

2.1 Introduction

Many institutions routinely collect and disseminate information. Although the collected information is instrumental to its consumers, the goal of dissemination is often to persuade. Persuading one's audience, however, requires the audience to believe what one says. In other words, the institution must be credible, capable of delivering both good and bad news. Delivering bad news might be especially difficult, requiring the institution to withstand pressure exerted by its superiors. The current paper studies how an institution's susceptibility to such pressures influences its persuasiveness and the quality of the information it provides.

We study a persuasion game between a receiver (R, he) and a sender (S, she) who cares only about R's chosen action. The game begins with S publicly announcing an official reporting protocol, which is a Blackwell experiment about the state. After the announcement, S privately learns the state and whether her reporting protocol is credible. If credible, R observes a message drawn from the announced reporting protocol. Otherwise, S can freely choose the message that R sees. R then takes an action, not knowing the message's origin. Given state θ, reporting is credible with probability $\chi(\theta)$, a probability that we interpret as the strength of S's institution in said state.

As in the recent Bayesian persuasion literature (e.g., Kamenica and Gentzkow, 2011; Alonso and Câmara, 2016a; Ely, 2017), we view S as a principal, capable of steering R toward her preferred equilibrium. Our main result (Theorem 2.1) characterizes S's highest equilibrium payoff . The characterization is geometric and is based on S's *value function*, which specifies the highest value S can obtain from R responding optimally to a given posterior belief. Under full credibility ($\chi(\theta) = 1$ for all θ), our model is equivalent to the one studied by Kamenica and Gentzkow (2011). As such, in this case, S's highest equilibrium value is given by the concave envelope of S's value function. The value function's quasiconcave envelope gives S's highest value under cheap talk (see Lipnowski and Ravid (2019)), and therefore S's highest equilibrium value

under no credibility ($\chi(\theta) = 0$ for all θ). For intermediate credibility values, Theorem 2.1's characterization combines the quasiconcave envelope of S's value function and the concave envelope of S's *capped* value function, which captures S's incentive constraints.

Using our characterization, we analyze how S's and R's values change with $\chi(\cdot)$. To illustrate, consider a multinational firm (R) that can make a large investment ($a = 1$), a small investment ($a = \frac{1}{2}$), or no investment ($a = 0$) in a small open economy. Profits from each investment level depend on the state of the economy, θ, which can be good ($\theta = 1$), or bad ($\theta = 0$) with equal probability. In particular,[1]

$$u_R(a, \theta) = a\theta - \tfrac{1}{2}a^2.$$

Because the state of the world is binary, the firm's beliefs can be identified with the probability that the economy is good, μ. Given the above preferences, no investment is optimal when $\mu \leqslant \frac{1}{4}$; a large investment is optimal when $\mu \geqslant \frac{3}{4}$; and a small investment is optimal when $\mu \in \left[\frac{1}{4}, \frac{3}{4}\right]$. A local policymaker (S) wants to maximize the firm's investment, and receives a payoff of 0, 1, and 2 from no, small, and large investments, respectively. To persuade the firm, the policymaker publicly commissions a report by the central bank. Formally, a report is a Blackwell experiment producing a stochastic investment recommendation conditional on the economy's state.[2] The reliability of this recommendation is questionable, as it is produced by the announced experiment only with probability x, independent of the state. With probability $1 - x$, the bank succumbs to the policymaker's pressure, producing the policymaker's recommendation of choice.

Proposition 2.1 shows R is often better off with a less credible S. The proposition applies to the above example. To see this, suppose first that the bank's report is fully credible, that is $x = 1$. In this case, the optimal report recommends either a large or a small investment with equal ex-ante probability in a way that makes the firm just will-

[1] An alternative, behaviorally equivalent specification has $u_R(a, \theta) = -(a - \theta)^2$.

[2] Restricting attention to such experiments in this example turns out to be without loss.

ing to accept each recommendation. In other words, the firm's posterior belief that the state is good is uniformly distributed on $\{\frac{1}{4}, \frac{3}{4}\}$, with the firm making a large investment when its belief is $\frac{3}{4}$, and a small investment otherwise. In this case, the firm's expected utility is $\frac{1}{8}$. Consider now a weaker central bank, capable of resisting the policymaker's pressure with a lower probability of $x = \frac{2}{3}$. Take any report that leads to an incentive-compatible large investment recommendation with positive probability. Because the policymaker gets to secretly influence the report with probability $1 - x = \frac{1}{3}$, the report produces a large investment recommendation with a probability of at least $\frac{1}{3}$, *regardless of the state*. By Bayes' rule, conditional on such a recommendation, the firm's posterior belief that the state is good is no greater than $\frac{3}{4}$. Note this upper bound can be achieved only if the bank's official report fully reveals the state. Hence, the report must generate a "no investment" recommendation whenever the economy is bad and reporting is uninfluenced (which happens with probability $\frac{1}{3}$), and a "large investment" recommendation otherwise. This policy is strictly better for the policymaker than conveying no information (which yields a small investment with certainty), and so is the policymaker's unique preferred equilibrium. Thus, when $x = \frac{2}{3}$, the firm's expected utility is $\frac{1}{6}$. In particular, the firm strictly benefits from a weaker central bank; that is, productive mistrust occurs.

Our next result, Proposition 2.2, shows that small decreases in credibility lead to large drops in the sender's value for all interesting instances of our model. More precisely, we show such a collapse occurs at some full-support prior and some credibility level if and only if S can benefit from persuasion. Such a collapse is clearly present in our example: Given the preceding analysis, $\frac{2}{3}$ is the lowest credibility level that allows the bank to credibly recommend a large investment. For any $x < \frac{2}{3}$, the policymaker can do no better than have the bank provide no information to the firm, giving the policymaker a payoff of $\frac{1}{2}$. Because $\frac{2}{3}$ is the policymaker's payoff when $x = \frac{2}{3}$, even an infinitesimal decrease in credibility results in a discrete drop in the policymaker's value.

One can construct examples in which S's value collapses at full credibility. For example, suppose the firm can make a very large investment, which yields a payoff of 10 to the policymaker, and is optimal if and only if the firm is *certain* the economy is good. Under full credibility, the policymaker can obtain a payoff of 5 by revealing the state and having the central bank recommend no investment when the economy is bad and a very large investment when the economy is good. A very large investment recommendation, however, is never credible for any $x < 1$. If it were, the policymaker would always send it when influencing the bank's report, regardless of the economy's state, and so the firm could never be completely certain that the economy's state is good. As such, the policymaker's optimal equilibrium policy for any $x \in [\frac{3}{4}, 1)$ remains as it was in the unmodified example, giving her a payoff of $\frac{3}{4}$. Thus, even a tiny imperfection in the central bank's credibility causes the policymaker's payoff to drop from 5 to $\frac{3}{4}$.

One may suspect the non-robustness of the full-credibility solution in the above modified example is rather special. Proposition 2.3 confirms this suspicion. In particular, it shows S's value can collapse at full credibility if and only if R does not give S the benefit of the doubt; that is, to obtain her best feasible payoff, S must persuade R that some state is impossible. This property is clearly violated in the above modified example: The firm is willing to make a very large investment only if it assigns a zero probability to the economy's state being bad. Thus, although S's value often collapses due to small decreases in credibility, such collapses rarely occur at full credibility.

Section 2.5 abandons our general analysis in favor of a specific instance of public persuasion, which enables us to assess the relative value of credibility in different states. In this specification, S uses her weak institution to release a public report whose purpose is to sway a population of receivers to take a favorable binary action. For example, S may be a seller who markets her product by sending it to reviewers or a leader vying for the support of her populace using state-owned media. Each receiver's utility from taking S's favorite action is additively separable in the unknown state and his idiosyn-

cratic type, which follows a well-behaved single-peaked distribution. We show (Claim 2.1) it is S-optimal for the official report to take an upper-censorship form, characterized by a threshold below which states are fully separated. States above this threshold are pooled into a single message, which is always sent when S influences the report. We also show that concentrating the credibility of S's institution in low states uniformly increases S's payoffs across all type distributions (Claim 2.2). Hence, S especially prefers her institution to be resistant to pressure in bad states.

To conclude our analysis, we allow S to design her institution at a cost. More precisely, we let S publicly choose the probability with which reporting is credible in each state. S's credibility choice is made in ignorance of the state, and comes at a cost that is a continuous and increasing function of the institution's average credibility. We explain how to adjust our analysis to this setting, and observe that R may benefit from an increase in S's costs, echoing the productive-mistrust phenomenon of the fixed-credibility model. By contrast, an infinitesimal increase in S's costs never leads to a sizable decrease in S's value, suggesting collapses in trust are a byproduct of rigid institutional structures. Finally, we show that in the public-persuasion setting of Section 2.5, S always chooses an institution that is immune to influence in low states, and perfectly amenable otherwise.

Related Literature. This paper contributes to the literature on strategic information transmission. To place our work, consider two extreme benchmarks: full credibility and no credibility. Our full-credibility case is the model used in the Bayesian persuasion literature (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011; Kamenica, 2019),[3] which studies sender-receiver games in which a sender commits to an information-transmission strategy. By contrast, our no-credibility case reduces to cheap talk (Crawford and Sobel, 1982; Green and Stokey, 2007). In particular, we build on

[3]See also Aumann and Maschler (1966).

Lipnowski and Ravid (2019), who use the belief-based approach to study cheap talk under state-independent sender preferences.

Two recent papers (Min, 2018; Fréchette, Lizzeri, and Perego, 2019) study closely related models. Fréchette, Lizzeri, and Perego (2019) test experimentally the connection between the informativeness of the sender's communication and her credibility in the binary-state, binary-action, independent-credibility version of our model. Min (2018) looks at a generalization of the independent-credibility version of our model in which the sender's preferences can be state dependent. He shows the sender weakly benefits from a higher commitment probability. Applying Blume, Board, and Kawamura's (2007) insights, Min (2018) also shows allowing the sender to commit with positive probability strictly helps both players in Crawford and Sobel's (1982) uniform-quadratic example.

Our paper is related to the literature on cheap talk with lying costs. In Kartik (2009), each message includes a reported state, and the cost of a message is measured via the distance between the reported and true states; as the cost increases, the sender's strategy becomes (in some sense) more truthful. In Guo and Shmaya (2019a), each communicated message is a distribution of states, and the sender faces a miscalibration cost that increases in the distance between the message and its induced equilibrium posterior belief. They obtain a surprising result: When costs are sufficiently large, the sender attains her full-commitment payoff under any extensive-form rationalizable play. Therefore, like our work, Guo and Shmaya's (2019a) model bridges the cheap talk and the Bayesian persuasion models.

Another related paper is Nguyen and Tan (2019). In Nguyen and Tan (2019), a sender has the opportunity to privately change the publicly observed outcome of a previously announced experiment. Such a change comes at a cost that may depend on the outcome. They find conditions under which the sender does not alter the experiment's

outcome in the sender-optimal equilibrium, and identify examples under which the sender obtains her commitment payoff.

We also speak to the literature that studies Bayesian persuasion under additional sender incentive constraints. In Salamanca (2019), a sender can use a mediator to design a communication protocol, but cannot commit to her own reporting strategy, and therefore must satisfy truth-telling constraints. Best and Quigley (2017) and Mathevet, Pearce, and Stacchetti (2019) both study a long-lived sender who interacts with a sequence of short-lived receivers via cheap talk. Each shows how enriching the environment can restore the sender's commitment value: in Best and Quigley (2017), by coarsening receivers' information via a review aggregator, and in Mathevet, Pearce, and Stacchetti (2019), via a reputational concern for the sender. A number of papers (Perez-Richet, 2014; Hedlund, 2017; Alonso and Câmara, 2018) study persuasion by a privately informed sender who might face exogenous constraints in her choice of signal. Perez-Richet (2014) studies the information-design analogue of an informed-principal (Myerson, 1983) problem. In Alonso and Câmara (2018) and Hedlund (2017), the sender is imperfectly informed. The former compares the value of expertise with the uninformed case and shows that private information cannot be beneficial if the sender's private information is (sequentially) redundant relative to the set of available signals. The latter shows that in a two-state model with state-independent preferences, the sender's behavior in any D1 equilibrium reveals either the sender's private information or the state. Perez-Richet and Skreta (2018) introduce the possibility of falsification in the context of test design, where a sender can make each state produce the conditional signal distribution associated with the other. Thus, their sender can manipulate a Blackwell experiment's input, whereas our sender manipulates the experiment's output.

Our productive-mistrust result relates to Ichihashi (2019), who analyzes the effect of bounding the informativeness of the sender's experiment in the binary-action

specialization of Kamenica and Gentzkow (2011). Ichihashi's (2019) main result characterizes the equilibrium outcome set as a function of said upper bound. He also shows that, whereas such a bound often helps the receiver, the receiver is always harmed from such a bound when the state is binary. By contrast, productive mistrust can occur with any number of states.

The model we analyze in Section 2.5 concerns persuasion of a population, and so relates thematically to the literature on persuasion with multiple receivers (e.g., Alonso and Câmara, 2016a; Bardhi and Guo, 2018; Chan, Gupta, Li, and Wang, 2019). Because our sender's motive is separable across audience members, the model in that section can be reinterpreted as communication to a single receiver who holds private information. Consequently, it relates to work by Kolotilin (2018), Guo and Shmaya (2019b), and Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), all of whom study information design under full commitment. We contribute to this literature by studying the effects of limited credibility.

Whereas our sender derives credibility through an institution, credibility can also arise via hard evidence. The effect of evidence on communication has been the subject of many studies (Glazer and Rubinstein, 2006; Sher, 2011; Hart, Kremer, and Perry, 2017; Ben-Porath, Dekel, and Lipman, 2019; Rappoport, 2017). Many such studies share our assumption of sender state-independent preferences but focus on receiver- (rather than sender-) optimal equilibria. The equivalence between such equilibria and the receiver's commitment outcome is a common point of inquiry.

Weak institutions often serve as a justification for examining mechanism design under limited commitment (Bester and Strausz, 2001; Skreta, 2006; Deb and Said, 2015; Liu, Mierendorff, Shi, and Zhong, 2019). We complement this literature by relaxing a principal's commitment power in the control of information rather than of mechanisms.

2.2 A Weak Institution

There are two players: a sender (S, she) and a receiver (R, he). Whereas both players' payoffs depend on R's action, $a \in A$, R's payoff also depends on an unknown state, $\theta \in \Theta$. Thus, S and R have objectives $u_S : A \to \mathbb{R}$ and $u_R : A \times \Theta \to \mathbb{R}$, respectively, and each aims to maximize expected payoffs.

The game begins with S commissioning a report, $\xi : \Theta \to \Delta M$, to be delivered by a research institution. The state then realizes, and R receives a message $m \in M$ (without observing θ). Given θ, S is credible with probability $\chi(\theta)$, meaning m is drawn according to the official reporting protocol, $\xi(\cdot|\theta)$. With probability $1 - \chi(\theta)$, S is not credible, in which case S decides which message to send after privately observing θ. Only S learns her credibility type, and she learns it only after announcing the official reporting protocol.

We now introduce some notation, which we use throughout. For a compact metrizable space, Y, we denote by ΔY the set of all Borel probability measures over Y, endowed with the weak* topology. If $f : Y \to \mathbb{R}$ is bounded and measurable and $\zeta \in \Delta Y$, define the measure $f\zeta$ on Y via $f\zeta(\hat{Y}) := \int_{\hat{Y}} f \, \mathrm{d}\zeta$ for each Borel $\hat{Y} \subseteq Y$. When the domain is not ambiguous, we use $\mathbf{1}$ and $\mathbf{0}$ to denote constant functions taking value 1 and 0, respectively.

We impose some technical restrictions on our model. Both A and Θ are compact metrizable spaces with at least two elements, the objectives u_R and u_S are continuous, and $\chi : \Theta \to [0,1]$ is measurable. We say the model is **finite** when referring to the special case in which both A and Θ are finite. The state, θ, is assumed to follow some full-support prior distribution $\mu_0 \in \Delta\Theta$, which is known to both players. Finally, we assume the message space M is an uncountable compact metrizable space.[4]

[4]This richness condition enables our complete characterization of equilibrium outcomes (Lemma A.3.1). If Θ is finite, our characterization of sender-optimal equilibrium values (Theorem 2.1) and our applied propositions hold without change for all M such that $|M| \geqslant 2|\Theta|$.

We now define an equilibrium, which consists of four objects: S's official reporting protocol, $\xi : \Theta \to \Delta M$, executed whenever S cannot influence reporting; the strategy that S employs when not committed, $\sigma : \Theta \to \Delta M$; R's strategy, $\alpha : M \to \Delta A$; and R's belief map, $\pi : M \to \Delta\Theta$, assigning a posterior to each message. A χ-equilibrium is an official reporting policy announced by S, ξ, together with a perfect Bayesian equilibrium of the subgame following S's announcement. Formally, a **χ-equilibrium** is a tuple $(\xi, \sigma, \alpha, \pi)$ of measurable maps such that

1. $\pi : M \to \Delta\Theta$ is derived from μ_0 via Bayes' rule, given message policy

$$\chi\xi + (1-\chi)\sigma : \Theta \to \Delta M,$$

   whenever possible;

2. $\alpha(m)$ is supported on $\operatorname{argmax}_{a\in A} \int_\Theta u_R(a,\cdot)\, \mathrm{d}\pi(\cdot|m)$ for all $m \in M$;

3. $\sigma(\theta)$ is supported on $\operatorname{argmax}_{m\in M} \int_A u_S \,\mathrm{d}\alpha(\cdot|m)$ for all $\theta \in \Theta$.

We view S as a principal capable of steering R toward her favorite χ-equilibria. Because such equilibria automatically satisfy S's incentive constraints on choice of ξ, we omit said constraints for the sake of brevity.

2.3 Persuasion with Partial Credibility

This section presents Theorem 2.1, which geometrically characterizes S's optimal χ-equilibrium value. To prove the theorem, we adopt a belief-based approach by using R's ex-ante belief distribution, $p \in \Delta\Delta\Theta$, to summarize equilibrium communication. When communication is sufficiently flexible, the sole restriction imposed on an induced belief distribution is Bayes plausibility: R's average posterior belief equals his prior belief; that is, $\int_{\Delta\Theta} \mu \,\mathrm{d}p(\mu) = \mu_0$. We refer to any such p as an **information policy** and denote the set of all information policies by $\mathcal{R}(\mu_0)$.

We represent each of S's messages with the posterior belief it induces in equilibrium and use S's **value correspondence**,

$$V : \Delta\Theta \rightrightarrows \mathbb{R}$$
$$\mu \mapsto \text{co } u_S\left(\operatorname*{argmax}_{a\in A} \int u_R(a,\cdot)\,\mathrm{d}\mu\right),$$

to account for R's incentive constraints. In words, $V(\mu)$ is the set of payoffs that S can attain when R behaves optimally given posterior belief μ. Note that (appealing to Berge's theorem) V is a Kakutani correspondence, that is, a nonempty-compact-convex-valued, upper hemicontinuous correspondence. As such, S's **value function**, $v(\mu) := \max V(\mu)$, which identifies S's highest continuation payoff from inducing posterior μ, is a well-defined, upper semicontinuous function.

When S is fully credible ($\chi(\cdot) = \mathbf{1}$), only S's official reporting protocol matters. Because S publicly commits to this rule at the beginning of the game, Bayes plausibility is the only constraint imposed on equilibrium communication. Hence, R may as well break ties in S's favor, reducing the maximization of S's equilibrium value to the maximization of v's expected value across all information policies. Aumann and Maschler (1995) and Kamenica and Gentzkow (2011) show the highest such value is given by the pointwise lowest concave upper semicontinuous function that majorizes v.[5] This function, which we denote by $\hat{v}$, is known as v's **concave envelope**.

Under no credibility ($\chi(\cdot) = \mathbf{0}$), the official reporting protocol plays no role, because S always influences the report. Therefore, S's messages must satisfy her incentive constraints, which take a very simple form due to S's state-independent payoffs: All on-path messages must give S the same continuation payoff. Lipnowski and Ravid (2019) show the maximal value that S can attain subject to this constraint is given by v's **quasicon-**

[5] In the case in which Θ is finite, the qualifier "upper semicontinuous" may be ommited in the definition of the (quasi)concave envelope. For instance, see Lipnowski and Ravid (2019), Corollary 4.

cave envelope, which is the lowest quasiconcave upper semicontinuous function that majorizes v. We denote this function by $\bar{v}$.

Theorem 2.1 shows that for intermediate $\chi(\cdot)$, S's highest χ-equilibrium value is characterized by an object that combines the concave and quasiconcave envelopes. For $\gamma \in \Delta\Theta$, define

$$\begin{aligned} v_{\wedge\gamma} : \Delta\Theta &\to \mathbb{R} \\ \mu &\mapsto \min\{\bar{v}(\gamma), v(\mu)\}. \end{aligned}$$

Theorem 2.1's characterization is based on the concave envelope of $v_{\wedge\gamma}$, which we denote by $\hat{v}_{\wedge\gamma}$. Figure 2.1 below visualizes the construction of $\hat{v}_{\wedge\gamma}$ in the binary-state case.

(a) $\bar{v}$ (b) $\hat{v}$ (c) $\hat{v}_{\wedge\gamma}$

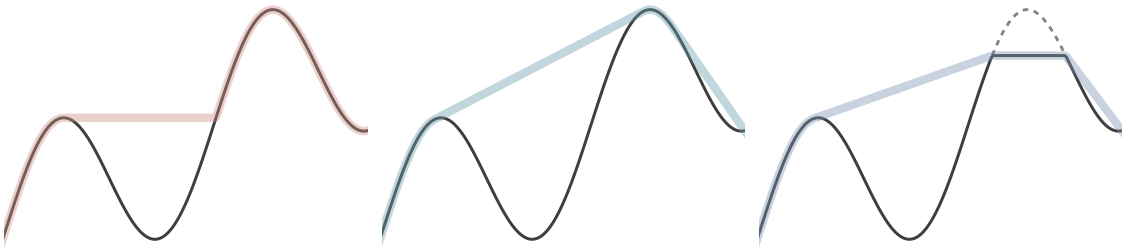


FIGURE 2.1: Quasiconcave envelope, concave envelope, and concave envelope with a cap.

With the relevant building blocks in hand, we now state our main result.

Theorem 2.1. *A sender-optimal χ-equilibrium exists and yields ex-ante sender payoff*

$$\begin{aligned} v_\chi^*(\mu_0) = \max_{\beta,\gamma\in\Delta\Theta,\ k\in[0,1]} \quad & k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) \\ \text{s.t.} \quad & k\beta + (1-k)\gamma = \mu_0, && \text{(R-BP)} \\ & (1-k)\gamma \geqslant (\mathbf{1}-\chi)\mu_0. && (\chi\text{-BP}) \end{aligned}$$

To understand Theorem 2.1, note that every χ-equilibrium partitions the messages R sees into two sets: the messages that are sometimes sent under influenced reporting, M_γ (messages that are "good" for S), and the messages that are not, M_β (those that are "bad" for S). Official reporting can send messages from either set. The theorem follows from maximizing S's expected payoffs from M_γ and M_β, holding R's expected posterior conditional on M_γ and M_β fixed at γ and β, respectively. As we explain below, this maximization yields a value of $k\hat{v}_{\wedge\gamma}(\beta)+(1-k)\bar{v}(\gamma)$, where k is the probability that the realized message is in M_β. All that remains is to maximize this value over the set of feasible triplets, (β,γ,k), which are constrained by Bayes plausibility in two ways, corresponding to (R-BP) and (χ-BP), respectively. First, the average posteriors must be equal to the prior, yielding (R-BP). Second, the ex-ante probability that R sees a message from M_γ and an event $\hat{\Theta}$ occurs is at least the ex-ante probability that $\hat{\Theta}$ occurs and reporting is influenced.

We now explain the characterization of S's optimal values from M_γ and M_β, which is based on the no-credibility and full-credibility cases, respectively. Because all messages in M_γ are sent under influenced reporting, they must satisfy the same constraints as in the no-credibility case. By Lipnowski and Ravid's (2019) arguments, $\bar{v}(\gamma)$ is the highest payoff that S can obtain from sending a message under these constraints. For S to send such messages, though, S's payoff from M_γ must be above her continuation payoff from any message in M_β. This requirement restricts M_β in two ways: (1) It caps S's continuation payoff from any feasible posterior, and (2) it restricts the set of feasible posteriors in M_β, precluding posteriors from which S must obtain too high a continuation payoff. In the proof, we argue the second constraint is automatically satisfied at the optimum. As such, one can apply the same arguments as in the full-credibility case, but with v replaced by $v_{\wedge\gamma}$. That S's highest payoff from M_β is given by $\hat{v}_{\wedge\gamma}(\beta)$ follows.

2.4 Varying Credibility

This section uses Theorem 2.1 to conduct general comparative statics in the model's finite version. First, we study how a decrease in S's credibility affects R's value. In particular we provide sufficient conditions for R to benefit from a less credible S. Second, we show that small reductions in S's credibility often lead to a large drop in S's payoffs. Finally, we note that these drops rarely occur at full credibility. In other words, the full credibility value is robust to small decreases in S's commitment power.

Productive Mistrust We now study how a decrease in S's credibility impacts R's value and the informativeness of S's equilibrium communication. In general, the less credible the sender, the smaller the set of equilibrium information policies.[6] However, that the set of equilibrium policies shrinks does not mean less information is transmitted in S's preferred equilibrium. Our introductory example is a case in point, showing that lowering S's credibility can result in a more informative equilibrium (à la Blackwell, 1953). Moreover, this additional information is used by R, who obtains a strictly higher value when S's credibility is lower. In what follows, we refer to this phenomenon as productive mistrust, and provide sufficient conditions for it to occur.

Our key sufficient condition involves S's optimal information policy under full credibility. Given prior μ, an information policy $p \in \mathcal{R}(\mu)$ is a **show-or-best** (SOB) policy if it is supported on $\{\delta_\theta\}_{\theta\in\Theta} \cup \operatorname{argmax}_{\mu'\in\Delta[\operatorname{supp}(\mu)]} v(\mu')$. In words, p is an SOB policy if it either shows the state to R, or brings R to a posterior that attains S's best feasible value. Say S is a **two-faced SOB** if, for every binary-support prior $\mu \in \Delta\Theta$, every $p \in \mathcal{R}(\mu)$ is outperformed by an SOB policy $p' \in \mathcal{R}(\mu)$; that is, $\int_{\Delta\Theta} v \,\mathrm{d}p \leqslant \int_{\Delta\Theta} v \,\mathrm{d}p'$. Figure 2.2 depicts an example in which S is a two-faced SOB. Note that productive mistrust cannot occur in this example. Indeed, one can show that, if S's favorite equilibrium policy

[6] See Lemma A.3.1 in the appendix.

changes as credibility declines, it must switch to no information. As such, R prefers a more credible S.

Finally, say a model is **generic** if R is (i) not indifferent between any two actions at any degenerate belief, and (ii) not indifferent between any three actions at any binary-support belief.[7]

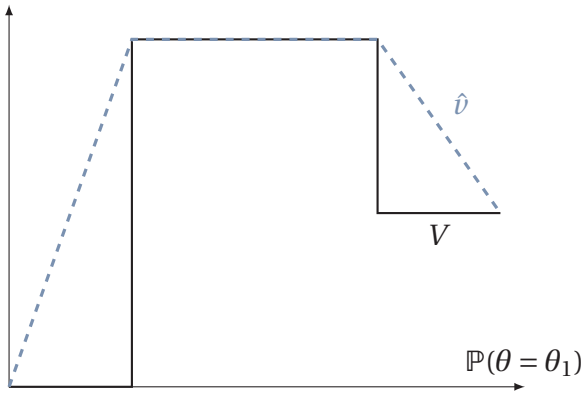


FIGURE 2.2: Sender is a two-faced SOB

Proposition 2.1 below shows that, in generic finite settings, S not being a two-faced SOB is sufficient for productive mistrust to occur for some full-support prior. Intuitively, S being an SOB means that a highly credible S has no bad information to hide: under full credibility, S's bad messages are maximally informative, subject to keeping R's posterior fixed following S's good messages. S not being an SOB at some prior means that S's bad messages optimally hide some instrumental information. By reducing S's credibility just enough to make the full-credibility solution infeasible, one can push S to reveal some of that information to R. In other words, S commits to potentially revealing more-extreme bad information in order to preserve the credibility of her good messages. Proposition 2.1 below formalizes this intuition.

[7]Given a fixed finite A and Θ, genericity holds for (Lebesgue) almost every $u_R \in \mathbb{R}^{A\times\Theta}$. In particular, it holds if $u_R(a,\theta) \neq u_R(a',\theta)$ for all distinct $a, a' \in A$ and all $\theta \in \Theta$, and $\frac{u_R(a_1,\theta_1)-u_R(a_2,\theta_1)}{u_R(a_1,\theta_2)-u_R(a_2,\theta_2)} \neq \frac{u_R(a_2,\theta_1)-u_R(a_3,\theta_1)}{u_R(a_2,\theta_2)-u_R(a_3,\theta_2)}$ for all distinct $a_1, a_2, a_3 \in A$ and all distinct $\theta_1, \theta_2 \in \Theta$.

Proposition 2.1. *Consider a finite and generic model in which S is not a two-faced SOB. Then, a full-support prior and credibility functions $\chi'(\cdot) < \chi(\cdot)$ exist such that every sender-optimal χ'-equilibrium is strictly better for R than every sender-optimal χ-equilibrium.*[8]

We should emphasize that Proposition 2.1's conditions are not necessary. We provide a necessary and sufficient condition for productive mistrust to occur at a given prior for the binary-state, finite-action case in the appendix. In particular, we weaken the SOB condition by requiring only that S wants to withhold information at the lowest credibility level at which she can beat her no-credibility payoff. We refer the reader to Lemma A.3.2 in the appendix for precise details. We do not know an analogous tight characterization of when productive mistrust occurs in the many-state model.

Collapse of Trust Theorem 2.1 immediately implies lowering S's credibility can only decrease her value.[9] Below we show this decrease is often discontinuous. In other words, small decreases in S's credibility often result in a large drop in S's benefits from communication.

Proposition 2.2. *In a finite model, the following are equivalent:*

(i) A collapse of trust never occurs:[10]

$$\lim_{\chi'(\cdot) \nearrow \chi(\cdot)} v^*_{\chi'}(\mu_0) = v^*_{\chi}(\mu_0)$$

for every $\chi(\cdot) \in [0,1]^{\Theta}$ and every full-support prior μ_0.

(ii) Commitment is of no value: $v^*_{\mathbf{1}} = v^*_{\mathbf{0}}$.

(iii) No conflict occurs: $v(\delta_\theta) = \max v(\Delta\Theta)$ for every $\theta \in \Theta$.

[8]Moreover, when $|\Theta| = 2$, every sender-optimal χ'-equilibrium is more Blackwell-informative than every sender-optimal χ-equilibrium.

[9]It also implies value *increases* have a continuous payoff effect: A sufficiently small increase in S's credibility never results in a large gain in S's benefits from communication.

[10]Convergence of $\chi'(\cdot) \to \chi(\cdot)$ is in the Euclidean topology on $\mathbb{R}^{\Theta}$.

Proposition 2.2 establishes that, in most finite examples, S's value collapses discontinuously when credibility decreases. In particular, such collapses are absent for all priors if and only if S wants to tell R all that she knows, or if, equivalently, commitment is immaterial to S.

Robustness of the Commitment Case Given the large and growing literature on optimal persuasion with commitment, wondering whether the commitment solution is robust to small decreases in S's credibility is natural. The answer turns out to be almost never. Thus, although small decreases in credibility often lead to a collapse in S's value, these collapses rarely occur at $\chi(\cdot) = \mathbf{1}$.

Proposition 2.3. *In a finite model, the following are equivalent:*

(i) *The full commitment value is robust:* $\lim_{\chi(\cdot) \nearrow \mathbf{1}} v^*_\chi(\mu_0) = v^*_{\mathbf{1}}(\mu_0)$ *for every full-support* μ_0.

(ii) *S gets the benefit of the doubt: Every* $\theta \in \Theta$ *is in the support of some member of* $\operatorname{argmax}_{\mu \in \Delta\Theta} v(\mu)$.

Proposition 2.3 shows that the full-credibility value is robust if and only if S can persuade R to take her favorite action without ruling out any states. In other words, robustness of the commitment solution is equivalent to S getting the benefit of the doubt.

2.5 Persuading the Public

This section considers a single sender interested in persuading a population of receivers to take a favorable action. For example, S could be a government of a small open economy trying to encourage foreigners to invest in the local market, a seller advertising to entice consumers to buy her product, or a leader vying for the support of her populace. To persuade the receivers, S commissions a weak institution (e.g., a central bank,

product reviewer company, or state-owned media outlet) to issue a public report. In this section, we analyze the S-optimal report under partial credibility, and identify the states at which credibility is most valuable for S.

We modify our model as follows. The report of S's institution is now publicly revealed to a unit mass of receivers. After observing the institution's report, receivers simultaneously take a binary action. Each receiver i cares only about his own action, $a_i \in A = \{0,1\}$. Receiver i's payoff from a_i is given by $a_i(\theta - \omega_i)$, where $\theta \in \Theta = [0,1]$ is the unknown state, distributed according to an atomless, full-support prior μ_0, and $\omega_i \in \mathbb{R}$ is receiver i's type. The mass of receivers whose type is below ω is given by $H(\omega)$, an absolutely continuous cumulative distribution function whose density h is continuous, strictly quasiconcave, and strictly positive on $(0,1)$. S's objective is to maximize the proportion of receivers taking action 1.

An equilibrium of the modified game is tuple, $(\xi, \sigma, \alpha, \pi)$, where $\xi : \Theta \to \Delta M$, $\sigma : \Theta \to \Delta M$, and $\pi : M \to \Delta\Theta$ respectively represent S's official report, S's strategy when not committed, and the public's belief mapping, as in the original game. We let $\alpha : M \to [0,1]$ represent the proportion of receivers taking action 1 conditional on the realized message. Observe action 1 is optimal for receiver i if and only if $\omega_i \leqslant E\mu$, where $\mu \in \Delta\Theta$ is the publicly held posterior about θ, and E maps beliefs to their associated expectations.[11] As such, given a posterior μ, the proportion of receivers taking action 1 is given by $H(E\mu)$. Thus, a **χ-equilibrium** is a tuple $(\xi, \sigma, \alpha, \pi)$ where π is derived from μ_0 via Bayes' rule, $\alpha(\cdot) = H(E\pi(\cdot))$, and $\sigma(\theta)$ is supported on $\arg\max_{m \in M} \alpha(m)$ for all θ.

Theorem 2.1 applies readily to the current setting. Because $H(E\mu)$ is the proportion of the population taking action 1 given posterior $\mu \in \Delta\Theta$, S's continuation payoff from a public message inducing μ is $v(\mu) := H(E\mu)$. Taking v to be S's value function, we can directly apply Theorem 2.1 to the current game.

[11] That is, $E\mu := \int \theta \, \mathrm{d}\mu(\theta)$ for all $\mu \in \Delta\Theta$.

Next, we use Theorem 2.1 to find S's optimal χ-equilibrium. We begin with the extreme credibility levels. Suppose first S has no credibility; that is, $\chi = \mathbf{0}$. In this case, S's optimal value is given by the quasiconcave envelope of S's value function evaluated at the prior, $\bar{v}(\mu_0)$. Because an increasing transformation of an affine function is quasiconcave, $v = H \circ E = \bar{v}$. Hence, with no credibility, S cannot benefit from communication.

Suppose now that S has full credibility; that is, $\chi = \mathbf{1}$. In this case, S's maximal χ-equilibrium value equals v's maximal expected value across all information policies, $p \in \mathcal{R}(\mu_0)$. Notice that a given information policy p yields an expected value of $\int H(\cdot)\,\mathrm{d}\mu$, where $\mu = p \circ E^{-1} \in \Delta\Theta$ is the distribution of the population's posterior mean. As such, maximizing S's value across all information policies is the same as maximizing the expectations of $H(\cdot)$ across all posterior mean distributions produced by some information policy. Such posterior mean distributions are characterized via the notion of mean-preserving spreads.[12] Formally, we say $\mu \in \Delta\Theta$ is a **mean-preserving spread** of $\tilde{\mu} \in \Delta\Theta$, denoted by $\mu \succcurlyeq \tilde{\mu}$, if

$$\int_0^{\hat{\theta}} \mu[0,\theta]\,\mathrm{d}\theta \geqslant \int_0^{\hat{\theta}} \tilde{\mu}[0,\theta]\,\mathrm{d}\theta,\ \forall \hat{\theta} \in [0,1],\ \text{with equality at } \hat{\theta} = 1. \tag{MPS}$$

As is well known,[13] μ_0 being a mean-preserving spread of μ is both necessary and sufficient for μ to arise as the posterior mean distribution of some information policy. Thus, S's value under full credibility is given by

$$\hat{v}(\mu_0) = \max_{\mu \in \Delta\Theta:\ \mu \preccurlyeq \mu_0} \int H(\cdot)\,\mathrm{d}\mu.$$

The solution to the above program is dictated by the shape of the CDF H. Because the CDF's density, h, is strictly quasiconcave, H is a convex-concave function over $[0,1]$. Said differently, an $\omega^* \in [0,1]$ exists such that H is strictly convex on $[0,\omega^*]$, and strictly

[12] See Blackwell and Girshick (1954) and Rothschild and Stiglitz (1970).
[13] See Gentzkow and Kamenica (2016) and references therein.

concave on $[\omega^*, 1]$. As noted by Kolotilin (2018) and Dworczak and Martini (2019), when H is convex-concave, the above program can be solved via θ^* upper censorship, which we now formally define. Under full credibility, θ^* upper censorship arises whenever S's official report reveals (pools) all states below (above) θ^*. Given such an official reporting protocol, it is optimal for S to say the state is above θ^* whenever she influences the report. Thus, we say (ξ, σ) is a θ^***-upper-censorship pair** if every $\theta \in \Theta$ has $\sigma(\cdot|\theta) = \delta_1$ and

$$\xi(\cdot|\theta) = \begin{cases} \delta_\theta & \text{if } \theta \in [0, \theta^*), \\ \delta_1 & \text{if } \theta \in [\theta^*, 1]. \end{cases}$$

Given a θ^*-upper-censorship pair, we refer to the resulting posterior mean distribution,[14]

$$\mathbf{1}_{[0,\theta^*)}\mu_0 + \mu_0[\theta^*, 1]\delta_{\mathbb{E}_{\mu_0}[\theta|\theta \geqslant \theta^*]},$$

as a θ^* **upper censorship** of μ_0. That upper censorship solves the full-credibility problem has been discussed by the aforementioned papers under slightly different assumptions. Still, we provide an elementary proof in the appendix for completeness.

We find upper-censorship pairs are also optimal when credibility is partial, although the reasoning is more delicate. One complication is that not every upper-censorship pair induces a χ-equilibrium. The reason is that under partial credibility, the posterior mean following message 1 can be strictly below the posterior mean induced by other messages, thereby violating S's incentive constraints. To avoid such a violation, the mean induced by message 1 must be above the upper-censorship cutoff, θ^*, which is

[14]Recall our notational convention: For bounded measurable $f : \Theta \to \mathbb{R}_+$ and $\mu \in \Delta\Theta$, we let $f\mu$ represent the measure defined via $f\mu(\hat{\Theta}) = \int_{\hat{\Theta}} f \, \mathrm{d}\mu$.

equivalent to[15]

$$\int (\theta - \theta^*)(1 - \mathbf{1}_{[0,\theta^*)}\chi(\theta))\mathrm{d}\mu_0(\theta) \geqslant 0. \qquad (\theta^*\text{-IC})$$

Observe that with intermediate credibility,[16] the left-hand side of (θ^*-IC) is continuous and strictly decreasing in θ^*, strictly positive for $\theta^* = 0$, and strictly negative for $\theta^* = 1$.[17] As such, (θ^*-IC) holds whenever θ^* is below the unique upper-censorship cutoff at which it holds with equality, a cutoff that we denote by $\bar{\theta}_\chi$.

Another complication arising from partial credibility is that a θ^*-upper-censorship pair does not typically yield an upper censorship of μ_0 as its posterior mean distribution. Instead, every θ^*-upper-censorship pair with $\theta^* \leqslant \bar{\theta}_\chi$ turns out to yield a θ^* upper censorship of

$$\bar{\mu}_\chi = \mathbf{1}_{[0,\bar{\theta}_\chi)}\chi\mu_0 + \left(1 - \chi\mu_0[0,\bar{\theta}_\chi)\right)\delta_{\bar{\theta}_\chi},$$

which is the posterior mean distribution induced by the $\bar{\theta}_\chi$-upper-censorship pair.

Claim 2.1 below shows that upper censorship always yields an S optimal χ-equilibrium. Moreover, to find the optimal censorship cutoff, one can solve the full-credibility problem with the modified prior $\bar{\mu}_\chi$.

Claim 2.1. A $\theta^* \in [0,\bar{\theta}_\chi]$ exists such that the θ^* upper censorship of $\bar{\mu}_\chi$, denoted by μ_{χ,θ^*}, satisfies

$$v^*_\chi(\mu_0) = \hat{v}(\bar{\mu}_\chi) = \int H(\cdot)\,\mathrm{d}\mu_{\chi,\theta^*}.$$

Moreover, the corresponding θ^*-upper-censorship pair is an S-optimal χ-equilibrium that induces μ_{χ,θ^*} as its posterior mean distribution.

Using Claim 2.1, we can compare the value of credibility in different states. Indeed, the claim makes it obvious that, regardless of the population's type distribution,

[15]To see this equivalence, note that R's posterior mean conditional on seeing message 1 from a θ^*-upper-censorship pair equals $\frac{\int\theta[\mathbf{1}_{[\theta^*,1]}\chi(\theta)+1-\chi(\theta)]\mathrm{d}\mu_0}{\int[\mathbf{1}_{[\theta^*,1]}\chi(\theta)+1-\chi(\theta)]\mathrm{d}\mu_0} = \frac{\int\theta[1-\mathbf{1}_{[0,\theta^*)}\chi(\theta)]\mathrm{d}\mu_0}{\int[1-\mathbf{1}_{[0,\theta^*)}\chi(\theta)]\mathrm{d}\mu_0}$, which is larger than θ^* only if (θ^*-IC) holds.

[16]That is, if $\mu_0\{\chi = 0\}, \mu_0\{\chi = 1\} < 1$.

[17]Recall μ_0 is assumed to be an atomless, full-support distribution over $[0,1]$.

S prefers the credibility distribution χ over $\tilde{\chi}$ whenever $\bar{\mu}_\chi$ is a mean-preserving spread of $\bar{\mu}_{\tilde{\chi}}$. One can then show by construction that the converse is also true; that is, S prefers χ to $\tilde{\chi}$ for all population type distributions *only if* $\bar{\mu}_{\tilde{\chi}}$ is a mean-preserving spread of $\bar{\mu}_\chi$. We present this result in Claim 2.2 below.

Claim 2.2. $v^*_\chi(\mu_0) \geqslant v^*_{\tilde{\chi}}(\mu_0)$ for all type distributions[18] if and only if $\bar{\mu}_\chi \succcurlyeq \bar{\mu}_{\tilde{\chi}}$.

The economic intuition behind the claim is that credibility is most valuable when the conflict between S's ex-ante and ex-post incentives is large. Indeed, it is useful to notice that $\bar{\mu}_\chi \succcurlyeq \bar{\mu}_{\tilde{\chi}}$ holds if and only if[19]

$$\int_0^{\hat{\theta}} \int_0^{\theta} (\chi - \tilde{\chi}) \, \mathrm{d}\mu_0 \, \mathrm{d}\theta \geqslant 0 \text{ for all } \theta \in [0, \bar{\theta}_{\tilde{\chi}}].$$

Thus, the claim shows a sense in which S prefers to have more credibility in low states. Intuitively, low states are those which S benefits from revealing ex ante but would like to hide ex post. The more credibility S has in those states, the less S's ex-post incentives interfere with his ex-ante payoffs, and so the higher is S's value.

2.6 Investing in Credibility

In this section, we extend our model to endogenize S's credibility χ. Specifically, suppose S can choose any measurable $\chi : \Theta \to [0,1]$ at a cost of $c\left(\int \chi \, \mathrm{d}\mu_0\right)$ prior to the persuasion game, where $c : [0,1] \to \mathbb{R}_+$ is continuous and strictly increasing. Then, S chooses χ to solve

$$v^{**}_c(\mu_0) = \max_\chi \left[v^*_\chi(\mu_0) - c\left(\int \chi \, \mathrm{d}\mu_0 \right) \right].$$

[18] That is, for all H admitting a continuous, quasiconcave density.

[19] To see the equivalence, one can verify that $\hat{\theta} \in [0, \bar{\theta}_\chi]$ has $\int_0^{\hat{\theta}} \bar{\mu}_\chi[0,\theta] \, \mathrm{d}\theta = \int_0^{\hat{\theta}} \int_0^{\theta} \chi \, \mathrm{d}\mu_0 \geqslant \theta - E\mu_0$, and each $\hat{\theta} \in [\bar{\theta}_\chi, 1]$ has $\int_0^{\hat{\theta}} \bar{\mu}_\chi[0,\theta] \, \mathrm{d}\theta = \theta - E\mu_0$ — and similarly for $\tilde{\chi}$. Therefore, the ranking (MPS) holds vacuously above $\bar{\theta}_{\tilde{\chi}}$ and reduces to the given equation below $\bar{\theta}_{\tilde{\chi}}$.

Clearly, S never invests in greater credibility than is necessary to induce her equilibrium information. As such, S always chooses (χ, k, β, γ) so that (χ-BP) holds with equality. Combining this observation with (R-BP) yields

$$\int \chi \, \mathrm{d}\mu_0 = k\beta(\Theta) = k.$$

S's problem therefore reduces to

$$\begin{aligned} v_c^{**}(\mu_0) = \max_{\beta,\gamma\in\Delta\Theta,\ k\in[0,1]} \quad & k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) - c(k) \\ \text{s.t.} \quad & k\beta + (1-k)\gamma = \mu_0. \end{aligned}$$

We now discuss how our results change when credibility is endogenized as above. We begin by revisiting productive mistrust. Similar to R's ability to benefit from a decrease in exogenous credibility, R can also benefit from an increase in S's credibility costs. Recall our introductory example, and suppose the cost function is given by $c(k) = \frac{\lambda}{2}k^2$ for some $\lambda > 0$. For any $\lambda \in [2,3)$, one can verify S has a unique optimal investment choice, leading to equilibrium distribution of posteriors

$$\left[1 - \left(\frac{6}{\lambda} - 2\right)\right]\left(\frac{1}{3}\delta_0 + \frac{2}{3}\delta_{\frac{3}{4}}\right) + \left[\frac{6}{\lambda} - 2\right]\left(\frac{1}{2}\delta_{\frac{1}{4}} + \frac{1}{2}\delta_{\frac{3}{4}}\right).$$

It is straightforward that this equilibrium information structure is Blackwell-monotone in λ — higher λ leads to a mean-preserving spread in posterior beliefs. Consequently, R's equilibrium payoff $(\frac{1}{4} - \frac{1}{4\lambda})$ is increasing in λ.

Whereas reducing χ in our main model often leads to a discontinuous drop in S's payoff (Proposition 2.2), a uniformly small increase in c cannot. The reason is that the set of feasible (β, γ, k) in Theorem 2.1's program is independent of the cost, and the

cost enters S's objective separably. Therefore,

$$|v_c^{**}(\mu_0) - v_{\tilde{c}}^{**}(\mu_0)| \leqslant ||c - \tilde{c}||_\infty.$$

Thus, in the endogenous-credibility model, small cost changes have small effects on S's value.

In our public-persuasion application (Section 2.5), we saw that optimal communication takes an upper-censorship form and S especially benefits from credibility in low states. These observations, together with the observation that S never invests in extraneous credibility, lead us to simple institutions when credibility is endogenous. In particular, S's optimal institution is fully immune to influence below a cutoff state, fully susceptible above, and fully informative in its official report. See Appendix A.3.6 for the formal result.

2.7 Conclusion

This paper studies a sender who uses a weak institution to disseminate information with the aim of persuading a receiver. An institution is weaker if it succumbs to external pressures with higher probability. Specifically, the weaker the institution is, the higher is the probability that its report reflects the sender's agenda rather than the truth. We analyze the value that the sender derives from communication through such an institution, as well as the information that it provides to the receiver.

Our analysis shows an institution's weakness reduces the sender's value through two channels: Restricting the kind of information the institution can disseminate, and reducing the value that the sender can extract from said information. Together, these channels lead to collapses of trust, whereby a slight decrease in an institution's strength yields a large drop in the sender's value. Moreover, these channels often result in productive mistrust, whereby the receiver benefits from the sender employing a weaker

institution. Intuitively, to credibly communicate the information the sender wishes to convey, a weaker institution must reveal information the sender would otherwise hide. Through these effects, our model highlights the role that weak institutions play in persuasion.

Our model also allows us to analyze the value of an institution's strength in different states. As a demonstration, we study a public-persuasion setting where a single sender attempts to persuade a population of receivers to take a favorable action. In this setting, the sender commissions her institution to reveal bad states, but hides those states when influencing the report. Accordingly, the sender prefers institutions that are immune to pressure in bad states, where the conflict between her ex-post and ex-ante incentives is largest.

Chapter 3

Screening with Frames

joint with Franz Ostrizek

3.1 Introduction

Ample evidence, casual empiricism and introspection suggest that framing effects are common in choice.[1] In particular, the way a product is presented and the setting of the sales interaction can have a strong impact on consumer valuations.[2] 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.

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

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

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

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 3.1). 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 either 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 3.1. 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. 3.1a). 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.[3]

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

To see how the principal achieves this, consider Fig. 3.1b. 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

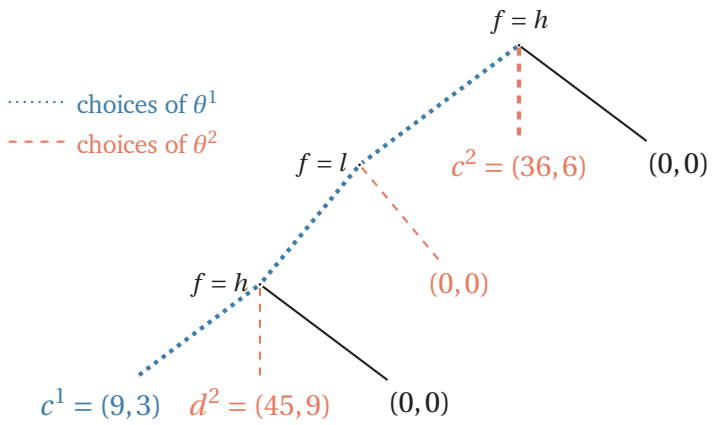
[3] In particular, the optimal contracts are $(p^1, q^1) = (8, 2)$ and $(p^2, q^2) = (32, 6)$. Note that with these functional forms, $q^i = \theta_f^i$ is efficient for frame f and the quality of type 2 is distorted downward compared to the efficient quality for *both* frames.

FIGURE 3.1: Example 3.1

(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.



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 3.1, we identify an equivalent simple optimization problem in contract space (Theorem 3.2). 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 3.3).[4]

For the main sections, we assume that consumers are sophisticated. They correctly anticipate their choices, but cannot commit to a course of action.[5] 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 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

[4]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.

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

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 3.3)[6]. 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 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 3.2. In Section 3.3, 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 3.4, 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

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

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:[7] 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 screening problems with dynamically inconsistent agents (Eliaz and Spiegler, 2006, 2008; Esteban, Miyagawa, and Shum, 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).[8] These papers consider situations when the taste changes

[7]That is, comparing their Section 3 with our Section 3.4.2, 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.

[8]Eliaz and Spiegler (2006, 2008) screen dynamically inconsistent agents by their degree of sophistication and optimistic agents by their degree of optimism, respectively. Esteban, Miyagawa, and Shum (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

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.

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, Gennaioli, and Shleifer, 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).

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.

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.

3.2 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.

3.2.1 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

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

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

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

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

$$u_{\theta_f}(c_\theta) \gtrless 0, \text{ and} \qquad (\mathrm{P}_\theta^f)$$

$$u_{\theta_f}(c_\theta) \gtrless u_{\theta_f}(c_{\theta'}) \qquad (\mathrm{IC}_{\theta\theta'}^f)$$

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 3.1 (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 3.2 (Comonotonic Environment)**.** *For any $\theta, \theta' \in \Theta$, $f, f' \in F$:*

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

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 3.2 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 3.1, it also rules out that certain frames are specific to certain types. We discuss how we can relax our assumptions in Section 3.3.5.

3.2.2 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 sit-

uation 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 $\mathbf{0} := (0,0)$. The set of 1-EDPs is $\mathcal{E}^1 := \mathcal{P}(\mathcal{C}) \times F$, 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.[9] Finally, the set $\mathcal{E}$ of all finite EDPs is given by

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

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 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.

[9] Set $\mathcal{E}^0 = \mathcal{C}$ 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 3.3.5.

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 $\forall \theta \in \Theta$

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

3.2.3 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

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

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.[10]

[10]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 $\Pi^* - \epsilon$.

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

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

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 $\widehat{c}_{\theta_f} := \left(v_{\theta_f}(\widehat{q}_{\theta_f}), \widehat{q}_{\theta_f}\right)$.

3.3 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.

In both cases, the optimal menu corresponds to the solution of the classic monopolistic screening problem with the set of types $\{\theta_h\}_{\theta\in\Theta}$.[11]

Observation 3.1. *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* $\left(\mathbf{0} \cup \{c^*_\theta\}_{\theta\in\Theta}, h\right)$ *solves* (GP) *if*

1. *the firm is constrained to* 1*-EDPs, or*

[11] 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. *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.

3.3.1 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:

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

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. 3.2):

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

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.

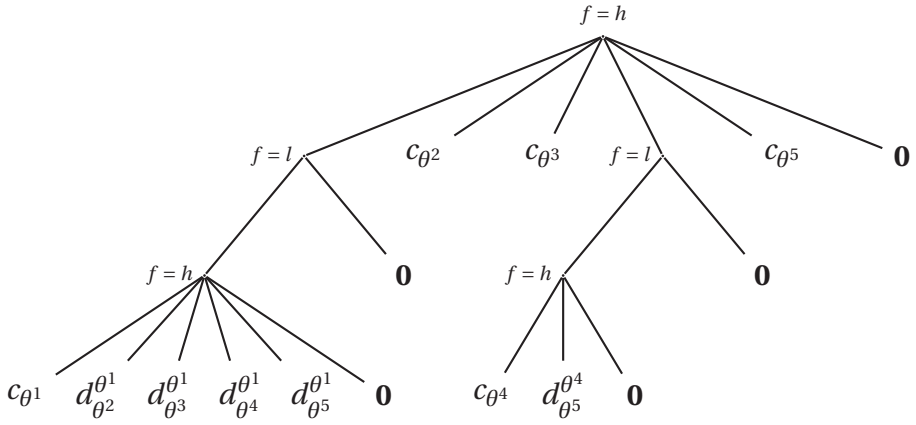
Definition 3.1. An EDP e is a *standard* EDP for a vector of contracts $\mathbf{c}$ if there exists a partition $\{\Theta_C, \Theta_R\}$ of Θ, and decoy contracts $\{d^{\theta}_{\theta'}\}_{\theta \in \Theta_C, \theta' > \theta}$, such that

$$e = \left(\{e_\theta\}_{\theta \in \Theta_C} \cup \{c_\theta\}_{\theta \in \Theta_R} \cup \{\mathbf{0}\}, h\right), \text{ where} \tag{3.1}$$

$$e_\theta = \left(\left\{\left(\{c_\theta, \mathbf{0}\} \cup \{d^{\theta}_{\theta'}\}_{\theta' > \theta}, h\right), \mathbf{0}\right\}, l\right), \forall \theta \in \Theta_C. \tag{3.2}$$

The extensive form in Example 3.1 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 3.2: 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 $\mathbf{c}$.[12]

Standard EDPs are sufficient to achieve the optimum.

Theorem 3.1. *If* $\mathbf{c}$ *is an optimal vector of contracts in* (GP)*, then it is implemented by a standard EDP.*

[12]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 $\mathbf{c}$.

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 Observation 3.1, 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.

3.3.2 Necessary and Sufficient Conditions for Implementation

In order to provide foundations for Theorem 3.1, 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 3.1. *If* **c** *is implemented by an EDP, then it satisfies constraints* $\{\mathrm{P}^h_\theta\}_{\theta\in\Theta_R}$, $\{\mathrm{P}^l_\theta\}_{\theta\in\Theta_C}$, *and* $\{\mathrm{IC}^h_{\theta\theta'}\}_{\theta\in\Theta,\theta'\in\Theta_R}$ *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 **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 3.2. *If* $\mathbf{c}$ *satisfies the constraints* $\{\mathrm{P}^h_\theta\}_{\theta\in\Theta_R}$, $\{\mathrm{P}^l_\theta\}_{\theta\in\Theta_C}$, $\{\mathrm{IC}^h_{\theta\theta'}\}_{\theta<\theta'}$, *and* $\{\mathrm{IC}^h_{\theta\theta'}\}_{\theta\in\Theta,\theta'\in\Theta_R}$ *for some partition* $\{\Theta_C,\Theta_R\}$ *of* Θ, *then* $\mathbf{c}$ *is implemented by a standard EDP.*

As in Example 3.1, 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 3.1 (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* (3.2) *has an outcome* σ *that satisfies*

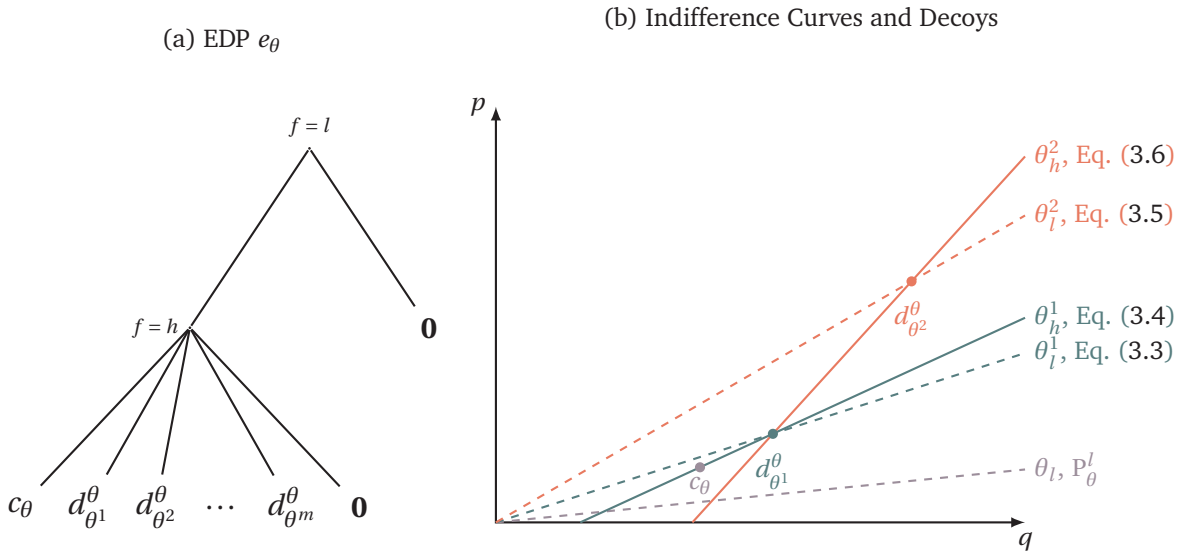
(i) $\sigma(\theta')\in\{\mathbf{0}, c_\theta\}$ *for all* $\theta'\in\Theta$,

(ii) $\sigma(\theta')=\mathbf{0}$ *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. 3.3. 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

FIGURE 3.3: The construction of e_θ.



greater 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

$$u_{\theta^1_l}(\mathbf{0}) = u_{\theta^1_l}(d^\theta_{\theta^1}). \tag{3.3}$$

$$u_{\theta^1_h}(c_\theta) = u_{\theta^1_h}(d^\theta_{\theta^1}) \tag{3.4}$$

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

$$u_{\theta^2_l}(\mathbf{0}) = u_{\theta^2_l}(d^\theta_{\theta^2}). \tag{3.5}$$

$$u_{\theta^2_h}(d^\theta_{\theta^1}) = u_{\theta^2_h}(d^\theta_{\theta^2}) \tag{3.6}$$

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 $\{c_\theta, d^\theta_{\theta_1}, \ldots, d^\theta_{\theta_m}, \mathbf{0}\}$ 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 Appendix A.4.2. □

3.3.3 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 3.1). 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.

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

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

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

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

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 3.2.

Theorem 3.2. *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)

$$\Pi^* = \Pi^R,$$

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 3.2 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.[13] 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 3.1.

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 3.3. *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:

[13] 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 Appendix A.4.2.

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 3.4. *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*

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

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 (3.4)), it is costless to conceal the type.

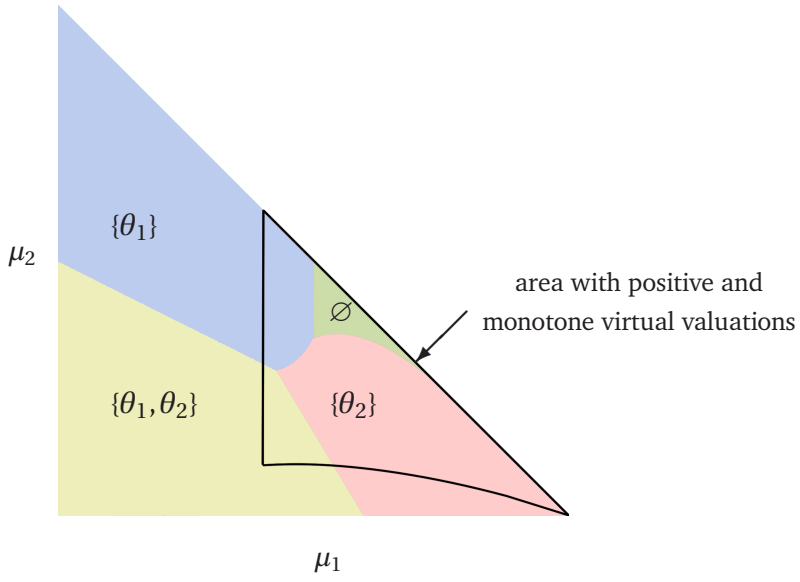
3.3.4 Optimal Concealed Types

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.

Observation 3.2. *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. 3.4 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 3.4: 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 3.4). 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. 3.4 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 3.5. *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 3.6. *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 3.7. *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 $\Theta_C = \Theta \setminus \max\Theta$.*

3.3.5 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.[14]

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

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

[14]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

$$\exists l(\theta) \in \underset{f \in F \setminus \{h\}}{\operatorname{argmax}} \theta_f, \text{ such that } \forall \theta',$$
$$\theta'_h < \theta_h \implies \theta'_{l(\theta)} \leqslant \theta_{l(\theta)},$$
$$\theta_h < \theta'_h \implies \theta_{l(\theta)} < \theta'_{l(\theta)} < \theta'_h.$$

Then our results generalize.[15] 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, $\Pi^* = \Pi((\widehat{c}_\theta)_{\theta \in \Theta})$. 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 3.2, 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

$$\forall i, \theta^i_h = \theta^i_{l_i},$$
$$\forall i, j, \theta^i_h > \theta^i_{l_j}.$$

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

[15] 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.

the 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.[16] 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. Further-

[16]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).

more, ex-ante profit is strictly greater as the "revealed" contract for the intermediate type is more profitable than concealing him.

3.4 Extensions

3.4.1 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.[17] 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

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

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

$$s_\theta|_{\mathcal{E}^0} = \mathbf{id}$$

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

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

[17] A related idea is projection bias. (Loewenstein, O'Donoghue, and Rabin, 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.

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.

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. $\nu(\theta) = (s_\theta \underset{k \text{ times}}{\circ \cdots \circ} s_\theta)(e)$ 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

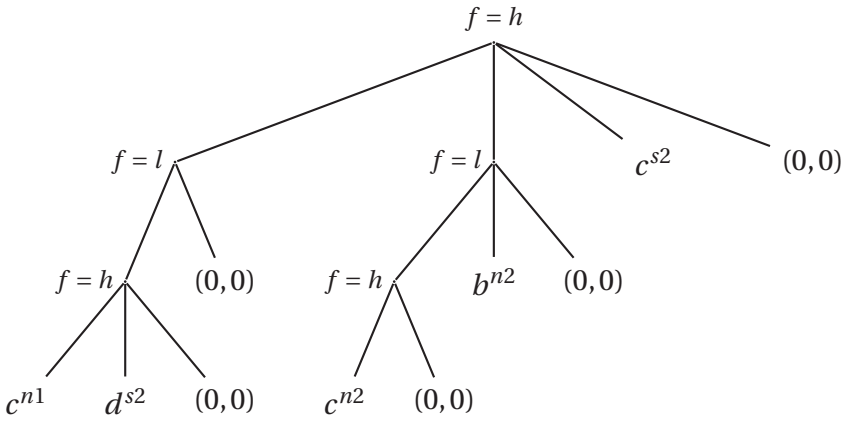
$$\begin{aligned} \Pi^* := \max_{e \in \mathcal{E}} & \Pi(\mathbf{c}) \\ \text{s.t. } & c_\theta \in \Sigma^e(\theta), \forall \theta \in \Theta_S, \\ & c_\theta \in N^e(\theta), \forall \theta \in \Theta_N. \end{aligned} \tag{3.7}$$

Optimal Structure and Contracts

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

Example 3.2. Recall from Example 3.1 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. 3.5. It implements $c^{n1} = (16, 4)$, $c^{n2} = (36, 6)$, $c^{s2} = (36, 6)$.

FIGURE 3.5: The optimal extensive-form decision problem in Example 3.2.



First, consider the sophisticated type. As in Example 3.1, the contract c^{n1} is more attractive than the implemented c^{s2}, but it is concealed using the decoy d^{s2}.

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.[18] 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

[18]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.

option in the continuation problem. Like θ^{n1} he reconsiders at the terminal node and ends up with c^{n2}. ▲

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 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 3.2.

Theorem 3.3. *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*

$$\sigma(\theta) = \sigma_S(\theta), \quad \forall \theta \in \Theta_S$$
$$\nu(\theta) = \widehat{c}_{\theta_h}, \quad \forall \theta \in \Theta_N.$$

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 3.1. 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.[19] In the following sense sophistication partially protects consumers from exploitation through the use of framing.

Observation 3.3. *For all types, the contract under sophistication is weakly preferred to the contract under naivete from the perspective of every frame.*[20]

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

[19] 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.

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

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 3.2). 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 (3.3) 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 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.

3.4.2 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.[21] 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

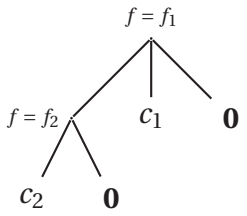
[21] E.g. directive 2011/83/EU: "the consumer should have the right of withdrawal because of the potential surprise element and/or psychological pressure".

in Salant and Siegel, 2018). One can interpret this decision as an additional *ex-post* participation decision at every terminal decision stage.

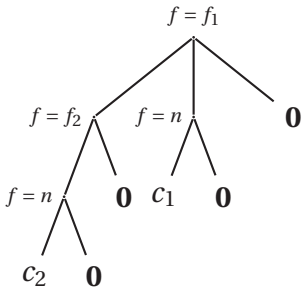
Formally, denote the neutral frame by $n \in F$, $n < h$.[22] This is the frame the consumer is in when unaffected by direct sales pressure by the firm.[23] We call $\bar{e} := (\{e, \mathbf{0}\}, n)$ an *interim modification*[24] of e. Then e is an *EDP with an interim participation constraint* if it 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 3.6: Interim and Ex-post Participation Constraints in Frame n

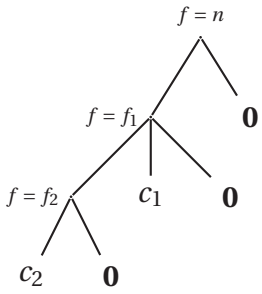
(a) An EDP e



(b) The ex-post modification $\underline{e}$



(c) The interim modification $\bar{e}$



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^n_θ. The following observation shows that the firm implements the efficient allocation associated with frame n and leaves no information rent to consumers.

Observation 3.4. *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,*

[22] It is immediate that additional participation decisions in frame h don't change the result, as all implemented contracts satisfy them.

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

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

are such that

$$\bar{\sigma}(\theta) = \underline{\sigma}(\theta) = \widehat{c}_{\theta_n}.$$

This observation is immediate from Theorem 3.1. The principal can remove all incoming IC constraints at the cost of an additional participation constraint in a lower frame. As 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.[25]

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.

Observation 3.5. *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 $\bar{\nu}$ and $\underline{\nu}$ satisfy*

[25] 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.

$\forall \theta \in \Theta$,

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

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

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.

3.5 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.[26]

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

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 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 3.2, 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). Sophis-

ticated 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.[27]

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

[27]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.

[28]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.

Appendix A

Supplemental Material

A.1 Appendix for Chapter 1: Proofs

Proof of Lemma 1.1 on page 16: Suppose that μ is such that $I_\mu = I$.

The threshold type must be indifferent between disclosing and not disclosing evidence, which implies

$$\bar{\theta}_{q,I} = \frac{(1-q)\mathbb{E}\mu + qF_\mu(\bar{\theta}_{q,I})\mathbb{E}_\mu(\theta|\theta \leqslant \bar{\theta}_{q,I})}{1-q+qF_\mu(\bar{\theta}_{q,I})}.$$

By rearranging and integrating by parts in $\mathbb{E}_\mu(\theta|\theta \leqslant \bar{\theta}_{q,I}) = \frac{1}{F_\mu(\bar{\theta}_{q,I})}\int_0^{\bar{\theta}_{q,I}} \theta \,\mathrm{d}F_\mu(\theta) = \bar{\theta}_{q,I} - \frac{I_\mu(\bar{\theta}_{q,I})}{F_\mu(\bar{\theta}_{q,I})}$, we obtain

$$(1-q+qF_\mu(\bar{\theta}_{q,I}))\bar{\theta}_{q,I} = (1-q)\mathbb{E}\mu + qF_\mu(\bar{\theta}_{q,I})\bar{\theta}_{q,I} - qI_\mu(\bar{\theta}_{q,I})$$
$$qI_\mu(\bar{\theta}_{q,I}) = (1-q)(\mathbb{E}\mu - \bar{\theta}_{q,I}).$$

Note that $\xi_{q,I} := I - \frac{1-q}{q}(\theta_0 - \mathbf{id})$ is a continuous, differentiable, strictly increasing function. Moreover, $\xi_{q,I}(0) \geqslant 0$ and $\xi_{q,I}(\theta_0) \leqslant 0$, with both inequalities strict if and only if $q \neq 1$ or $\inf\mathrm{supp}\,\mu = 0$.

In addition, since $\xi_{q,I}$ is strictly increasing in q and increasing in I (with respect to $\geqslant$), it follows that $\bar{\theta}_{q,I}$ is also strictly increasing in q and increasing in I.

□

Proof of Lemma 1.2 on page 21:

$$
\begin{aligned}
v(\mathscr{D}_q^V I) &= \int_0^1 (\mathscr{D}_q^V I - \underline{I})\,\mathrm{d}h \\
&= \int_0^{\bar{\theta}_{q,I}} (\mathscr{D}_q^V I - \underline{I})\,\mathrm{d}h + \int_{\bar{\theta}_{q,I}}^{\theta_0} (\mathscr{D}_q^V I - \underline{I})\,\mathrm{d}h + \int_{\theta_0}^1 (\mathscr{D}_q^V I - \underline{I})\,\mathrm{d}h \\
&= \int_{\bar{\theta}_{q,I}}^{\theta_0} (qI + (1-q)(\mathbf{id} - \theta_0))\,\mathrm{d}h + \int_{\theta_0}^1 \big(qI + (1-q)(\mathbf{id} - \theta_0) - (\mathbf{id} - \theta_0)\big)\,\mathrm{d}h \\
&= q\int_{\bar{\theta}_{q,I}}^{\theta_0} \left(I - \frac{1-q}{q}(\theta_0 - \mathbf{id})\right)\mathrm{d}h + q\int_{\theta_0}^1 (I - \underline{I})\,\mathrm{d}h \\
&= q\left(-\int_{\bar{\theta}_{q,I}}^{\theta_0} \frac{1-q}{q}(\theta_0 - \mathbf{id})\,\mathrm{d}h + \int_{\bar{\theta}_{q,I}}^1 I\,\mathrm{d}h - \int_0^1 \underline{I}\,\mathrm{d}h\right) \\
&= q\left(-\int_{\bar{\theta}_{q,I}}^{\theta_0} \frac{1-q}{q}(\theta_0 - \mathbf{id})\,\mathrm{d}h - \int_0^{\bar{\theta}_{q,I}} I\,\mathrm{d}h + \int_0^1 (I - \underline{I})\,\mathrm{d}h\right) \\
&= q\big(v(I) - L_q(I)\big)
\end{aligned}
$$

□

We now establish the following lemma, which will be useful in proving the main results.

Lemma A.1.1. *Fix any q. For each optimal I^*, there exists θ, such that I^* coincides with θ upper censorship I_θ on $[\bar{\theta}_{q,I^*}, 1]$. Moreover, I^* is disclosure-equivalent to I_θ: $\mathscr{D}_q^V I^* = \mathscr{D}_q^V I_\theta$.*

Proof. Take any optimal I^*. Consider two cases:

Case 1: $\bar{\theta}_{q,\bar{I}} \geqslant \hat{\omega}$. For any $I \in \mathcal{I}$, we have

$$
\begin{aligned}
v(\mathscr{D}_q^V I) &= q\int_{\bar{\theta}_{q,I}}^1 \left(I - \frac{1-q}{q}(\theta_0 - \mathbf{id})^+ - \underline{I}\right)\mathrm{d}h \\
&\leqslant 0 = q\int_{\theta_0}^1 (\underline{I} - 0 - \underline{I})\,\mathrm{d}h \\
&= q\int_{\bar{\theta}_{q,\underline{I}}}^1 \left(\underline{I} - \frac{1-q}{q}(\theta_0 - \mathbf{id})^+ - \underline{I}\right)\mathrm{d}h \\
&= v(\mathscr{D}_q^V \underline{I})
\end{aligned}
$$

Note that since I is continuous and h is strictly negative on $(\bar{\theta}_{q,I}, 1]$, it follows that the inequality strict if $I \neq \underline{I}$. Letting $\theta = 0$, yields $I^* = \underline{I} = I_0$.

Case 2: $\bar{\theta}_{q,\bar{I}} < \widehat{\omega}$. Apply Lemma A.3.4 to I^* to construct θ, such that $I_\theta - I^*$ is nonnegative on $[0, \omega]$ and nonpositive on $[\omega, 1]$. Note that this implies that $v(\mathscr{D}_q^V I_\theta) \geqslant v(\mathscr{D}_q^V I^*)$. Suppose, by contradiction, that $I_\theta \neq I^*$ on $(\bar{\theta}_{q,I}, 1]$. Then since h is strictly increasing on $[0, \omega]$ and strictly decreasing on $[\omega, 1]$, which implies $v(\mathscr{D}_q^V I_\theta) > v(\mathscr{D}_q^V I^*)$, a violation optimality of I^*.

Note that the image of $\mathscr{D}_q^V$ does not depend on values of an evidence structure below the disclosure threshold, which gives the second part. □

Proof of Theorem 1.1 on page 25: First, note that an optimum exists since $\mathcal{I}$ is compact and both v and $\mathscr{D}_q^V$ are continuous.

Fix some q and suppose I^* is an optimum for q. By Lemma A.1.1, there exists θ, such that I^* is disclosure-equivalent to the θ upper censorship I_θ. Therefore, one can reduce the sender's problem to finding optimal values of θ, which allows to recover I^* as the $\bar{\theta}_{q,\bar{I}_\theta}$ lower censorship of I_θ.

Formally, we have the following one-dimensional problem

$$\max_{\theta \in [0,1]} \tilde{v}_q(\theta), \tag{***}$$

where we define function

$$\begin{aligned} \tilde{v} : [0,1] \times [0,1) &\to \mathbb{R}_+, \\ (\theta, p) &\mapsto \tilde{v}_q(\theta) = v(I_\theta) - L_q(I_\theta). \end{aligned}$$

Note that $\tilde{v}_q$ is continuous and, therefore, attains maximum on $[0,1]$.

The following lemmata establish useful properties of the objective function v_q.

Lemma A.1.2. *There exists $\theta_1^* \in (0, \widehat{\omega}]$, such that $\tilde{v}_1$ is strictly increasing (decreasing) below (above) θ_1^*.*

Proof. We have

$$\begin{aligned}\tilde{v}_1(\theta) &= v(I_\theta) \\ &= \int_0^1 (I_\theta - \underline{I})\,\mathrm{d}h \\ &= \int_0^1 H\,\mathrm{d}(I_\theta)'_+ - H(\theta_0) \\ &= \int_0^\theta H\,\mathrm{d}\bar{I}' + (1-\bar{I}'(\theta))H\big(y(\theta)\big) - H(\theta_0),\end{aligned}$$

where $y(\theta) = \mathbb{E}(\mu_0|[\theta,1]) = \frac{\theta_0 + \bar{I}(\theta) - \theta I'(\theta)}{1-\bar{I}'(\theta)}$. The derivative of $\tilde{v}_1$ is given by

$$\tilde{v}_1'(\theta) = \bar{I}''(\theta)\big(H(\theta) - H(y(\theta)) - h(y(\theta))(y(\theta)-\theta)\big).$$

Consider equation $H(\theta) - H(x) - h(x)(x-\theta) = 0$. Since H is strictly convex over $[0,\widehat{\omega}]$ and strictly concave over $[\widehat{\omega},1]$, this equation has the unique solution in $[\widehat{\omega},1]$, denote it $x(\theta)$. Thus, we have two continuous functions x and y, such that x is strictly decreasing and y is strictly increasing. Let $\theta_1^* := \max\{\theta \in [0,1] : x(\theta) > y(\theta)\}$ and note that since $\operatorname{sign}(x-y) = \operatorname{sign}(\tilde{v}_1')$, $\tilde{v}_1$ is strictly increasing on $[0,\theta_1^*]$ and strictly decreasing on $[\theta^*,1]$.

Notice that $\theta_1^* \in (0,\widehat{\omega}]$, since $\theta_0 = y_0 < x_0 \in [\widehat{\omega},1]$ and $\widehat{\omega} = x(\widehat{\omega}) \leqslant y(\widehat{\omega}) = \mathbb{E}(\mu_0|[\widehat{\omega},1]) \geqslant \widehat{\omega}$. □

Lemma A.1.3. *$\tilde{v}$ has increasing marginal differences property in (θ,q): $\frac{\partial^2 \tilde{v}}{\partial q \partial \theta} \geqslant 0$. Moreover, it is strict on $(0,\bar{\theta}_{q,I}) \times (0,q]$ for any $q \in (0,1]$.*

Proof. Let $\ell_q^I := \min\left(I, \frac{1-q}{q}(\theta_0 - \mathbf{id})\right)$. For any $1 \geqslant q_1 \geqslant q_2 > 0$, we have

$$\begin{aligned}
\tilde{v}'_{q_1} - \tilde{v}'_{q_2} &= \frac{\mathrm{d}}{\mathrm{d}\theta}\left(L_{q_2}(I_\theta) - L_{q_1}(I_\theta)\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}\theta}\int_{\bar{\theta}_{q_1,I_\theta}}^{\theta_0}\left(\ell_{q_2}^{I_\theta} - \frac{q_1}{q_1}(\theta_0 - \mathbf{id})^+\right)\mathrm{d}h \\
&= \int_{\bar{\theta}_{q_1,I_\theta}}^{\bar{\theta}_{q_2,I_\theta}} \frac{\mathrm{d}I_\theta}{\mathrm{d}\theta}\,\mathrm{d}h \\
&= \int_{\bar{\theta}_{q_1,I_\theta}}^{\bar{\theta}_{q_2,I_\theta}} \bar{I}''(\theta)(\mathbf{id} - \theta)^+\,\mathrm{d}h \\
&\geqslant 0
\end{aligned}$$

Note that if $\theta \in (0, \bar{\theta}_{q_1,\bar{I}})$, then the integrand is strictly positive, which gives the strict condition.

△

We can establish the proof of Theorem 1.1. For any $q \in (0,1]$ denote the set of solutions to the one-dimensional program (***) as

$$\Theta_q^* := \operatorname*{argmax}_{\theta\in[0,1]} \tilde{v}_q(\theta).$$

Note that, by Berge's Maximum Theorem, $q \mapsto \Theta_q^*$ is upper hemi-continuous. Lemma A.1.3 allows to invoke Theorem 2.8.1 from Topkis (1998). This implies that $q \mapsto \Theta_q^*$ is non-decreasing with respect to the strong set order (Veinott order).

Note that the implication of Lemma A.1.2 is twofold. First, it implies that $\Theta_1^* = \{\theta_1^*\}$. Second, since $L_q(I_\theta)$ is constant in θ on $[\bar{\theta}_{q,\bar{I}}, 1]$, Lemma A.1.2 also implies that $\Theta_q^* \cap [\bar{\theta}_{q,\bar{I}}, 1] \subseteq \{\theta_1^*\}$ for any $q \in (0,1]$. In words, if there is a solution above the disclosure threshold, then it must be θ_1^*. Now define the threshold $\bar{q}$ as the the greatest lower bound on the values of q at which θ_1^* is the unique solution

$$\bar{q} := \inf\{q \in [0,1] : \Theta_q^* = \{\theta_1^*\}\}.$$

Note that if $\bar{q} = 0$, we are done, so assume $\bar{q} > 0$.

Next, we show that $\bar{q}$ must be strictly below 1. Suppose, by contradiction, that $\bar{q} = 1$. Take any sequence $\{q_n\}, \lim_{n\to\infty} q_n = 1, q_n \in (0,1)$. Since $\Theta^*_q \cap [\bar{\theta}_{q,I_\theta}, 1] \subseteq \{\theta^*_1\}$ for any q and $\lim_{n\to\infty} \bar{\theta}_{q_n,I_\theta} = 0$ for any $\theta \in [0,1]$, it follows that $\limsup \Theta^*_{q_n} < \bar{\theta}_{q,I_\theta}$, which violates upper hemi-continuity of $q \mapsto \Theta^*_q$.

It is left to show that, for every $q < \bar{q}$, any optimum is a binary certification. By upper hemi-continuity of $q \mapsto \Theta^*_q$, the set $\Theta^*_{\bar{q}}$ must contain some $\theta^*_{\bar{q}} < \theta^*_1$.

The derivations in the proof of Lemma A.1.3 imply that

$$\frac{\partial}{\partial q-} L_q(\theta^*_{\bar{q}}) < \frac{\partial}{\partial q} L_q(\theta^*_1).$$

Thus, both θ^*_1 and $\theta_{\bar{q}}$ are optimal at $\bar{q}$ and the marginal reduction in the concealment loss is strictly higher for $\theta^*_{\bar{q}}$ than for θ^*_1 if q decreases. Therefore, there exists $\varepsilon > 0$, such that θ^*_1 cannot be optimal for any $q \in (\bar{q} - \varepsilon, \bar{q}]$. But then because $q \mapsto \Theta^*_q$ is non-decreasing in the strong set order, it means that θ^*_1 is never optimal for $q < \bar{q}$.

Finally, since $\Theta^*_q \subseteq [0, \bar{\theta}_{q,\bar{I}}]$ for any $q < \bar{q}$, the optimal I is a lower censorship of the upper censorship with a threshold below $\bar{\theta}_{q,\bar{I}}$, which is a binary certification.

□

Proof of Theorem 1.2 on page 28: The result follows from the following lemma, which that any selection from $q \mapsto \Theta^*_q$ is strictly increasing on $[0, \bar{q}]$. Since there exist $\theta^*_{\bar{q}} \in \Theta^*_{\bar{q}} \cap [0, \bar{\theta}_{q,\bar{I}}]$ and $q \mapsto \bar{\theta}_{q,\bar{I}}$ is strictly decreasing, the lemma then implies that for any $q < \bar{q}$, $\sup \Theta^*_q \leqslant \bar{\theta}_{q,\bar{I}}$, which means that the optimum is a binary certification.

Lemma A.1.4. *Any selection from $q \mapsto \Theta^*_q$ is strictly increasing on $[\bar{q}, 1)$.*

Proof. Notice that since $\Theta_q^* \setminus \{\theta_1^*\} \subseteq [0, \bar{\theta}_{q,\bar{I}}]$ for $q \geqslant \bar{q}$, if one makes the objective function smaller on $(\bar{\theta}_{q,\bar{I}}, 1]$, it will not change the set of maximizers. Define

$$\widehat{v}_q(\theta) = \begin{cases} \tilde{v}_q(\theta), & \theta \leqslant \bar{\theta}_{q,\bar{I}} \\ \tilde{v}_q(\theta) - q\left(\theta - \bar{\theta}_{q,\bar{I}}\right)^2, & \theta > \bar{\theta}_{q,\bar{I}}, \end{cases}$$

so that for $q > \bar{q}$,

$$[0, \bar{\theta}_{q,\bar{I}}] \cap \operatorname*{argmax}_{\theta \in [0,1]} \widehat{v}_q(\theta) = [0, \bar{\theta}_{q,\bar{I}}] \cap \operatorname*{argmax}_{\theta \in [0,1]} \tilde{v}_q(\theta).$$

Using Lemma A.1.3 and strict monotonicity of $q \mapsto \bar{\theta}_{q,\bar{I}}$, we conclude that $\tilde{w}$ satisfies strictly increasing marginal differences property in (θ, q) and, therefore, Strict Monotonicity Theorem 1 from Edlin and Shannon (1998) applies. Since 0 is never optimal for any $q > 0$, it implies that any selection from $q \mapsto \Theta_q^*$ is strictly increasing on $(0, \bar{q}]$, which gives the desired result.

△

□

Proof of Proposition 1.2 on page 31: We will show that $\frac{v_q^*}{q}$ is strictly increasing in q, which implies that v_q^* is strictly increasing in q.

By Lemma 1.2, we have

$$\frac{v_q^*}{q} = \max_{I \in \mathcal{I}} v(I) - L_q(I).$$

Invoking the Envelope Theorem, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}q}\frac{v_q^*}{q} = -\frac{\mathrm{d}L_q(I)}{\mathrm{d}q}\Big|_{I=I_q^*} = -\int_0^{\theta_0} \frac{\mathrm{d}\ell_q^I}{\mathrm{d}q}\,\mathrm{d}h \,\Big|_{I=I_q^*} = \int_{\bar{\theta}_{q,I}}^{\theta_0} \frac{1}{q^2}(\theta_0 - \mathbf{id})\,\mathrm{d}h \,\Big|_{I=I_q^*} > 0,$$

where the inequality holds for any optimal I_q^*. □

Proof of Proposition 1.3 on page 33: We will show that $\frac{w^*_q}{q}$ is strictly increasing in q, which implies that w^*_q is strictly increasing in q.

It follows from the sender's one-dimensional problem (***), given in the proof of Theorem 1.1, that is is enough to show that $\frac{w(\mathscr{D}^V_q I_{\theta^*_q})}{q}$ is strictly increasing in q, where I_θ denotes the θ upper censorship. We have

$$\begin{aligned}\frac{w(\mathscr{D}^V_q I_{\theta^*_q})}{q} &= w(I_{\theta^*_q}) - L_q(I_{\theta^*_q}) \\ &= \int_{\bar\theta_{q,I_{\theta^*_q}}}^1 (I_{\theta^*_q} - \underline{I})\,\mathrm{d}H \\ &= \int_0^{\bar\theta_{\bar q,\bar I}} (I_{\theta^*_q} - \underline{I})\,\mathrm{d}H + \int_{\bar\theta_{\bar q,\bar I}}^{\theta_0} (I_{\theta^*_q} - \underline{I})\,\mathrm{d}H + \int_{\theta_0}^1 (I_{\theta^*_q} - \underline{I})\,\mathrm{d}H,\end{aligned}$$

where all three terms are increasing in q. Then $\frac{w(\mathscr{D}^V_q I_{\theta^*_q})}{q}$ is strictly increasing, since $I_{\theta^*_q}|_{[\bar\theta_{\bar q,\bar I},\theta_0]}$ is $\gg$-increasing in q and $[\bar\theta_{\bar q,\bar I},\theta_0] \subseteq [\bar\theta_{q,I_{\theta^*_q}},1]$.

□

A.2 Appendix for Chapter 2: Proof Exposition

This appendix provides exposition for the paper's proofs. The exposition is not formally necessary, and so a reader interested solely in our formal arguments may proceed directly to Appendix A.3.

We begin by explaining how to visualize Theorem 2.1's program. Using this visualization, we provide intuitions for Proposition 2.1, Proposition 2.2 and Proposition 2.3. Finally, we elaborate on the main text's exposition for Claim 2.1.

A.2.1 Visualizing Theorem 2.1

We now explain how to use Theorem 2.1 to graphically solve for S's optimal equilibrium value when Θ is binary, $\Theta = \{\theta_1, \theta_2\}$. Consider Figure A.1, which visualizes constraints (R-BP) and (χ-BP) for the binary-state case. In this figure, the horizontal axis is the mass on θ_1, and the vertical axis is the mass on θ_2. Because μ_0, β, and γ assign a total probability of 1 to both states, each of them can be represented as a point on the line connecting the two atomistic beliefs δ_{θ_1} and δ_{θ_2}. Every point underneath this line represents the product $(1-k)\gamma$ for some k and γ. The drawn box represents the constraints in Theorem 2.1's program. By (χ-BP), $(1-k)\gamma$ must be pointwise larger than $[\mathbf{1} - \chi(\cdot)]\mu_0$, which is the box's bottom-left corner. The box's top-right corner, which corresponds to the prior, must be pointwise larger than $(1-k)\gamma$ by (R-BP).[1] In other words, $(1-k)\gamma$ must lie within the drawn box. Once $(1-k)\gamma$ is chosen, one can recover γ and β by finding the unique points on the line $[\delta_{\theta_1}, \delta_{\theta_2}]$ that lie in the same direction as $(1-k)\gamma$ and $\mu_0 - (1-k)\gamma$, respectively.

Figure A.2 shows how to simultaneously visualize the constraint illustrated in Figure A.1 and S's value for the introduction's example, where χ is a constant $x\mathbf{1}$. Such a visualization enables us to solve for S's optimal equilibrium value. To do so, we start by

[1] To see this requirement, rearrange (R-BP) to obtain that $\mu_0 - (1-k)\gamma = k\beta \geqslant 0$.

(a) Construction of γ and β for a given $(1-k)\gamma$ (b) γ' is infeasible

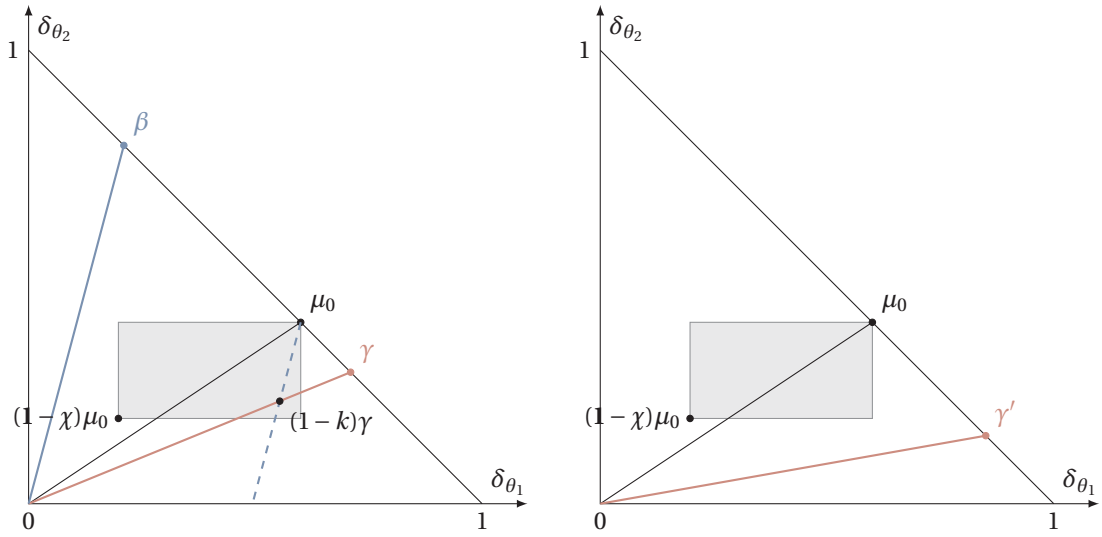


FIGURE A.1: Constraints (R-BP) and (χ-BP) and construction of γ and β for a given $(1-k)\gamma$.

drawing $\bar{v}$, the quasiconcave envelope of S's value function. For each feasible candidate $(1-k)\gamma$, we find the corresponding β, as in Figure A.1. To calculate S's value from the resulting (β, γ, k), we simply find the value above μ_0 of the line connecting the points $(\beta, \widehat{v}_{\wedge\gamma}(\beta))$ and $(\gamma, \bar{v}(\gamma))$.

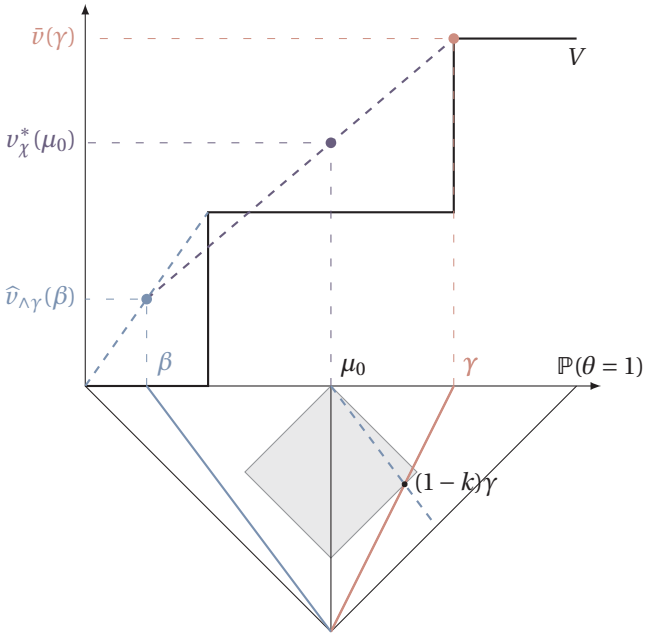


FIGURE A.2: An illustration of the solution to Theorem 2.1's program for the example from the introduction, with a constant credibility level between $\frac{2}{3}$ and $\frac{3}{4}$.

A.2.2 Exposition for Proposition 2.1

This section sketches the argument behind Proposition 2.1. The proposition builds on the binary-state case. In this case, genericity implies $\bar{v}$ has a non-degenerate interval of maximizers, and S not being an SOB implies $\widehat{v}$ has a kink somewhere outside of this interval. Fixing a prior near this interval, but toward the nearest kink, we then find the lowest constant $x \in [0,1]$ such that S still obtains her full credibility value at $\chi(\cdot) = x\mathbf{1}$. At this $\chi(\cdot)$, S's favorite equilibrium information policy is unique and is supported on the beliefs (γ, β) that solve Theorem 2.1's program. These beliefs are interior, and $\widehat{v}$ has a kink at β. Although γ remains optimal in Theorem 2.1's program for any additional small reduction in credibility, (χ-BP) forces the optimal β to move away from the prior. Relying on the set of beliefs being one-dimensional, we show the only incentive-compatible way of attaining S's new optimal value is to spread the original β between γ and a further posterior that gives S an even lower continuation value than under β. Hence, S provides R with more information. The reduction in S's value indicates a change in R's optimal behavior. In other words, the additional information is instrumental, strictly increasing R's utility. Figure A.3 illustrates the argument using our introductory example.

A.2.3 Exposition for Proposition 2.2

This section describes the proof of Proposition 2.2. Notice that two of the proposition's three implications are immediate. First, whenever no conflict occurs, S can reveal the state in an incentive-compatible way while obtaining her first-best payoff (given R's incentives), meaning commitment is of no value; that is, (iii) implies (ii). Second, because S's highest equilibrium value increases with her credibility, commitment having no value means S's best equilibrium value is constant (and, a fortiori, continuous) in the credibility level; that is, (ii) implies (i).

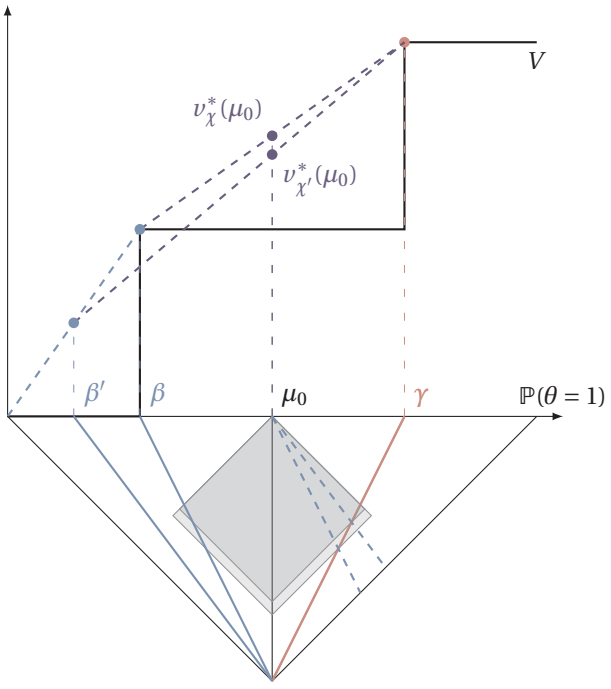


FIGURE A.3: An illustration of Proposition 2.1's proof for two states. The argument begins by identifying a μ_0 as above. Given μ_0, we find two constant $\chi(\cdot) > \chi'(\cdot)$ as above, yielding the constraints depicted by the light and dark boxes, respectively. Whereas γ is optimal under both credibility levels, β is optimal under χ, whereas β' is optimal under χ'. One can then deduce R is strictly better off under χ' than under χ.

To show that (i) implies (iii), we show that any failure of (iii) implies the failure of (i). To do so, we fix a full-support prior μ_0 at which $\bar{v}$ is minimized. Because conflict occurs, $\bar{v}$ is nonconstant and thus takes values strictly greater than $\bar{v}(\mu_0)$. By Theorem 2.1, one has that $v^*_\chi(\mu_0) > \bar{v}(\mu_0)$ if and only if some feasible triplet (β, γ, k), with $k < 1$ exists such that $\bar{v}(\gamma) > \bar{v}(\mu_0)$. Using upper semicontinuity of $\bar{v}$, we show such a triplet is feasible for a constant credibility $\chi(\cdot) = x\mathbf{1}$ if and only if x is weakly greater than some strictly positive x^*. We thus have that for all $x < x^*$,

$$v^*_{x^*\mathbf{1}}(\mu_0) \geqslant k\bar{v}(\mu_0) + (1-k)\bar{v}(\gamma) > \bar{v}(\mu_0) = v^*_{x\mathbf{1}}(\mu_0),$$

where the first inequality follows from μ_0 minimizing $\bar{v}$; that is, a collapse of trust occurs. Figure A.4 below illustrates the argument in the context of our leading example.

The figure depicts a prior that minimizes S's payoff under no credibility. The depicted constraint set is drawn for $\chi^* = x^*\mathbf{1}$, the lowest constant credibility for which a (β, γ, k) satisfying both $k < 1$ and $\bar{v}(\gamma) > \bar{v}(\mu_0)$ is feasible. In other words, $\chi^*(\cdot)$ is the lowest constant credibility at which S's value is strictly above $\bar{v}(\mu_0)$. Therefore, $v^*_{x^*\mathbf{1}}(\mu_0) > v^*_{x\mathbf{1}}(\mu_0)$ for any x strictly below x^*.

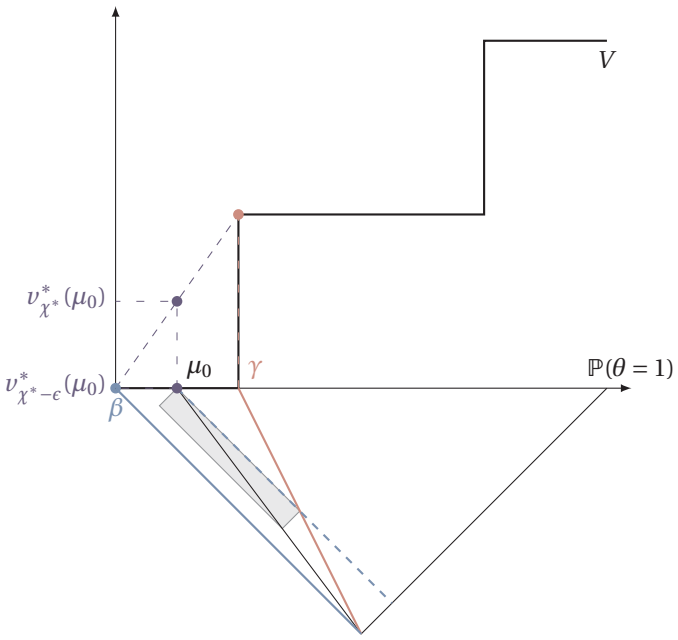


FIGURE A.4: An illustration of the Proposition 2.2's proof in the context of the introduction's example. The proof starts with choosing a prior minimizing the payoff S obtains under no credibility. We then identify x^*, the lowest credibility level for which a (β, γ, k) attaining a value strictly above $\bar{v}(\mu_0)$ is feasible at $\chi^* = x^*\mathbf{1}$. By choice of x^*, $v^*_{x^*\mathbf{1}}(\mu_0) > v^*_{x\mathbf{1}}(\mu_0)$ must hold for any $x = x^* - \epsilon < x^*$; that is, S's value collapses.

A.2.4 Exposition for Proposition 2.3

This section discusses the proof of Proposition 2.3 that is based on establishing a four-way equivalence between (a) S getting the benefit of the doubt, (b) $\bar{v}$ being maximized by a full-support prior γ, (c) a full-support γ existing such that $\widehat{v}_{\wedge\gamma}$ and $\widehat{v}$ agree over all full-support prior(s), (d) robustness to limited credibility. That (a) is equivalent to (b) follows from the arguments of Lipnowski and Ravid (2019). For the equivalence of (b) and (c), note that in finite models $\widehat{v}$ and $\widehat{v}_{\wedge\gamma}$ are both continuous. Therefore, the

two functions agree over all full-support priors if and only if they are equal, which is equivalent to the cap on $v_{\wedge\gamma}$ being non-binding; that is, γ maximizes $\bar{v}$. To see why (c) is equivalent to (d), fix some full-support μ_0, and consider two questions about Theorem 2.1's program. First, which beliefs can serve as γ for $\chi(\cdot) \ll \mathbf{1}$ large enough?[2] Second, how do the optimal (k, β) for a given γ change as $\chi(\cdot)$ goes to $\mathbf{1}$? Figure A.5 illustrates the answer to both questions for the two-state case. For the first question, the answer is that γ is feasible for some $\chi(\cdot) \ll \mathbf{1}$ if and only if γ has full support. For the second question, one can show it is always optimal to choose (k, β) so as to make (χ-BP) bind while still satisfying (R-BP).[3] Direct computation reveals that, as $\chi(\cdot)$ goes to $\mathbf{1}$, every such (k, β) must converge to $(1, \mu_0)$. Combined, one obtains that, as $\chi(\cdot)$ increases, S's optimal value converges to $\max_{\gamma \in \text{int}(\Delta\Theta)} \widehat{v}_{\wedge\gamma}(\mu_0)$. Thus, S's value is robust to limited credibility if and only if some full-support γ exists for which $\widehat{v}_{\wedge\gamma} = \widehat{v}$ for all full-support priors; that is, (c) is equivalent to (d). The proposition follows.

A.2.5 Exposition for Claim 2.1

This section provides some intuition for Claim 2.1. Let us first explain why $v^*_\chi(\mu_0) \geqslant \widehat{v}(\bar{\mu}_\chi)$. As explained in the main text, $\widehat{v}(\bar{\mu}_\chi) = \int H \, \mathrm{d}\mu_{\chi,\theta^*}$, where μ_{χ,θ^*} is a θ^* upper censorship of $\bar{\mu}_\chi$ for some $\theta^* \in [0,1]$. Because $\bar{\mu}_\chi$'s support is in $[0, \bar{\theta}_\chi]$, any θ upper censorship of $\bar{\mu}_\chi$ for a θ above $\bar{\theta}_\chi$ is just $\bar{\mu}_\chi$ itself. Thus, assuming θ^* is in $[0, \bar{\theta}_\chi]$ is without loss. Given such a θ^*, one can induce the posterior mean distribution μ_{χ,θ^*} in a χ-equilibrium (with the original prior μ_0) using a θ^*-upper-censorship pair. As such,

[2] By $\chi(\cdot) \ll \mathbf{1}$, we mean $\chi(\theta) < 1$ for all $\theta \in \Theta$.

[3] To see why, for any feasible (k, β, γ), a (k', β') exists such that (k', β', γ) is feasible, (χ-BP) binds, and $k' \geqslant k$. By (R-BP), $\beta' = \frac{k}{k'}\beta + \left(1 - \frac{k}{k'}\right)\gamma$. Because $\widehat{v}_{\wedge\gamma}$ is concave and $\widehat{v}_{\wedge\gamma}(\gamma) = \bar{v}(\gamma)$,

$$\begin{aligned} k'\widehat{v}_{\wedge\gamma}(\beta') + (1-k')\,\bar{v}(\gamma) &= k'\widehat{v}_{\wedge\gamma}\left(\tfrac{k}{k'}\beta + \left(1-\tfrac{k}{k'}\right)\gamma\right) + (1-k')\,\bar{v}(\gamma) \\ &\geqslant k\widehat{v}_{\wedge\gamma}(\beta) + (k'-k)\,\widehat{v}_{\wedge\gamma}(\gamma) + (1-k')\,\bar{v}(\gamma) = k\widehat{v}_{\wedge\gamma}(\beta) + (1-k)\,\bar{v}(\gamma). \end{aligned}$$

(a) The set of feasible $(1-k)\gamma$ as $\chi(\cdot) \to \mathbf{1}$ (b) (k, β) converges to $(1, \mu_0)$

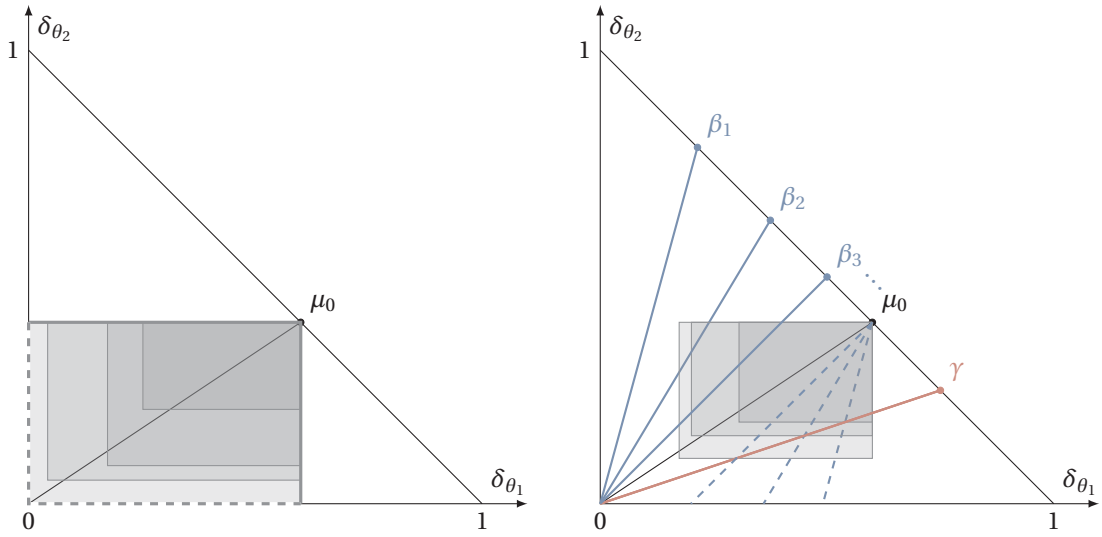


FIGURE A.5: Robustness to limited credibility

S's maximal χ-equilibrium value is at least as high as the value generated by μ_{χ,θ^*}; that is, $v^*_\chi(\mu_0) \geqslant \int H \,\mathrm{d}\mu_{\chi,\theta^*} = \widehat{v}(\bar{\mu}_\chi)$.

We now sketch the reasoning behind $v^*_\chi(\mu_0) \leqslant \widehat{v}(\bar{\mu}_\chi)$. Suppose (β, γ, k) solves Theorem 2.1's program. Because cheap talk is equivalent to no information (as explained earlier in this section), one can attain $\bar{v}(\gamma)$ with a single message that induces a posterior mean of $E\gamma$. Therefore, $v_{\wedge\gamma}(\mu) = H(E\gamma) \wedge H(E\mu)$, meaning $\widehat{v}_{\wedge\gamma}(\beta)$ is given by

$$\widehat{v}_{\wedge\gamma}(\beta) = \max_{\tilde{\beta} \preccurlyeq \beta} \int H(E\gamma) \wedge H(\cdot) \,\mathrm{d}\tilde{\beta}.$$

Using optimality of (β, γ, k), one can show the above program is solved by a $\tilde{\beta}$ whose support lies in $[0, E\gamma]$. As such, H's expected value according to $\tilde{\mu} := k\tilde{\beta} + (1-k)\delta_{E\gamma}$ equals S's maximal χ-equilibrium value; that is, $v^*_\chi(\mu_0) = \int H \,\mathrm{d}\tilde{\mu}$. Hence, a sufficient condition for $v^*_\chi(\mu_0) \leqslant \widehat{v}(\bar{\mu}_\chi)$ is that $\tilde{\mu} \preccurlyeq \bar{\mu}_\chi$. In other words, it suffices to establish that (MPS) holds for $\bar{\mu}_\chi$ and $\tilde{\mu}$ for all $\widehat{\theta}$. To establish (MPS) for $\widehat{\theta} \geqslant E\gamma$, we use two facts. First, $\tilde{\mu}[0, \theta] = 1 \geqslant \bar{\mu}_\chi[0, \theta]$ holds for all $\theta \geqslant E\gamma$. And, second, both $\bar{\mu}_\chi$ and $\tilde{\mu}$ admit μ_0 as a mean-preserving spread. As such, $\int_0^{\tilde{\theta}} (\tilde{\mu}[0, \theta] - \bar{\mu}_\chi[0, \theta]) \,\mathrm{d}\theta$ decreases in $\widehat{\theta}$ over $[E\gamma, 1]$

and reaches a value of zero at $\tilde{\theta} = 1$. It follows that (MPS) holds for $\bar{\mu}_\chi$ and $\tilde{\mu}$ for all $\tilde{\theta} \geqslant E\gamma$. To establish (MPS) for $\widehat{\theta} < E\gamma$, notice that $\tilde{\mu}[0,\theta] = k\tilde{\beta}[0,\theta]$ whenever $\theta < E\gamma$. Therefore, if $\widehat{\theta} < E\gamma$,

$$\begin{aligned}\int_0^{\widehat{\theta}} \tilde{\mu}[0,\theta]\,\mathrm{d}\theta &= k\int_0^{\widehat{\theta}} \tilde{\beta}[0,\theta]\,\mathrm{d}\theta \leqslant k\int_0^{\widehat{\theta}} \beta[0,\theta]\,\mathrm{d}\theta \\ &= \int_0^{\widehat{\theta}} (\mu_0 - (1-k)\gamma)[0,\theta]\,\mathrm{d}\theta \leqslant \int_0^{\widehat{\theta}} \chi\mu_0[0,\theta]\,\mathrm{d}\theta \leqslant \int_0^{\widehat{\theta}} \bar{\mu}_\chi[0,\theta]\,\mathrm{d}\theta,\end{aligned}$$

where the first inequality follows from $\beta \succcurlyeq \tilde{\beta}$, the second equality from (R-BP), and the second inequality from (χ-BP).

A.3 Appendix for Chapter 2: Proofs

We first introduce some convenient notation that we will use below. For a compact metrizable space, Y, and $f : Y \to \mathbb{R}$ bounded and measurable, let $f(\gamma) := \int_Y f \, \mathrm{d}\gamma$.

A.3.1 Toward the Proof of the Main Theorem

To present unified proofs, we adopt the notational convention that $\frac{0}{0} = 1$ wherever it appears.

Characterization of All Equilibrium Outcomes

En route to our characterization of the sender-preferred equilibrium outcomes, we characterize the full range of equilibrium outcomes.

Definition A.3.1. $(p, s_o, s_i) \in \Delta\Delta\Theta \times \mathbb{R} \times \mathbb{R}$ is a χ-**equilibrium outcome** if there exists a χ-equilibrium $(\xi, \sigma, \alpha, \pi)$ such that, letting $P_o := \frac{1}{\chi(\mu_0)} \int_\Theta \chi\xi \, \mathrm{d}\mu_0$ and $P_i := \frac{1}{1-\chi(\mu_0)} \int_\Theta (\mathbf{1} - \chi)\sigma \, \mathrm{d}\mu_0$ be the equilibrium distributions over M conditional on official and influenced reporting, respectively, we have: $p = [\chi(\mu_0)P_o + [1 - \chi(\mu_0)]P_i] \circ \pi^{-1}$, $s_o = u_S\left(\int_M \alpha \, \mathrm{d}P_o\right)$, and $s_i = u_S\left(\int_M \alpha \, \mathrm{d}P_i\right)$.

The following lemma adopts a belief-based approach, directly characterizing χ-equilibrium outcomes of our game.

Lemma A.3.1. *Fix $(p, s_o, s_i) \in \Delta\Delta\Theta \times \mathbb{R} \times \mathbb{R}$. Then (p, s_o, s_i) is a χ-equilibrium outcome if and only if there exists $k \in [0, 1]$, b, $g \in \Delta\Delta\Theta$ such that*

(i) $kb + (1 - k)g = p \in \mathcal{R}(\mu_0)$;

(ii) $(1 - k) \int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu) \geqslant (\mathbf{1} - \chi)\mu_0$;

(iii) $g\{\mu \in \Delta\Theta : s_i \in V(\mu)\} = b\{\mu \in \Delta\Theta : \min V(\mu) \leqslant s_i\} = 1$;

(iv) $\left[1-\chi(\mu_0)\right] s_i + \chi(\mu_0) s_o \in (1-k)\, s_i + k \int_{\mathrm{supp}\,(b)} s_i \wedge V \,\mathrm{d}b$.[4]

Proof. As M is an uncountable Polish space, Kuratowski's theorem says M is isomorphic (as a measurable space) to $\{0,1\} \times \Delta\Theta$. We can therefore assume without loss that $M = \{0,1\} \times \Delta\Theta$.

First, suppose $k \in [0,1]$, $g, b \in \Delta\Delta\Theta$ satisfy the four listed conditions. Let ϕ be a measurable selector of $s_i \wedge V|_{\mathrm{supp}\,(b)}$ with $s_o = \left(1 - \frac{k}{\chi(\mu_0)}\right) s_i + \frac{k}{\chi(\mu_0)} \int_{\mathrm{supp}\,(b)} \phi \,\mathrm{d}b$.

Define $D := \mathrm{supp}\,(p)$, $\beta := \int_{\Delta\Theta} \mu \,\mathrm{d}b(\mu)$, and $\gamma := \int_{\Delta\Theta} \mu \,\mathrm{d}g(\mu)$. Let measurable $\eta_g, \eta_b : \Theta \to \Delta\Delta\Theta$ be signals that induce belief distribution g for prior γ and belief distribution b for prior β, respectively.[5] That is, for every Borel $\widehat{\Theta} \subseteq \Theta$ and $\widehat{D} \subseteq \Delta\Theta$,

$$\int_{\widehat{\Theta}} \eta_b(\widehat{D}|\cdot)\,\mathrm{d}\beta = \int_{\widehat{D}} \mu(\widehat{\Theta})\,\mathrm{d}b(\mu) \text{ and } \int_{\widehat{\Theta}} \eta_g(\widehat{D}|\cdot)\,\mathrm{d}\gamma = \int_{\widehat{D}} \mu(\widehat{\Theta})\,\mathrm{d}g(\mu).$$

Take some Radon-Nikodym derivative $\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0} : \Theta \to \mathbb{R}_+$; changing it on a μ_0-null set, we may assume that $\mathbf{0} \leqslant \frac{k}{\chi} \frac{\mathrm{d}\beta}{\mathrm{d}\mu_0} \leqslant \mathbf{1}$ since $(1-k)\gamma \geqslant (\mathbf{1}-\chi)\mu_0$.

Next, define the sender's influenced strategy and reporting protocol $\sigma, \xi : \Theta \to \Delta M$ by letting, for every Borel $\widehat{M} \subseteq M$,

$$\begin{aligned}
\sigma(\widehat{M}|\cdot) &:= \eta_g\left(\{\mu \in D :\ (0,\mu) \in \widehat{M}\} \;\middle|\; \cdot\right), \\
\xi(\widehat{M}|\cdot) &:= \left[\mathbf{1} - \tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right] \eta_g\left(\{\mu \in D :\ (0,\mu) \in \widehat{M}\} \;\middle|\; \cdot\right) \\
&\quad + \tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0} \eta_b\left(\{\mu \in D :\ (1,\mu) \in \widehat{M}\} \;\middle|\; \cdot\right).
\end{aligned}$$

[4]Here, $s_i \wedge V : \Delta\Theta \rightrightarrows \mathbb{R}$ is the correspondence with $s_i \wedge V(\mu) = (-\infty, s_i] \cap V(\mu)$; it is a Kakutani correspondence (because V is) on the restricted domain $\mathrm{supp}\,(b)$. The integral is the (Aumann) integral of a correspondence:

$$\int_{\mathrm{supp}\,(b)} s_i \wedge V \,\mathrm{d}b = \left\{ \int_{\mathrm{supp}\,(b)} \phi \,\mathrm{d}b :\ \phi \text{ is a measurable selector of } s_i \wedge V|_{\mathrm{supp}\,(b)} \right\}.$$

[5]These are the partially informative signals about $\theta \in \Theta$ such that it is Bayes-consistent for the listener's posterior belief to equal the message.

Now, fix some $\widehat{\mu} \in D$ and $\widehat{a} \in \operatorname{argmax}_{a \in A} u_R(a, \widehat{\mu})$ with $u_S(\widehat{a}) \leqslant s_i$; we can then define a receiver belief map as

$$\pi : M \to \Delta\Theta$$
$$m \mapsto \begin{cases} \mu & : \ m \in \{0,1\} \times \{\mu\} \text{ for } \mu \in D \\ \widehat{\mu} & : \ m \notin \{0,1\} \times D. \end{cases}$$

Finally, by Lipnowski and Ravid (2019, Lemma 2), there are some measurable $\alpha_b, \alpha_g : \operatorname{supp}(p) \to \Delta A$ such that:[6]

- $\alpha_b(\mu), \alpha_g(\mu) \in \operatorname{argmax}_{\tilde{\alpha} \in \Delta A} u_R(\tilde{\alpha}, \mu) \ \forall \mu \in \operatorname{supp}(p)$;
- $u_S(\alpha_b(\mu)) = \phi(\mu) \ \forall \mu \in \operatorname{supp}(b)$, and $u_S(\alpha_g(\mu)) = s_i \ \forall \mu \in \operatorname{supp}(g)$.

From these, we can define a receiver strategy as

$$\alpha : M \to \Delta A$$
$$m \mapsto \begin{cases} \alpha_b(\mu) & : \ m = (1, \mu) \text{ for } \mu \in D \\ \alpha_g(\mu) & : \ m = (0, \mu) \text{ for } \mu \in D \\ \delta_{\widehat{a}} & : \ m \notin \{0,1\} \times D. \end{cases}$$

We want to show that the tuple $(\xi, \sigma, \alpha, \pi)$ is a χ-equilibrium resulting in outcome (p, s_o, s_i). It is immediate from the construction of (σ, α, π) that sender incentive compatibility and receiver incentive compatibility hold, and that the expected sender payoff is s_i given influenced reporting.

Recall $\chi\xi : \Theta \to \Delta M$ is defined as the pointwise product, i.e. for every $\theta \in \Theta$ and Borel $\widehat{M} \subseteq M$, we have $(\chi\xi)(\widehat{M}|\theta) = \chi(\theta)\xi(\widehat{M}|\theta)$; and similarly for $(\mathbf{1} - \chi)\sigma$. To see that

[6]The cited lemma will exactly deliver $\alpha_b|_{\operatorname{supp}(b)}, \alpha_g|_{\operatorname{supp}(g)}$. Then, as $\operatorname{supp}(p) \subseteq \operatorname{supp}(b) \cup \operatorname{supp}(g)$, we can extend both functions to the rest of their domains by making them agree on $\operatorname{supp}(p) \setminus [\operatorname{supp}(b) \cap \operatorname{supp}(g)]$.

the Bayesian property holds, observe that every Borel $\widehat{D} \subseteq D$ satisfies

$$\begin{aligned}
[(\mathbf{1}-\chi)\sigma + \chi\xi](\{1\} \times \widehat{D}|\cdot) &= k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\eta_b(\widehat{D}|\cdot) \\
[(\mathbf{1}-\chi)\sigma + \chi\xi](\{0\} \times \widehat{D}|\cdot) &= \left[(\mathbf{1}-\chi) + \chi\left(\mathbf{1} - \tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\right]\eta_g(\widehat{D}|\cdot) \\
&= \left(\mathbf{1} - k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\eta_g(\widehat{D}|\cdot).
\end{aligned}$$

Now, take any Borel $\widehat{M} \subseteq M$ and $\widehat{\Theta} \subseteq \Theta$, and let $D_z := \{\mu \in D : \ (z,\mu) \in \widehat{M}\}$ for $z \in \{0,1\}$. Observe that

$$\begin{aligned}
&\int_\Theta \int_{\widehat{M}} \pi(\widehat{\Theta}|\cdot)\,\mathrm{d}[(1-\chi(\theta))\sigma + \chi(\theta)\xi](\cdot|\theta)\,\mathrm{d}\mu_0(\theta) \\
=\ &\int_\Theta \int_{\widehat{M}\cap[\{0,1\}\times D]} \pi(\widehat{\Theta}|\cdot)\,\mathrm{d}[(1-\chi(\theta))\sigma + \chi(\theta)\xi](\cdot|\theta)\,\mathrm{d}\mu_0(\theta) \\
=\ &\int_\Theta \left(\int_{\{1\}\times D_1} + \int_{\{0\}\times D_0}\right) \pi(\widehat{\Theta}|\cdot)\,\mathrm{d}[(1-\chi(\theta))\sigma + \chi(\theta)\xi](\cdot|\theta)\,\mathrm{d}\mu_0(\theta) \\
=\ &\int_\Theta \left[k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}(\theta)\int_{D_1}\mu(\widehat{\Theta})\,\mathrm{d}\eta_b(\mu) + \left(1 - k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}(\theta)\right)\int_{D_0}\mu(\widehat{\Theta})\,\mathrm{d}\eta_g(\mu)\right]\mathrm{d}\mu_0(\theta) \\
=\ &k\int_\Theta\int_{D_1}\mu(\widehat{\Theta})\,\mathrm{d}\eta_b(\mu)\,\mathrm{d}\beta + \int_\Theta\int_{D_0}\mu(\widehat{\Theta})\,\mathrm{d}\eta_g(\mu)\,\mathrm{d}[\mu_0 - k\beta] \\
=\ &k\int_\Theta\int_{D_1}\mu(\widehat{\Theta})\,\mathrm{d}\eta_b(\mu)\,\mathrm{d}\beta(\theta) + (1-k)\int_\Theta\int_{D_0}\mu(\widehat{\Theta})\,\mathrm{d}\eta_g(\mu)\,\mathrm{d}\gamma(\theta) \\
=\ &k\int_{D_1}\int_\Theta\mu(\widehat{\Theta})\,\mathrm{d}\mu(\theta)\,\mathrm{d}b + (1-k)\int_{D_0}\int_\Theta\mu(\widehat{\Theta})\,\mathrm{d}\mu(\theta)\,\mathrm{d}b(\mu) \\
=\ &k\int_{D_1}\mu(\widehat{\Theta})\,\mathrm{d}b(\mu) + (1-k)\int_{D_0}\mu(\widehat{\Theta})\,\mathrm{d}g(\mu) \\
=\ &k\int_{\widehat{\Theta}}\eta_b(D_1|\cdot)\,\mathrm{d}\beta + (1-k)\int_{\widehat{\Theta}}\eta_g(D_0|\cdot)\,\mathrm{d}\gamma \\
=\ &\int_{\widehat{\Theta}}\eta_b(D_1|\cdot)\,\mathrm{d}[k\beta] + \int_{\widehat{\Theta}}\eta_g(D_0|\cdot)\,\mathrm{d}[\mu_0 - k\beta] \\
=\ &\int_{\widehat{\Theta}} k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\eta_b(D_1|\cdot)\,\mathrm{d}\mu_0 + \int_{\widehat{\Theta}}\left(\mathbf{1} - k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\eta_g(D_0|\cdot)\,\mathrm{d}\mu_0 \\
=\ &\int_{\widehat{\Theta}}[(\mathbf{1}-\chi)\sigma + \chi\xi]\left(\widehat{M}\cap[\{0,1\}\times D]\ \middle|\ \cdot\right)\mathrm{d}\mu_0 \\
=\ &\int_{\widehat{\Theta}}[(\mathbf{1}-\chi)\sigma + \chi\xi](\widehat{M}|\cdot)\,\mathrm{d}\mu_0,
\end{aligned}$$

verifying the Bayesian property. So (ξ,σ,α,π) is a χ-equilibrium. Moreover, for any Borel $\widehat{D} \subseteq \Delta\Theta$, the equilibrium probability of the receiver posterior belief belonging to

$\widehat{D}$ is exactly (specializing the above algebra to $D_1 = D_0 = \widehat{D}$ and $\widehat{\Theta} = \Theta$)

$$\int_\Theta [(\mathbf{1}-\chi)\sigma + \chi\xi](\{0,1\}\times\widehat{D}|\cdot)\,\mathrm{d}\mu_0 = k\int_{\widehat{D}} \mathbf{1}\,\mathrm{d}b + (1-k)\int_{\widehat{D}} \mathbf{1}\,\mathrm{d}g = p(\widehat{D}).$$

Finally, the expected sender payoff conditional on reporting not being influenced—note the conditional distribution $\frac{\chi}{\chi(\mu_0)}\mu_0 \in \Delta\Theta$—is given by:

$$\begin{aligned}
&\int_\Theta\int_M u_S(\alpha(m))\,\mathrm{d}\xi(m|\cdot)\,\mathrm{d}\left[\tfrac{\chi}{\chi(\mu_0)}\mu_0\right] \\
&= \int_\Theta \left[\left(1-\tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\int_{\Delta\Theta} u_S\left(\alpha(0,\mu)\right)\mathrm{d}\eta_g(\mu|\cdot) + \tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\int_{\Delta\Theta} u_S\left(\alpha(1,\mu)\right)\mathrm{d}\eta_b(\mu|\cdot)\right]\mathrm{d}\left[\tfrac{\chi}{\chi(\mu_0)}\mu_0\right] \\
&= \int_\Theta \left[\left(\mathbf{1}-\tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\int_{\Delta\Theta} s_i\,\mathrm{d}\eta_g(\mu|\cdot) + \tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\int_{\mathrm{supp}\,(b)} \phi(\mu)\,\mathrm{d}\eta_b(\mu|\cdot)\right]\mathrm{d}\left[\tfrac{\chi}{\chi(\mu_0)}\mu_0\right] \\
&= s_i + \tfrac{k}{\chi(\mu_0)}\int_\Theta \left[-s_i + \int_{\mathrm{supp}\,(b)} \phi(\mu)\,\mathrm{d}\eta_b(\mu|\theta)\right]\mathrm{d}\beta(\theta) \\
&= \left[1-\tfrac{k}{\chi(\mu_0)}\right]s_i + \tfrac{k}{\chi(\mu_0)}\int_{\Delta\Theta}\int_\Theta \phi(\mu)\,\mathrm{d}\mu(\theta)\,\mathrm{d}b(\mu) \\
&= \tfrac{(1-k)-[1-\chi(\mu_0)]}{\chi(\mu_0)}s_i + \tfrac{k}{\chi(\mu_0)}\int_{\mathrm{supp}\,(b)} \phi\,\mathrm{d}b \\
&= s_o,
\end{aligned}$$

as required.

Conversely, suppose (ξ,σ,α,π) is a χ-equilibrium resulting in outcome (p,s_o,s_i). Let

$$\tilde{G} := \int_\Theta \sigma\,\mathrm{d}\mu_0 \text{ and } P := \int_\Theta [\chi\xi + (\mathbf{1}-\chi)\sigma]\,\mathrm{d}\mu_0 \in \Delta M$$

denote the probability measures over messages induced by non-committed behavior and by average sender behavior, respectively.

Let $M^* := \{m \in M :\ u_S(\alpha(m)) = s_i\}$ and $k := 1 - P(M^*)$. Sender incentive compatibility (which implies that $\sigma(M^*|\cdot) = \mathbf{1}$) tells us that $k \in [0,\chi(\mu_0)]$. Let $G := \frac{1}{1-k}P(\cdot\cap M^*)$ if $k<1$; and let $G := \tilde{G}$ otherwise. Let $B := \frac{1}{k}[P-(1-k)G]$ if $k>0$; and let $B := \int_\Theta \xi\,\mathrm{d}\mu_0$ otherwise. Both G and B are in ΔM because $(1-k)G \leqslant P$. Let $g := G\circ\pi^{-1}$ and $b := B\circ\pi^{-1}$,

both in $\Delta\Delta\Theta$. By construction, $kb + (1-k)g = P \circ \pi^{-1} = p \in \mathcal{R}(\mu_0)$. Moreover,

$$(1-k)\int_{\Delta\Theta} \mu \,\mathrm{d}g(\mu) = \int_M \pi \,\mathrm{d}[(1-k)G] = \int_{M^*} \pi \,\mathrm{d}P \geqslant (\mathbf{1} - \chi)\mu_0,$$

where the last inequality follows from the Bayesian property of π, together with the fact that σ almost surely sends a message from M^* on the path of play.

Next, for any $m \in M$ sender incentive compatibility tells us that $u_S(\alpha(m)) \leqslant s_i$, and receiver incentive compatibility tells us that $\alpha(m) \in V(\pi(m))$. If follows directly that $g\{V \ni s_i\} = b\{\min V \leqslant s_i\} = 1$.

Now viewing π, α as random variables on the probability space $\langle M, P\rangle$, define the conditional expectation $\phi_0 := \mathbb{E}_B[u_S(\alpha)|\pi] : M \to \mathbb{R}$. By Doob-Dynkin, there is a measurable function $\phi : \Delta\Theta \to \mathbb{R}$ such that $\phi \circ \pi =_{B-\text{a.e.}} \phi_0$. As $u_S(\alpha(m)) \in s_i \wedge V(m)$ for every $m \in M$, and the correspondence $s_i \wedge V$ is compact- and convex-valued, it must be that $\phi_0 \in_{B-\text{a.e.}} s_i \wedge V(\pi)$. Therefore, $\phi \in_{b-\text{a.e.}} s_i \wedge V$. Modifying ϕ on a b-null set, we may assume without loss that ϕ is a measurable selector of $s_i \wedge V$.

Observe now that $\tilde{G}(M^*) = G(M^*) = 1$ and

$$\int_{\text{supp}\,(b)} \phi \,\mathrm{d}b = \int_M \phi_0 \,\mathrm{d}B = \int_M \mathbb{E}_B[u_S(\alpha)|\pi] \,\mathrm{d}B = \int_M u_S \circ \alpha \,\mathrm{d}B.$$

Therefore,

$$\begin{aligned} s_o &= \int_M u_S \circ \pi \,\mathrm{d}\tfrac{P - [1-\chi(\mu_0)]\tilde{G}}{\chi} = \int_M u_S \circ \pi \,\mathrm{d}\tfrac{P - [1-\chi(\mu_0)]G}{\chi(\mu_0)} = \int_M u_S \circ \pi \,\mathrm{d}\tfrac{kB + (1-k)G - [1-\chi(\mu_0)]G}{\chi(\mu_0)} \\ &= \int_M u_S \circ \pi \,\mathrm{d}\left[\left(1 - \tfrac{k}{\chi(\mu_0)}\right)G + \tfrac{k}{\chi(\mu_0)}B\right] = \left(1 - \tfrac{k}{\chi(\mu_0)}\right)s_i + \tfrac{k}{\chi(\mu_0)}\int_{\text{supp}\,(b)} \phi \,\mathrm{d}b, \end{aligned}$$

as required. □

Proof of Theorem 2.1

Proof. By Lemma A.3.1, the supremum sender value over all χ-equilibrium outcomes is

$$\begin{aligned} v_\chi^*(\mu_0) := \sup_{b,g\in\Delta\Delta\Theta,\ k\in[0,1],\ s_o,s_i\in\mathbb{R}} & \left\{\chi(\mu_0)s_o + [1-\chi(\mu_0)]s_i\right\} \\ \text{s.t.} \quad & kb + (1-k)g \in \mathcal{R}(\mu_0),\ (1-k)\int_{\Delta\Theta}\mu\,\mathrm{d}g(\mu) \geqslant (\mathbf{1}-\chi)\mu_0, \\ & g\{V \ni s_i\} = b\{\min V \leqslant s_i\} = 1, \\ & s_o \in \left(1-\tfrac{k}{\chi(\mu_0)}\right)s_i + \tfrac{k}{\chi(\mu_0)}\int_{\mathrm{supp}\,(b)} s_i \wedge V\,\mathrm{d}b. \end{aligned}$$

Given any feasible (b, g, k, s_o, s_i) in the above program, replacing the associated measurable selector of $s_i \wedge V|_{\mathrm{supp}\,(b)}$ with the weakly higher function $s_i \wedge v|_{\mathrm{supp}\,(b)}$, and raising s_o to $\left(1-\frac{k}{\chi(\mu_0)}\right)s_i + \frac{k}{\chi(\mu_0)}\int_{\mathrm{supp}\,(b)} s_i \wedge v\,\mathrm{d}b$, will weakly raise the objectives and preserve all constraints. Therefore,

$$\begin{aligned} v_\chi^*(\mu_0) = \sup_{b,g\in\Delta\Delta\Theta,\ k\in[0,1],\ s_i\in\mathbb{R}} & \left\{\chi(\mu_0)\left[\left(1-\tfrac{k}{\chi(\mu_0)}\right)s_i + \tfrac{k}{\chi(\mu_0)}\int_{\mathrm{supp}\,(b)} s_i \wedge v\,\mathrm{d}b\right] + [1-\chi(1-\mu_0)]s_i\right\} \\ \text{s.t.} \quad & kb + (1-k)g \in \mathcal{R}(\mu_0),\ (1-k)\int_{\Delta\Theta}\mu\,\mathrm{d}g(\mu) \geqslant (\mathbf{1}-\chi)\mu_0, \\ & g\{V \ni s_i\} = b\{\min V \leqslant s_i\} = 1, \\ = \sup_{b,g\in\Delta\Delta\Theta,\ k\in[0,1],\ s_i\in\mathbb{R}} & \left\{(1-k)s_i + k\int_{\mathrm{supp}\,(b)} s_i \wedge v\,\mathrm{d}b\right\} \\ \text{s.t.} \quad & kb + (1-k)g \in \mathcal{R}(\mu_0),\ (1-k)\int_{\Delta\Theta}\mu\,\mathrm{d}g(\mu) \geqslant (\mathbf{1}-\chi)\mu_0, \\ & g\{V \ni s_i\} = b\{\min V \leqslant s_i\} = 1. \end{aligned}$$

Given any feasible (b, g, k, s_i) in the latter program, replacing (g, s_i) with any (g^*, s_i^*) such that $\int_{\Delta\Theta}\mu\,\mathrm{d}g^*(\mu) = \int_{\Delta\Theta}\mu\,\mathrm{d}g(\mu)$, $g^*\{V \ni s_i^*\} = 1$, and $s_i^* \geqslant s_i$ will preserve all constraints and weakly raise the objective. Moreover, Lipnowski and Ravid (2019, Lemma

1 and Theorem 2) tell us that any $\gamma \in \Delta\Theta$ has

$$\max_{g \in \mathcal{R}(\gamma), s_i \in \mathbb{R}:\ g\{V \ni s_i\}=1} s_i = \bar{v}(\gamma),$$

where $\bar{v}$ is the quasiconcave envelope of v.[7] Therefore,

$$\begin{aligned} v^*_\chi(\mu_0) = \sup_{b \in \Delta\Delta\Theta,\ \gamma \in \Delta\Theta,\ k \in [0,1]} & \left\{ (1-k)\,\bar{v}(\gamma) + k \int_{\Delta\Theta} \bar{v}(\gamma) \wedge v \,\mathrm{d}b \right\} \\ \text{s.t.} \qquad & k \int_{\Delta\Theta} \mu \,\mathrm{d}b(\mu) + (1-k)\gamma = \mu_0,\ (1-k)\gamma \geqslant (\mathbf{1} - \chi)\mu_0, \\ & b\{\min V \leqslant \bar{v}(\gamma)\} = 1. \end{aligned}$$

Claim: If $b \in \Delta\Delta\Theta$, $\gamma \in \Delta\Theta$, and $k \in [0,1]$ satisfy $k \int_{\Delta\Theta} \mu \,\mathrm{d}b(\mu) + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geqslant (\mathbf{1} - \chi)\mu_0$, then there exists (b^*, γ^*, k^*) feasible in the above program[8] such that $(1-k^*)\,\bar{v}(\gamma^*) + k^* \int_{\Delta\Theta} \bar{v}(\gamma^*) \wedge v \,\mathrm{d}b^* \geqslant (1-k)\,\bar{v}(\gamma) + k \int_{\Delta\Theta} \bar{v}(\gamma) \wedge v \,\mathrm{d}b$.

To prove the claim, let $\beta := \int_{\Delta\Theta} \mu \,\mathrm{d}b(\mu)$, and consider three exhaustive cases.

Case 1: $\bar{v}(\gamma) \leqslant v(\mu_0)$.

In this case, $(b^*, \gamma^*, k^*) := (\delta_{\mu_0}, \mu_0, 0)$ will work.

Case 2: $v(\mu_0) < \bar{v}(\gamma) \leqslant v(\beta)$.

In this case, Lipnowski and Ravid (2019, Lemma 3) delivers some $\beta^* \in \mathrm{co}\{\beta, \mu_0\}$ such that $V(\beta^*) \ni \bar{v}(\gamma)$. But then $\mu_0 \in \mathrm{co}\{\beta^*, \gamma\}$. As $\bar{v}$ is quasiconcave, $\bar{v}(\mu_0) \geqslant \min\{\bar{v}(\beta^*), \bar{v}(\gamma)\} \geqslant \min\{v(\beta^*), \bar{v}(\gamma)\} = \bar{v}(\gamma)$.

Therefore, $(b^*, \gamma^*, k^*) := (\delta_{\mu_0}, \mu_0, 0)$ will again work.

Case 3: $v(\beta) < \bar{v}(\gamma)$.

In this case, our aim is to show that there exists a $b^* \in \Delta\Delta\Theta$ such that:

- $b^* \in \mathcal{R}(\beta)$ and $b\{\min V \leqslant \bar{v}(\gamma)\} = 1$;
- $\int_{\Delta\Theta} \bar{v}(\gamma) \wedge v \,\mathrm{d}b^* \geqslant \int_{\Delta\Theta} \bar{v}(\gamma) \wedge v \,\mathrm{d}b$.

[7] Note that, $g\{V \ni s_i\} = 1$ implies $s_i \in \bigcap_{\mu \in \mathrm{supp}(g)} V(\mu)$ because V is upper hemicontinuous.
[8] That is, (b^*, γ^*, k^*) satisfy the same constraints, and further have $b^*\{\min V \leqslant \bar{v}(\gamma)\} = 1$.

Given such a measure, (b^*, γ, k) will be as required. We explicitly construct such a b^*.

Let $D := \operatorname{supp}(b)$, and define the measurable function,

$$\lambda : D \to [0,1]$$
$$\mu \mapsto \begin{cases} 1 & : \ v(\mu) \leqslant \bar{v}(\gamma) \\ \inf\left\{\widehat{\lambda} \in [0,1] : \ v\left((1-\widehat{\lambda})\gamma + \widehat{\lambda}\mu\right) \geqslant \bar{v}(\gamma)\right\} & : \text{otherwise.} \end{cases}$$

Lipnowski and Ravid (2019, Lemma 3) tells us that $\bar{v}(\gamma) \in V([1-\lambda(\mu)]\gamma + \lambda(\mu)\mu)$ for every $\mu \in D$ for which $v(\mu) > \bar{v}(\gamma)$. This implies that $\min V([1-\lambda(\mu)]\gamma + \lambda(\mu)\mu) \leqslant \bar{v}(\gamma)$ for *every* $\mu \in D$.

There must some number $\epsilon > 0$ such that $\lambda \geqslant \epsilon$ uniformly, because v is upper semi-continuous and $\bar{v}(\gamma) > v(\beta)$; and so $\frac{1}{\lambda} : D \to [1, \infty)$ is bounded. Moreover, by construction, $\lambda(\mu) < 1$ only for $\mu \in D$ with $v(\mu) > v(\gamma)$.

Now, define $b^* \in \Delta\Delta\Theta$ via

$$b^*(\widehat{D}) := \left(\int_{\Delta\Theta} \tfrac{1}{\lambda}\, \mathrm{d}b\right)^{-1} \cdot \int_{\Delta\Theta} \frac{1}{\lambda(\mu)} \mathbf{1}_{[1-\lambda(\mu)]\mu_0 + \lambda(\mu)\mu \in \widehat{D}}\, \mathrm{d}b(\mu), \ \forall \text{ Borel } \widehat{D} \subseteq \Delta\Theta.$$

Direct computation shows that $\int_{\Delta\Theta} \mu\, \mathrm{d}b^*(\mu) = \int_{\Delta\Theta} \mu\, \mathrm{d}b(\mu)$, i.e. $b^* \in \mathcal{R}(\beta)$. Moreover, by construction, $\min V([1-\lambda(\mu)]\gamma + \lambda(\mu)\mu) \leqslant \bar{v}(\gamma)\ \forall \mu \in D$. All that remains, then, is the

value comparison.

$$
\begin{aligned}
&\left(\int_{\Delta\Theta} \tfrac{1}{\lambda}\,\mathrm{d}b\right)\int_{\Delta\Theta} \bar{v}(\gamma)\wedge v\,\mathrm{d}[b^*-b]\\
&=\int_{\Delta\Theta}\left[\tfrac{1}{\lambda(\mu)}\,\bar{v}(\gamma)\wedge v\Big([1-\lambda(\mu)]\mu_0+\lambda(\mu)\mu\Big)-\left(\int_{\Delta\Theta}\tfrac{1}{\lambda}\,\mathrm{d}b\right)\bar{v}(\gamma)\wedge v(\mu)\right]\mathrm{d}b(\mu)\\
&=\int_{\Delta\Theta}\left(\tfrac{1}{\lambda(\mu)}-\int_{\Delta\Theta}\tfrac{1}{\lambda}\,\mathrm{d}b\right)\left[v(\mu)\mathbf{1}_{v(\mu)\leqslant\bar{v}(\gamma)}+\bar{v}(\gamma)\mathbf{1}_{v(\mu)>\bar{v}(\gamma)}\right]\mathrm{d}b(\mu)\\
&=\int_{\Delta\Theta}\left(\tfrac{1}{\lambda(\mu)}-\int_{\Delta\Theta}\tfrac{1}{\lambda}\,\mathrm{d}b\right)\left\{\bar{v}(\gamma)-[\bar{v}(\gamma)-v(\mu)]\mathbf{1}_{v(\mu)\leqslant\bar{v}(\gamma)}\right\}\mathrm{d}b(\mu)\\
&=0+\int_{\Delta\Theta}\left(\int_{\Delta\Theta}\tfrac{1}{\lambda}\,\mathrm{d}b-\tfrac{1}{\lambda(\mu)}\right)[\bar{v}(\gamma)-v(\mu)]\mathbf{1}_{v(\mu)\leqslant\bar{v}(\gamma)}\,\mathrm{d}b(\mu)\\
&=\int_{\{\mu\in\Delta\Theta:\ v(\mu)\leqslant\bar{v}(\gamma)\}}\left(\int_{\Delta\Theta}\tfrac{1}{\lambda}\,\mathrm{d}b-1\right)[\bar{v}(\gamma)-v(\mu)]\,\mathrm{d}b(\mu)\\
&=\left(\int_{\Delta\Theta}\tfrac{1-\lambda}{\lambda}\,\mathrm{d}b\right)\int_{\{\mu\in\Delta\Theta:\ v(\mu)\leqslant\bar{v}(\gamma)\}}[\bar{v}(\gamma)-v]\,\mathrm{d}b\\
&\geqslant 0,
\end{aligned}
$$

proving the claim.

In light of the claim, the optimal value is

$$
\begin{aligned}
v^*_\chi(\mu_0)&=\sup_{b\in\Delta\Delta\Theta,\ \gamma\in\Delta\Theta,\ k\in[0,1]}\left\{(1-k)\,\bar{v}(\gamma)+k\int_{\Delta\Theta}\bar{v}(\gamma)\wedge v\,\mathrm{d}b\right\}\\
&\qquad\text{s.t.}\qquad k\int_{\Delta\Theta}\mu\,\mathrm{d}b(\mu)+(1-k)\gamma=\mu_0,\ (1-k)\gamma\geqslant(\mathbf{1}-\chi)\mu_0,\\
&=\sup_{\beta,\gamma\in\Delta\Theta,\ k\in[0,1]}\left\{(1-k)\,\bar{v}(\gamma)+k\sup_{b\in\mathcal{R}(\beta)}\int_{\Delta\Theta}\bar{v}(\gamma)\wedge v\,\mathrm{d}b\right\}\\
&\qquad\text{s.t.}\qquad k\beta+(1-k)\gamma=\mu_0,\ (1-k)\gamma\geqslant(\mathbf{1}-\chi)\mu_0,\\
&=\sup_{\beta,\gamma\in\Delta\Theta,\ k\in[0,1]}\left\{(1-k)\,\bar{v}(\gamma)+k\widehat{v}_{\wedge\gamma}(\beta)\right\}\\
&\qquad\text{s.t.}\qquad k\beta+(1-k)\gamma=\mu_0,\ (1-k)\gamma\geqslant(\mathbf{1}-\chi)\mu_0.
\end{aligned}
$$

Finally, observe that the supremum is in fact a maximum because the constraint set is a compact subset of $(\Delta\Theta)^2\times[0,1]$ and the objective upper semicontinuous. □

Consequences of Lemma A.3.1 and Theorem 2.1

Corollary A.3.1. *As x ranges over $[0,1]$, the set of $x\mathbf{1}$-equilibrium outcomes (p, s_o, s_i) at prior μ_0 is a compact-valued, upper hemicontinuous correspondence of (μ_0, x).*

Proof. Let Y_G be the graph of V and Y_B be the graph of $[\min V, \max u_S(A)]$, both compact because V is a Kakutani correspondence.

Let X be the set of all $(\mu_0, p, g, b, x, k, s_o, s_i) \in (\Delta\Theta) \times (\Delta\Delta\Theta)^3 \times [0,1]^2 \times [\text{co } u_S(A)]^2$ such that:

- $kb + (1-k)g = p$;
- $(1-x)\int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu) + x\int_{\Delta\Theta} \mu \, \mathrm{d}b(\mu) = \mu_0$;
- $(1-k)\int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu) \geqslant (1-x)\mu_0$;
- $g \otimes \delta_{s_i} \in \Delta(Y_G)$ and $b \otimes \delta_{s_i} \in \Delta(Y_B)$;
- $k\int_{\Delta\Theta} \min V \, \mathrm{d}b \leqslant (k-x)\, s_i + x s_o \leqslant k\int_{\Delta\Theta} s_i \wedge v \, \mathrm{d}b$.

As an intersection of compact sets, X is itself compact. By Lemma A.3.1, the equilibrium outcome correspondence has a graph which is a projection of X, and so is itself compact. Therefore, it is compact-valued and upper hemicontinuous. □

Corollary A.3.2. *For any $\mu_0 \in \Delta\Theta$, the map*

$$\begin{aligned} [0,1] &\to \mathbb{R} \\ x &\mapsto v^*_{x\mathbf{1}}(\mu_0) \end{aligned}$$

is weakly increasing and right-continuous.

Proof. That it is weakly increasing is immediate from Theorem 2.1, given that increasing credibility expands the constraint set. That it is upper semicontinuous (and so, since nondecreasing, it is right-continuous) follows directly from Corollary A.3.1. □

Corollary A.3.3. *For any $x \in [0,1]$, the map $v^*_{x\mathbf{1}} : \Delta\Theta \to \mathbb{R}$ is upper semicontinuous.*

Proof. This is immediate from Corollary A.3.1. □

A.3.2 Productive Mistrust: Proofs

Toward verifying our sufficient conditions for productive mistrust to occur, we first study in some depth the possibility of productive mistrust in the binary-state world. We then leverage that analysis to study the same in many-state environments.

To this end, it useful to introduce a more detailed language for our key SOB condition. Given a prior $\mu \in \Delta\Theta$, say S is **an SOB at** μ if every $p \in \mathcal{R}(\mu)$ is outperformed by an SOB policy $p' \in \mathcal{R}(\mu)$.

Productive Mistrust with Binary States

Given binary states, finitely many actions, and a full-support prior μ_0, we know that the quasiconcave envelope function $\bar{v} : \Delta\Theta \to \mathbb{R}$ is upper semicontinuous, weakly quasiconcave, and piecewise constant. Therefore, if $\mu_0 \notin \operatorname{argmax}_{\mu\in\Delta\Theta} \bar{v}(\mu)$, there is then a unique $\mu_+ = \mu_+(\mu_0)$ closest to μ_0 with the property that $\bar{v}(\mu_+) > \bar{v}(\mu_0)$, and a unique $\theta = \theta(\mu_0) \in \Theta$ with $\mu_0 \in \operatorname{co}\{\mu_+(\mu_0), \delta_\theta\}$. In this case, for the rest of the subsection, we identify $\Delta\Theta \approxeq [0,1]$ by identifying $\nu \in \Delta\Theta$ with $1 - \nu(\theta(\mu_0))$.[9]

Lemma A.3.2. *Given finite A, binary Θ, and a full-support prior $\mu_0 \in \Delta\Theta$, the following are equivalent:*

1. *There exist credibility levels $\chi' < \chi$ such that, for every S-optimal χ-equilibrium outcome (p,s) and S-optimal χ'-equilibrium outcome (p',s'), the policy p' is strictly more Blackwell-informative than p.*

2. *$\mu_0 \notin \operatorname{argmax}_{\mu\in\Delta\Theta:\ \mu \text{ full-support}} \bar{v}(\mu)$, and there exists $\mu_- \in [0,\mu_0]$ such that $v(\mu_-) > v(0) + \frac{\mu_-}{\mu_+}\big[v(\mu_+) - v(0)\big]$.*

[9] So, under this normalization, $0 = \theta < \mu_0 < \mu_+$.

Moreover, in this case, every S-optimal χ'-equilibrium outcome gives the receiver a strictly higher payoff than any S-optimal χ-equilibrium.

Proof. First, suppose (2) fails. There are three ways it could fail:

(a) With $\mu_0 \in \operatorname{argmax}_{\mu\in\Delta\Theta} \bar{v}(\mu)$;

(b) With $\mu_0 \in \operatorname{argmax}_{\mu\in\Delta\Theta:\ \mu \text{ full-support}} \bar{v}(\mu) \setminus \operatorname{argmax}_{\mu\in\Delta\Theta} \bar{v}(\mu)$;

(c) With $\mu_0 \notin \operatorname{argmax}_{\mu\in\Delta\Theta:\ \mu \text{ full-support}} \bar{v}(\mu)$;

In case (a) or (b), pick some S-optimal 0-equilibrium information policy p_0. For any $\widehat{x} \in [0,1)$, we know $(p_0, \bar{v}(\mu_0))$ is a S-optimal 0-equilibrium outcome; and in case (a) it is also a S-optimal $\mathbf{1}$-equilibrium outcome.

For case (a), there is nothing left to show.

For case (b), we need only consider the case of $\chi = \mathbf{1}$. In case (b), that $\bar{v}$ is weakly quasiconcave implies it is monotonic. So $\mu_+ = 1$, and $\bar{v} : [0,1] \to \mathbb{R}$ is nondecreasing with $\bar{v}|_{[\mu_0,1)} = \bar{v}(\mu_0) < \bar{v}(1)$. As $v^*_{\mathbf{1}}$ is the concave envelope of $\bar{v}$, it must be that the support of any S-optimal $\mathbf{1}$-equilibrium information policy is contained in $\left[0,\ \min\{\mu \in [0,1] :\ v(\mu) = v(\mu_0))\}\right] \cup \{1\}$, so that (1) fails as well.

In case (c), failure of (2) tells us $v(\mu) \leqslant v(0) + \frac{\mu}{\mu_+}\left[v(\mu_+) - v(0)\right],\ \forall \mu \in [0,\mu_0]$. As $\bar{v}|_{[0,\mu_+)} \leqslant \bar{v}(\mu_0)$, it follows that

$$
\begin{aligned}
v^*_{\widehat{x}\mathbf{1}}(\mu_0) = &\max_{\beta,\gamma,k\in[0,1]} \left\{k\widehat{v}_{\wedge\gamma}(\beta) + (1-k)\,\bar{v}(\gamma)\right\} \\
&\qquad \text{s.t.} \qquad k\beta + (1-k)\gamma = \mu_0,\ (1-k)(\gamma, 1-\gamma) \geqslant (1-\widehat{x})(\mu_0, 1-\mu_0) \\
= &\max_{\gamma\in[\mu_0,1],k\in[0,1]} \left\{k v(0) + (1-k)\,\bar{v}(\gamma)\right\} \\
&\qquad \text{s.t.} \qquad k0 + (1-k)\gamma = \mu_0,\ (1-k)(\gamma, 1-\gamma) \geqslant (1-\widehat{x})(\mu_0, 1-\mu_0) \\
= &\max_{\gamma\in[\mu_0,1]} \left\{\left(1 - \tfrac{\mu_0}{\gamma}\right) v(0) + \tfrac{\mu_0}{\gamma}\,\bar{v}(\gamma)\right\} \\
&\qquad \text{s.t.} \qquad \tfrac{\mu_0}{\gamma}(1-\gamma) \geqslant (1-\widehat{x})(1-\mu_0).
\end{aligned}
$$

In particular, defining $\gamma(\widehat{x})$ to be the largest argmax in the above optimization problem, it follows that

$$p_{\widehat{x}} = \left(1 - \tfrac{\mu_0}{\gamma(\widehat{x})}\right)\delta_0 + \tfrac{\mu_0}{\gamma}\delta_{\gamma(\widehat{x})}$$

is a S-optimal $\widehat{x}\mathbf{1}$-equilibrium information policy for any $\widehat{x} \in [0,1]$, so that (1) does not hold.

Conversely, suppose (2) holds.

The function $v : [0,1] \to \mathbb{R}$ is upper semicontinuous and piecewise constant, which implies that its concave envelope $v_{\mathbf{1}}^*$ is piecewise affine. We may then define

$$\mu_-^* := \min\{\mu \in [0,\mu_0] : \ v_{\mathbf{1}}^* \text{ is affine over } [\mu,\mu_0]\}.$$

That (2) holds tells us that $\mu_-^* \in (0,\mu_0)$. It is then without loss to take $\mu_- = \mu_-^*$.

There are thus beliefs $\mu_-, \mu_+ \in [0,1]$ such that: $0 < \mu_- < \mu_0 < \mu_+$; $v_{\mathbf{1}}^*$ is affine on $[\mu_-,\mu_+]$ and on no larger interval; and $v_{\mathbf{1}}^*$ is strictly increasing on $[0,\mu_+]$. It follows that $\widehat{v}_{\wedge\mu_+} = v_{\mathbf{1}}^*$ on $[0,\mu_+]$. By definition of $\mu_+ = \mu_+(\mu_0)$, we know that $\bar{v}$ is constant on $[\mu_0,\mu_+)$. That is, (appealing to Lipnowski and Ravid (2019, Theorem 2)) $v_{\mathbf{0}}^*$ is constant on $[\mu_0,\mu_+)$. Then, since $v_{\mathbf{1}}^*$ strictly decreases there, it must be that $v_{\mathbf{1}}^* > v_{\mathbf{0}}^*$ on (μ_0,μ_+).

Let $x \in [0,1]$ be the smallest credibility level such that $v_{x\mathbf{1}}^*(\mu_0) = v_{\mathbf{1}}^*(\mu_0)$, which exists by Corollary A.3.2. That $v_{\mathbf{0}}^*(\mu_0) < v_{\mathbf{1}}^*(\mu_0)$ implies $\chi > 0$. That μ_+ has full support, which follows from (2), implies that $x < 1$.[10]

Consider now the following claim.

Claim: *Given $x' \in [0,x]$, suppose that*

$$(\beta',\gamma',k') \in \operatorname{argmax}_{(\beta,\gamma,k)\in[0,1]^3}\left\{k\widehat{v}_{\wedge\gamma}(\beta) + (1-k)\,\bar{v}(\gamma)\right\}$$
$$s.t. \qquad k\beta + (1-k)\gamma = \mu_0,\ (1-k)(\gamma,1-\gamma) \geqslant (1-x')(\mu_0,1-\mu_0),$$

[10]In particular, this follows from the hypothesis that there exists some *full-support* belief at which $\bar{v}$ takes a strictly higher value than $v(\mu_0)$. This implies $x < 1$ by the same argument employed to prove Proposition 2.3.

for a value strictly higher than $\bar{v}(\mu_0)$. Then:

- $\gamma' = \mu_+$ *and* $\beta' \leqslant \mu_-$.
- *If* $h' \in \mathcal{R}(\beta')$ *and* $\ell' \in \mathcal{R}(\gamma')$ *are such that* $p' = k'h' + (1-k')\ell'$ *is the information policy of a S-optimal* $x'\mathbf{1}$*-equilibrium, then* $h'[0,\mu_-] = \ell'\{\mu_+\} = 1$.

We now prove the claim.

If $\gamma' > \mu_+$, then let $k'' \in (0,k')$ be the unique solution to $k''\beta' + (1-k'')\mu_+ = \mu_0$. As $(1-k'')(\mu_+, 1-\mu_+) \geqslant (1-x')(\mu_0, 1-\mu_0)$ and

$$k''\widehat{v}_{\wedge\mu_+}(\beta') + (1-k'')\bar{v}(\mu_+) \geqslant k''\widehat{v}_{\wedge\gamma'}(\beta') + (1-k'')\bar{v}(\gamma') > k'\widehat{v}_{\wedge\gamma'}(\beta') + (1-k')\bar{v}(\gamma'),$$

the feasible solution (β', μ_+, k'') would strictly outperform (β', γ', k'). So optimality implies $\gamma' \leqslant \mu_+$.

Notice that $\bar{v}$—as a weakly quasiconcave function which is nondecreasing and nonconstant over $[\mu_0, \mu_+]$—is nondecreasing over $[0, \mu_+]$. Moreover, $\lim_{\mu \nearrow \mu_+} \bar{v}(\mu) = \bar{v}(\mu_0) < \bar{v}(\mu_+)$. Therefore, if $\gamma' < \mu_+$, it would follow that $k'\widehat{v}_{\wedge\gamma'}(\beta') + (1-k')\bar{v}(\gamma') \leqslant \bar{v}(\gamma') \leqslant \bar{v}(\mu_0)$. Given the hypothesis that (β', γ', k') strictly outperforms $\bar{v}(\mu_0)$, it follows that $\gamma' = \mu_+$. One direct implication is that

$$\begin{aligned}(\beta', k') \in \operatorname{argmax}_{(\beta,k)\in[0,1]^2} &\left\{k\widehat{v}_{\wedge\mu_+}(\beta) + (1-k)\max v[0,\mu_+]\right\} \\ \text{s.t.} \qquad & k\beta + (1-k)\mu_+ = \mu_0,\ (1-k)(1-\mu_+) \geqslant (1-x')(1-\mu_0).\end{aligned}$$

Let us now see why we cannot have $\beta' \in (\mu_-, \mu_0)$. As $\widehat{v}_{\wedge\mu_+}$ is affine on $[\mu_+, \mu_-]$, replacing such (k', β') with (k, μ_-) which satisfies $k\mu_- + (1-k)\mu_+ = \mu_0$ necessarily has $(1-k)(\mu_+, 1-\mu_+) \gg (1-x')(\mu_0, 1-\mu_0)$. This would contradict minimality of x. Therefore, $\beta' \leqslant \mu_-$.

We now prove the second bullet. First, every $\mu < \mu_+$ satisfies $v(\mu) \leqslant v^*_{\mathbf{1}}(\mu) < v^*_{\mathbf{1}}(\mu_+) = v(\mu_+)$. This implies that δ_{μ_+} is the unique $\ell \in \mathcal{R}(\mu_+)$ with $\inf v(\operatorname{supp} \ell) \geqslant v(\mu_+)$. Therefore, $\ell' = \delta_{\mu_+}$.

Second, the measure $h' \in \mathcal{R}(\beta')$ can be expressed as $h' = (1-\gamma)h_L + \gamma h_R$ for $h_L \in \Delta[0,\mu_-]$, $h_R \in \Delta(\mu_-, 1]$, and $\gamma \in [0,1)$. Notice that $(\mu_-, v(\mu_-))$ is an extreme point of the subgraph of $v^*_{\mathbf{1}}$, and therefore an extreme point of the subgraph of $\widehat{v}_{\wedge\mu_+}$. Taking the unique $\widehat{\gamma} \in [0,\gamma]$ such that $\widehat{h} := (1-\widehat{\gamma})h_L + \widehat{\gamma}\delta_{\mu_-} \in \mathcal{R}(\beta')$, it follows that $\int_{[0,1]} \widehat{v}_{\wedge\mu_+} \, \mathrm{d}\widehat{h} \geqslant \int_{[0,1]} \widehat{v}_{\wedge\mu_+} \, \mathrm{d}h'$, strictly so if $\widehat{\gamma} < \gamma$. But $\widehat{\gamma} < \gamma$ necessarily if $\gamma > 0$, since $\int_{[0,1]} \mu \, \mathrm{d}h_R(\mu) > \mu_-$. Optimality of h' then implies that $\gamma = 0$, i.e. $h'[0,\mu_-] = 1$. This completes the proof of the claim.

With the claim in hand, we can now prove the proposition. Letting $k^* \in (0,1)$ be the solution to $k^*\mu_- + (1-k^*)\mu_+ = \mu_0$, the claim implies that (μ_-, μ_+, k^*) is the unique solution to

$$\begin{aligned} &\max_{(\beta,\gamma,k)\in[0,1]^3} \left\{ k\widehat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) \right\} \\ &\quad \text{s.t.} \quad k\beta + (1-k)\gamma = \mu_0, \ (1-k)(\gamma, 1-\gamma) \geqslant (1-x)(\mu_0, 1-\mu_0), \end{aligned}$$

and that $p^* = k^*\delta_{\mu_-} + (1-k^*)\delta_{\mu_+}$ is the uniquely S-optimal $x\mathbf{1}$-equilibrium information policy. Moreover, the minimality property defining x implies that $(1-k^*)(1-\mu_+) = (1-x)(1-\mu_0)$.

Given $x' < x$ sufficiently close to x, one can verify directly that (β', μ_+, k') is feasible, where

$$k' := 1 - \tfrac{1-x'}{1-x}(1-k^*) \text{ and } \beta' := \tfrac{1}{k'}\left[\mu_0 - (1-k')\mu_+\right].$$

As $\widehat{v}_{\wedge\mu_+}$ is a continuous function, it follows that $v^*_{x'\mathbf{1}}(\mu_0) \nearrow v^*_{x\mathbf{1}}(\mu_0)$ as $x' \nearrow x$. In particular, $v^*_{x'\mathbf{1}}(\mu_0) > v^*_0(\mu_0)$ for $x' < x$ sufficiently close to x. Fix such a x'.

Let p' be any S-optimal $x'\mathbf{1}$-equilibrium information policy. Appealing to the claim, it must be that there exists some $h' \in \mathcal{R}(\beta') \cap \Delta[0, \mu_-]$ such that $p' \in \text{co}\{h', \delta_{\mu_+}\}$. Therefore, p' is weakly more Blackwell-informative than p^*. Finally, as $(1-k^*)(1-\mu_+) = (1-x)(1-\mu_0)$ and $x' < x$, feasibility of p' tells us that $p' \neq p^*$. Therefore (the Blackwell order being antisymmetric), p' is strictly more informative than p^*, proving (1).

Having shown that (2) implies (1), all that remains is to show that the receiver's optimal payoff is strictly higher given p' than given p^*. To that end, fix sender-preferred receiver best responses a_- and a_+ to μ_- and μ_+, respectively. As the receiver's optimal value given p^* is attainable using only actions $\{a_-, a_+\}$, and the same value is feasible given only information p' and using only actions $\{a_-, a_+\}$, it suffices to show that there are beliefs in the support of p' to which neither of $\{a_-, a_+\}$ is a receiver best response. But, at every $\mu \in [0, \mu_-)$ satisfies

$$v(\mu) \leqslant \bar{v}(\mu) < \bar{v}(\mu_-) = \min\{\bar{v}(\mu_-), \bar{v}(\mu+)\};$$

that is, $\max u_S\left(\text{argmax}_{a \in A} u_R(a, \mu)\right) < \min\{u_S(a_-), u_S(a_+)\}$. The result follows. □

The following Lemma is the specialization of Proposition 2.1 to the binary-state world. In addition to being a special case of the proposition, it will also be an important lemma for proving the more general result.

Lemma A.3.3. *Suppose $|\Theta| = 2$, the model is finite and generic, a full-support belief $\mu \in \Delta\Theta$ exists such that the sender is not an SOB at μ. Then there exists a full-support prior μ_0 and credibility levels $\chi' < \chi$ such that every S-optimal χ'-equilibrium is both strictly better for R and more Blackwell-informative than every S-optimal χ-equilibrium.*

Proof. First, notice that the genericity assumption delivers full-support μ', such that $V(\mu') = \{\max v(\Delta\Theta)\}$.

Name our binary-state space $\{0, 1\}$ and identify $\Delta\Theta = [0, 1]$ in the obvious way. The function $v : [0, 1] \to \mathbb{R}$ is piecewise constant, which implies that its concave envelope $v^*_{\mathbf{1}}$

is piecewise affine. That is, there exist $n \in \mathbb{N}$ and $\{\mu^i\}_{i=0}^n$ such that $0 = \mu^0 \leqslant \cdots \leqslant \mu^n = 1$ and $v_{\mathbf{1}}^*|_{[\mu^{i-1},\mu^i]}$ is affine for every $i \in \{1,\ldots,n\}$. Taking n to be minimal, we can assume that $\mu^0 < \cdots < \mu^n$ and the slope of $v_{\mathbf{1}}^*|_{[\mu^{i-1},\mu^i]}$ is strictly decreasing in i. Therefore, there exist $i_0, i_1 \in \{0,\ldots,n\}$ such that $i_1 \in \{i_0, i_0+1\}$ and $\operatorname{argmax}_{\mu\in[0,1]} v(\mu) = [\mu^{i_0}, \mu^{i_1}]$. That the sender is not an SOB at μ implies that $i_0 > 1$ or $i_1 < n-1$. Without loss of generality, say $i_0 > 1$. Now let $\mu_- := \mu^{i_0-1}$ and $\mu_+ := \mu^{i_0}$.

Finally, that $V(\mu') = \{\max v(\Delta\Theta)\}$, and V is (by Berge's theorem) upper hemicontinuous implies $\operatorname{argmax}_{\mu\in\Delta\Theta:\ \mu \text{ full-support}} \bar{v}(\mu) = \operatorname{argmax}_{\mu\in\Delta\Theta} \bar{v}(\mu)$. Therefore, considering any prior of the form $\mu_0 = \mu_+ - \epsilon$ for sufficiently small $\epsilon > 0$, Lemma A.3.2 applies. □

Productive Mistrust with Many States: Proof of Proposition 2.1

Given Lemma A.3.3, we need only prove the proposition for the case of $|\Theta| > 2$, which we do below. The proof intuition is as follows. Using the binary-state logic, one can always obtain a binary-support prior μ_0^∞ and constant credibility levels $\chi' < \chi$ such that R strictly prefers every S-optimal χ'-equilibrium to every S-optimal χ-equilibrium. We then find an interior direction through which to approach μ_0^∞, while keeping S's optimal equilibrium value under both credibility levels continuous. Genericity ensures that such a direction exists despite $\bar{v}$ being discontinuous. The continuity in S's value from the identified direction then ensures upper hemicontinuity of S's optimal equilibrium policy set; that is, the limit of every sequence of S-optimal equilibrium policies from said direction must also be optimal under μ_0^∞. Now, if the proposition were false, one could construct a convergent sequence of S-optimal equilibrium policies from said direction for each credibility level, $\{p_n^\chi, p_n^{\chi'}\}_{n\geqslant 0}$, such that R would weakly prefer p_n^χ to $p_n^{\chi'}$. As R's payoffs are continuous, R being weakly better off under χ than under χ' along the sequences would imply the same at the sequences' limits. Notice, though, that such limits must be S-optimal for the prior μ_0^∞ by the choice of direction, meaning that

productive mistrust fails at μ_0^∞; that is, we have a contradiction. Below, we proceed with the formal proof.

Proof. Let $\Theta_2 := \{\theta_1, \theta_2\}$ and $u := \max v(\Delta\Theta_2)$, and define the receiver value function $v_R : \Delta\Delta\Theta \to \mathbb{R}$ via $v_R(p) := \int_{\Delta\Theta} \max_{a \in A} u_R(a, \mu)\, \mathrm{d}p(\mu)$.

Appealing to Lemma A.3.3, there is some $\mu_0^\infty \in \Delta\Theta$ with support Θ_2 and credibility levels $\chi'' < \chi'$ such that every S-optimal χ''-equilibrium is strictly better for R than every S-optimal χ'-equilibrium.

Consider the following claim.

Claim: There exists a sequence $\{\mu_0^n\}$ of full-support priors converging to μ_0^∞ such that

$$\liminf_{n \to \infty} v_\chi^*(\mu_0^n) \geqslant v_\chi^*(\mu_0^\infty) \text{ for } \chi \in \{\chi', \chi''\}.$$

Before proving the claim, let us argue that it implies the proposition. Given the claim, assume for contradiction that: for every $n \in \mathbb{N}$, prior μ_0^n admits some S-optimal χ'-equilibrium and χ''-equilibrium, $\Psi'_n = (p'_n, s'_{i\,n}, s'_{o\,n})$ and $\Psi''_n = (p''_n, s''_{i\,n}, s''_{o\,n})$, respectively, such that $v_R(p'_n) \geqslant v_R(p''_n)$. Dropping to a subsequence if necessary, we may assume by compactness that $(\Psi'_n)_n$ and $(\Psi''_n)_n$ converge (in $\Delta\Delta\Theta \times \mathbb{R} \times \mathbb{R}$) to some $\Psi' = (p', s'_i, s'_o)$ and $\Psi'' = (p'', s''_i, s''_o)$ respectively. By Corollary A.3.1, for every credibility level χ, the set of χ-equilibria is an upper hemicontinuous correspondence of the prior. Therefore, Ψ' and Ψ'' are χ'- and χ''-equilibria, respectively, at prior μ_0^∞. Continuity of v_R (by Berge's theorem) then implies that $v_R(p') \geqslant v_R(p'')$. Finally, by the claim, it must be that Ψ' and Ψ'' are S-optimal χ'- and χ''-equilibria, respectively, contradicting the definition of μ_0^∞. Therefore, there is some $n \in \mathbb{N}$ for which the full-support prior μ_0^n is as required for the proposition.

So all that remains is to prove the claim. To do this, we construct the desired sequence.

First, the proof of Lemma A.3.3 delivers some $\gamma^\infty \in \Delta\Theta$ such that $\bar{v}(\gamma^\infty) = u$ and, for both $\chi \in \{\chi', \chi''\}$, some $(\beta, \gamma, k) \in \Delta\Theta \times \{\gamma^\infty\} \times [0,1]$ solves the program in Theorem 2.1 at prior μ_0^∞.

Let us now show that there exists a closed convex set $D \subseteq \Delta\Theta$ which contains γ^∞, has nonempty interior, and satisfies $\bar{v}|_D = u$. Indeed, for any $n \in \mathbb{N}$, let $B_n \subseteq \Delta\Theta$ be the closed ball (say with respect to the Euclidean metric) of radius $\frac{1}{n}$ around μ', and let $D_n := \text{co}\left[\{\gamma^\infty\} \cup B_n\right]$. As $v|_{\Delta\Theta_2} \leqslant u$ and constant functions are quasiconcave, Lipnowski and Ravid (2019, Theorem 2) tells us $\bar{v}|_{\Delta\Theta_2} \leqslant u$ as well. As V is upper hemicontinuous, the hypothesis on μ' ensures that $\bar{v}|_{B_n} \geqslant v|_{B_n} = u$ for sufficiently large $n \in \mathbb{N}$; quasiconcavity then tells us $\bar{v}|_{D_n} \geqslant u$. Assume now, for a contradiction, that every $n \in \mathbb{N}$ has $\bar{v}|_{D_n} \not\leqslant u$. That is, there is some $\lambda_n \in [0,1]$ and $\mu'_n \in B_n$ such that $\bar{v}\left((1-\lambda_n)\mu + \lambda_n \mu'_n\right) > u$. Dropping to a subsequence, we get a strictly increasing sequence $(n_\ell)_\ell^\infty$ of natural numbers such that (since $[0,1]$ is compact and $\bar{v}(\Delta\Theta)$ is finite) $\lambda_{n_\ell} \xrightarrow{\ell\to\infty} \lambda \in [0,1]$ and $\bar{v}\left((1-\lambda_{n_\ell})\mu + \lambda_{n_k}\mu'_{n_k}\right) = \hat{u}$ for some number $\hat{u} \in (u, \infty)$ and every $\ell \in \mathbb{N}$. As $\bar{v}$ is upper semicontinuous, this would imply that $\bar{v}\left((1-\lambda)\mu + \lambda\mu'\right) \geqslant \hat{u} > u$, contradicting the definition of u. Therefore, some $D \in \{D_{n_\ell}\}_{\ell=1}^\infty$ is as desired. In what follows, let $\gamma_1 \in D$ be some interior element with full support.

Now, for each $n \in \mathbb{N}$, define $\mu_0^n := \frac{n-1}{n}\mu_0^\infty + \frac{1}{n}\gamma_1$. We will show that the sequence $(\mu_0^n)_{n=1}^\infty$—a sequence of full-support priors converging to μ_0^∞—is as desired. To that end, fix $\chi \in \{\chi', \chi''\}$ and some $(\beta, k) \in \Delta\Theta \times [0,1]$ such that $(\beta, \gamma^\infty, k)$ solves the program in Theorem 2.1 at prior μ_0^∞. Then, for any $n \in \mathbb{N}$, let:

$$\begin{aligned}
\epsilon_n &:= \tfrac{1}{n-(n-1)k} \in (0,1], \\
\gamma_n &:= (1-\epsilon_n)\gamma^\infty + \epsilon_n\gamma_1 \in D, \\
k_n &:= \tfrac{n-1}{n}k \in [0,k).
\end{aligned}$$

Given these definitions,

$$\begin{aligned}
(1-k_n)\gamma_n &= \tfrac{1}{n}\left[n-(n-1)k\right]\gamma_n \\
&= \tfrac{1}{n}\left\{\left[n-(n-1)k-1\right]\gamma^\infty+\gamma_1\right\} \\
&= \tfrac{n-1}{n}(1-k)\gamma^\infty+\tfrac{1}{n}\gamma_1 \\
&\geqslant \tfrac{n-1}{n}(\mathbf{1}-\chi)\mu_0^\infty+\tfrac{1}{n}\gamma_1 \geqslant (\mathbf{1}-\chi)\mu_0^n, \text{ and} \\
k_n\beta+(1-k_n)\gamma_n &= \tfrac{n-1}{n}k\beta+\tfrac{n-1}{n}(1-k)\gamma^\infty+\tfrac{1}{n}\gamma_1 \\
&= \tfrac{n-1}{n}\mu_0^\infty+\tfrac{1}{n}\gamma_1=\mu_0^n.
\end{aligned}$$

Therefore, (β,γ_n,k_n) is χ-feasible at prior μ_0^n. As a result,

$$\begin{aligned}
v_\chi^*(\mu_0^n) &\geqslant k_n\widehat{v}_{\wedge\gamma_n}(\beta)+(1-k_n)\bar{v}(\gamma_n) \\
&= k_n\widehat{v}_{\wedge\gamma}(\beta)+(1-k_n)\bar{v}(\gamma) \text{ (since } \bar{v}(\gamma_n)=u) \\
&\xrightarrow{n\to\infty} k\widehat{v}_{\wedge\gamma}(\beta)+(1-k)\bar{v}(\gamma)=v_\chi^*(\mu_0^\infty).
\end{aligned}$$

This proves the claim, and so too the proposition. □

A.3.3 Collapse of Trust: Proof of Proposition 2.2

Proof. Two of three implications are easy given Corollary A.3.2. First, if there is no conflict, then Lipnowski and Ravid (2019, Lemma 1) tells us that there is a 0-equilibrium with full information that generates sender value $\max v(\Delta\Theta)\geqslant v_{\mathbf{1}}^*$; in particular, $v_{\mathbf{0}}^*=v_{\mathbf{1}}^*$. Second, if $v_{\mathbf{0}}^*=v_{\mathbf{1}}^*$, then v_χ^* is constant in χ, ruling out a collapse of trust. Below we show that any conflict whatsoever implies a collapse of trust.

Suppose there is conflict; that is, $\min_{\theta\in\Theta}v(\delta_\theta)<\max v(\Delta\Theta)$. Taking a positive affine transformation of u_S, we may assume without loss that $\min v(\Delta\Theta)=0$ and (since $v(\Delta\Theta)\subseteq u_S(A)$ is finite) $\min[v(\Delta\Theta)\setminus\{0\}]=1$. The set $D:=\operatorname{argmin}_{\mu\in\Delta\Theta}v(\mu)=v^{-1}(-\infty,1)$ is then open and nonempty. We can then consider some full-support prior $\mu_0\in D$. For

any scalar $\widehat{x} \in [0,1]$, let

$$\Gamma(\widehat{x}) := \left\{(\beta, \gamma, k) \in \Delta\Theta \times (\Delta\Theta \setminus D) \times [0,1] : \ k\beta + (1-k)\gamma = \mu_0, \ (1-k)\gamma \geqslant (1-\widehat{x})\mu_0\right\},$$

and $K(\widehat{x})$ be its projection onto its last coordinate. As the correspondence Γ is upper hemicontinuous and decreasing (with respect to set containment), K inherits the same properties. Next, notice that $K(1) \ni 1$ (as v is nonconstant by hypothesis, so that $\Delta\Theta \neq D$) and $K(0) = \varnothing$ (as $\mu_0 \in D$). Therefore, $x := \min\{\widehat{x} \in [0,1] : \ K(\widehat{x}) \neq \varnothing\}$ exists and belongs to $(0,1]$.

Given any scalar $x' \in [0,x)$, it must be that $K(x') = \varnothing$. That is, if $\beta, \gamma \in \Delta\Theta$ and $k \in [0,1]$ with $k\beta + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geqslant (1-\widehat{x})\mu_0$, then $\gamma \in D$. By Theorem 2.1, then, $v^*_{x'\mathbf{1}}(\mu_0) = v(\mu_0) = 0$.

There is, however, some $k \in K(x)$. By Theorem 2.1 and the definition of Γ, there is therefore a $x\mathbf{1}$-equilibrium generating ex-ante sender payoff of at least $k \cdot 0 + (1-k) \cdot 1 = (1-k) \geqslant (1-x)$. If $x < 1$, a collapse of trust occurs at credibility level x.

The only remaining case is the case that $x = 1$. In this case, there is some $\epsilon \in (0,1)$ and $\mu \in \Delta\Theta \setminus D$ such that $\epsilon\mu \leqslant \mu_0$. Then

$$v^*_{x\mathbf{1}}(\mu_0) \geqslant \epsilon\, v(\mu) + (1-\epsilon)\, v\left(\tfrac{\mu_0 - \epsilon\mu}{1-\epsilon}\right) \geqslant \epsilon.$$

So again, a collapse of trust occurs at credibility level x. □

A.3.4 Robustness: Proof of Proposition 2.3

Proof. By Lipnowski and Ravid (2019, Lemma 1 and Theorem 2), S gets the benefit of the doubt (i.e. every $\theta \in \Theta$ is in the support of some member of $\operatorname{argmax}_{\mu \in \Delta\Theta} v(\mu)$) if and only if there is some full-support $\gamma \in \Delta\Theta$ such that $\bar{v}(\gamma) = \max v(\Delta\Theta)$.

First, given a full-support prior μ_0, suppose $\gamma \in \Delta\Theta$ is full-support with $\bar{v}(\gamma) = \max v(\Delta\Theta)$. It follows immediately that $\widehat{v}_{\wedge\gamma} = \widehat{v} = v_{\mathbf{1}}^*$.

Let $r_0 := \min_{\theta\in\Theta} \frac{\mu_0\{\theta\}}{\gamma\{\theta\}} \in (0,\infty)$ and $r_1 := \max_{\theta\in\Theta} \frac{\mu_0\{\theta\}}{\gamma\{\theta\}} \in [r_0,\infty)$. Then Theorem 2.1 tells us that, for $\chi \in \left[\frac{r_1-r_0}{r_1},1\right)^\Theta$, letting $\underline{x} := \min_{\theta\in\Theta}\chi(\theta) \in \left[\frac{r_1-r_0}{r_1},1\right)$:

$$\begin{aligned}
v_\chi^*(\mu_0) &\geqslant \sup_{\beta\in\Delta\Theta,\ k\in[0,1]} \left\{k v_{\mathbf{1}}^*(\beta) + (1-k)\, v(\gamma)\right\} \\
&\qquad \text{s.t.} \qquad k\beta + (1-k)\gamma = \mu_0,\ (1-k)\gamma \geqslant (1-\underline{x})\mu_0 \\
&= \sup_{k\in[0,1]} \left\{k v_{\mathbf{1}}^*\left(\tfrac{\mu_0-(1-k)\gamma}{k}\right) + (1-k)\, v(\gamma)\right\} \\
&\quad \text{s.t.} \ \ (1-\underline{x})\mu_0 \leqslant (1-k)\gamma \leqslant \mu_0 \\
&\geqslant \sup_{k\in[0,1]} \left\{k v_{\mathbf{1}}^*\left(\tfrac{\mu_0-(1-k)\gamma}{k}\right) + (1-k)\, v(\gamma)\right\} \\
&\quad \text{s.t.} \ \ (1-\underline{x})r_1 \leqslant (1-k) \leqslant r_0 \\
&\geqslant \sup_{k\in[0,1]} \left\{k v_{\mathbf{1}}^*\left(\tfrac{\mu_0-(1-k)\gamma}{k}\right) + (1-k)\, v(\gamma)\right\} \\
&\quad \text{s.t.} \ \ (1-\underline{x})r_1 = (1-k) \\
&= \left[1-(1-\underline{x})r_1\right] v_{\mathbf{1}}^*\left(\tfrac{\mu_0-(1-\underline{x})r_1\gamma}{1-(1-\underline{x})r_1}\right) + (1-\underline{x})r_1\, v(\gamma).
\end{aligned}$$

But notice that $v_{\mathbf{1}}^*$, being a concave function on a finite-dimensional space, is continuous on the interior of its domain. Therefore, $v_{\mathbf{1}}^*\left(\frac{\mu_0-(1-\underline{x})r_1\gamma}{1-(1-\underline{x})r_1}\right) \to v_{\mathbf{1}}^*(\mu_0)$ as $\chi \to \mathbf{1}$,[11] implying $\liminf_{\chi\nearrow 1} v_\chi^*(\mu_0) \geqslant v_{\mathbf{1}}^*(\mu_0)$. Finally, monotonicity of $\chi \mapsto v_\chi^*(\mu_0)$ implies $v_\chi^*(\mu_0) \to v_{\mathbf{1}}^*(\mu_0)$ as $\chi \to \mathbf{1}$. That is, persuasion is robust to limited commitment.

Conversely, suppose that S does not get the benefit of the doubt (which of course implies v is non-constant). Taking an affine transformation of u_S, we may assume without loss that $\max v(\Delta\Theta) = 1$ and (since $v(\Delta\Theta) \subseteq u_S(A)$ is finite) $\max[\bar{v}(\Delta\Theta)\setminus\{1\}] = 0$.

Consider any full-support prior μ_0. We will now prove a slightly stronger robustness result, that $v_\chi^*(\mu_0) \not\to v_{\mathbf{1}}^*(\mu_0)$ as $\chi \to \mathbf{1}$ *even if* we restrict attention to imperfect credibility

[11] Note that Θ is finite, so that $\chi(\cdot) \to \mathbf{1}$ is equivalent to $\underline{x} \to 1$.

which is independent of the state. To that end, take any constant $\chi \in [0,1)$. For any $\beta, \gamma \in \Delta\Theta$, $k \in [0,1]$ with $k\beta + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geqslant (\mathbf{1} - \chi)\mu_0$, that S does not get the benefit of the doubt implies (say by Lipnowski and Ravid (2019, Theorem 1)) that $\bar{v}(\gamma) \leqslant 0$, and therefore that $k\widehat{v}_{\wedge\gamma}(\beta) + (1-k)v(\gamma) \leqslant 0$. Theorem 2.1 then implies that $v^*_\chi(\mu_0) \leqslant 0$.

Fix some full-support $\mu_1 \in \Delta\Theta$ and some $\gamma \in \Delta\Theta$ with $v(\gamma) = 1$. For any $\epsilon \in (0,1)$, the prior $\mu_\epsilon := (1-\epsilon)\gamma + \epsilon\mu_1$ has full support and satisfies

$$v^*_{\mathbf{1}}(\mu_\epsilon) \geqslant (1-\epsilon)v(\gamma) + \epsilon v(\mu_1) \geqslant (1-\epsilon) + \epsilon \cdot \min v(\Delta\Theta).$$

For sufficiently small ϵ, then, $v^*_{\mathbf{1}}(\mu_\epsilon) > 0$. Persuasion is therefore not robust to limited commitment at prior μ_ϵ. □

A.3.5 Persuading the Public: Proofs from Section 2.5

Mathematical preliminaries

In this subsection, we document some notations and basic properties that are useful for the present case of $\Theta = [0,1]$, with the sender's value depending only on the receiver's posterior expectation of the state. This environment is studied by Gentzkow and Kamenica (2016) and others. Throughout the subsection, let $\theta_0 := E\mu_0$ be the prior mean; let

$$\mathcal{I} := \{I : \mathbb{R}_+ \to \mathbb{R}_+ : \ I \text{ convex}, \ I(0) = 0, \ I|_{[1,\infty)} \text{ affine}\};$$

let I' denote the right-hand-side derivative of I for any $I \in \mathcal{I}$; and let

$$\mathcal{I}(I) := \{\widehat{I} \in \mathcal{I} : \ I'(1) = \widehat{I}'(1), \ I(1) = \widehat{I}(1), \ \widehat{I} \leqslant I\}$$

for any $I \in \mathcal{I}$.

Fact A.1. Let $\mathcal{M}$ be the set of finite, positive, countably additive Borel measures on Θ.

1. For any $\eta \in \mathcal{M}$, the function $I_\eta : \mathbb{R}_+ \to \mathbb{R}_+$ given by $\bar{\theta} \mapsto \int_0^{\bar{\theta}} \eta[0,\theta] \, \mathrm{d}\theta$ is a member of $\mathcal{I}$.

2. For any $I \in \mathcal{I}$, the function I' is the CDF of some $\eta \in \mathcal{M}$ such that $I_\eta = I$.

3. Any $\eta \in \mathcal{M}$ has total mass $I'_\eta(1)$ and, if $\eta \in \Delta\Theta$, has barycenter $1 - I_\eta(1)$.

The proof of the above fact is immediate, invoking the fundamental theorem of calculus for the second point and integration by parts for the third.

Fact A.2. Given $\mu, \widehat{\mu} \in \Delta\Theta$, the following are equivalent:

1. $\widehat{\mu} = p \circ E^{-1}$ for some $p \in \mathcal{R}(\mu)$.

2. μ is a mean-preserving spread of $\widehat{\mu}$.

3. $I_{\widehat{\mu}} \in \mathcal{I}(I_\mu)$.

That the last two points are equivalent is immediate from the definition of a mean-preserving spread. Equivalence between these conditions and the first is as described in Gentzkow and Kamenica (2016). To apply their results, given $\mu \in \Delta\Theta$, notice that:

- A convex function $I : [0,1] \to \mathbb{R}$ with $I(\theta) \leqslant I_\mu(\theta)$ and $I(\theta) \geqslant (\theta - E\mu)_+$ for every $\theta \in [0,1]$ extends (by letting it take slope 1 on $[1,\infty)$) to a member of $\mathcal{I}(I_\mu)$.

- Every element $I \in \mathcal{I}(I_\mu)$ has, for each $\theta \in [0,1]$,

$$I(\theta) - (\theta - E\mu) = \int_\theta^1 [1 - I'(\tilde{\theta})]\, \mathrm{d}\tilde{\theta} \geqslant 0,$$

  so that $I(\theta) \geqslant (\theta - E\mu)_+ = \max\{I_\mu(1) - I'_\mu(1)(1-\theta),\ 0\}$.

Characterizing S-optimal equilibrium

Lemma A.3.4. *Suppose $\bar{I} \in \mathcal{I}$, $I \in \mathcal{I}(\bar{I})$, and $\omega \in [0,1]$. Then there exist $\theta^* \in [0,\omega]$, $\theta^{**} \in [\omega,1]$ and $I^* \in \mathcal{I}(\bar{I})$ such that:*

- *$I^* = \bar{I}$ on $[0,\theta^*]$, I is affine on $[\theta^*,\theta^{**}]$, and $I^{*'}(\theta) = 1$ on $[\theta^{**},1]$;*

- *$I^* - I$ is nonnegative on $[0,\omega]$ and nonpositive on $[\omega,1]$.*

The proof of the lemma is constructive. While tedious to formally verify that the construction is as desired, it is intuitive to picture. We illustrate in Figure A.6. Given the curves I and $\bar{I}$, we wish to construct the curve $I^* \in \mathcal{I}(\bar{I})$. In order to ensure that I^* has the required level and slope at $\theta = 1$, we will construct it to lie above the tangent line $\theta \mapsto \theta - \theta_0$ of $\bar{I}$ at 1. Now, consider positively sloped lines through the point $(\omega, I(\omega))$. Convexity of $\bar{I}$ ensures that some such line lies everywhere below the graph of $\bar{I}$, whence continuity delivers such a line of shallowest slope. This line is necessarily tangent to $\bar{I}$ somewhere to the left of ω: this point will be our θ^*. The same line intersects the tangent line $\theta \mapsto \theta - \theta_0$ to the right of ω: this will be our θ^{**}. Finally, we construct I^* to

coincide with upper bound function $\bar{I}$ to the left of θ^*, the θ^* tangent line on $[\theta^*, \theta^{**}]$, and the 1 tangent line $\theta \mapsto \theta - \theta_0$ to the right of θ^{**}.

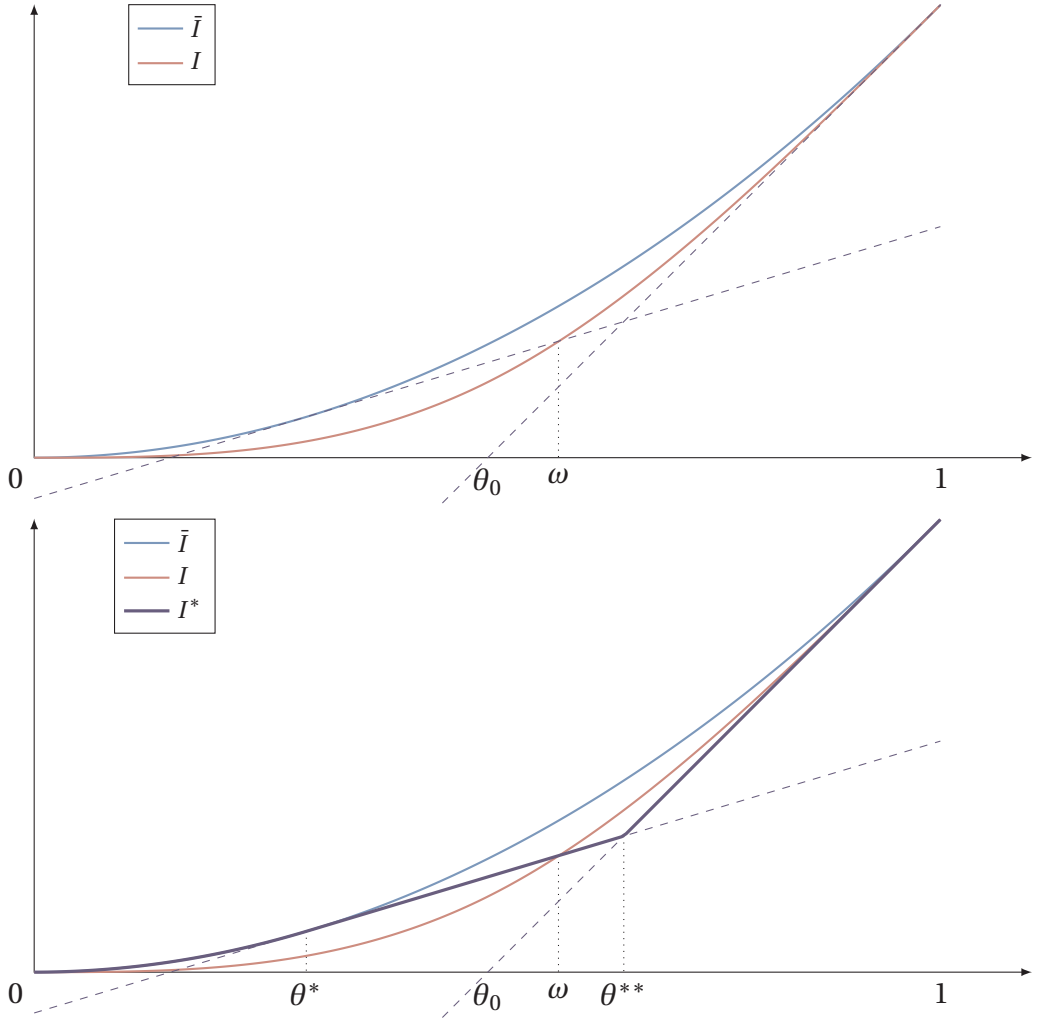


FIGURE A.6: Construction of $(\theta^*, \theta^{**}, I^*)$ in Lemma A.3.4

Proof. Let $\Lambda := \{\lambda \in [0, I'(\omega)] : I(\omega) - \lambda(\omega - \theta) \leqslant \bar{I}(\theta) \text{ for all } \theta \in [0, \omega]\}$. The set Λ is closed because $\bar{I}$ is continuous, and it contains $I'(\omega)$ because I is convex and below $\bar{I}$. So let $\lambda := \min \Lambda$.

Let us now show that there is some $\theta^* \in [0, \omega]$ such that $I(\omega) - \lambda(\omega - \theta^*) = \bar{I}(\theta^*)$. First, if $\lambda = 0$, then $0 \leqslant I(\omega) \leqslant \bar{I}(0) = 0$; and so $\theta^* = 0$ is as desired. Focus now on the case that $\lambda > 0$. The compact subset $\{\bar{I}(\theta) - [I(\omega) - \lambda(\omega - \theta)] : 0 \leqslant \theta \leqslant \omega\}$ of $\mathbb{R}_+$ attains a minimum,

which we wish to show is zero. If the minimum were $\epsilon > 0$, then $\max\{\lambda - \epsilon, 0\} \in \Lambda$ too, a contradiction to $\lambda = \min \Lambda$. So 0 is in the set as desired.

Construct now the function

$$I^* : \mathbb{R}_+ \to \mathbb{R}_+$$
$$\theta \mapsto \begin{cases} \bar{I}(\theta) & : 0 \leqslant \theta \leqslant \theta^* \\ I(\omega) - \lambda(\omega - \theta) & : \theta^* \leqslant \theta \leqslant \omega \\ \max\{I(\omega) + \lambda(\theta - \omega),\ I(1) - I'(1)(1-\theta)\} & : \omega \leqslant \theta. \end{cases}$$

The definition of θ^* ensures $I^*(\theta^*)$ is well-defined. That I is convex implies $I(\omega) + \lambda(\omega - \omega) \geqslant I(1) - (1-\omega)I'(1)$, which in particular ensures that $I(\omega)$ is well-defined. That I is convex and $\lambda \leqslant I'(\omega)$ implies $\max\{I(\omega) + \lambda(1-\omega) \leqslant I(1) - I'(1)(1-1)$. So there is some $\theta^{**} \in [\omega, 1]$ such that $I^*(\theta)$ is equal to $I(\omega) + \lambda(\theta - \omega)$ for $\theta \in [\omega, \theta^{**}]$ and equal to $I(1) - I'(1)(1-\theta)$ for $\theta \in [\theta^{**}, \infty)$. This verifies the first bullet.

It remains to verify that $I^* - I$ is nonpositive on $[\omega, 1]$ and nonnegative on $[0, \omega]$, and that $I^* \in \mathcal{I}(\bar{I})$.

To see that $I^* - I$ is nonpositive above ω, consider any $\theta \in [\omega, 1]$ and use convexity of I. Specifically, first observe that $I(\theta) \geqslant I(1) - I'(1)(1-\theta) = \bar{I}(1) - (1-\theta)\bar{I}'(1)$. Next, that $\lambda \leqslant I'(\omega)$ implies $I(\theta) \geqslant I(\omega) + \lambda(\theta - \omega)$. So $I(\theta) \geqslant I^*(\theta)$. Moreover, $I^{*'} = 1 = I'$ on $[1, \infty)$, so the ranking holds everywhere above ω.

It is immediate that $I^* - I$ is nonnegative on $[0, \theta^*]$, so we turn to showing it is nonnegative on $(\theta^*, \omega]$ too; focus on the nontrivial case with $\theta^* < \omega$. That $I^* \leqslant \bar{I}$ on $(\theta^*, \omega]$ by definition of λ implies $\lambda = \bar{I}'(\theta^*)$. Assume then, for a contradiction, that some $\theta \in (\theta^*, \omega]$ has $I(\theta) > I^*(\theta)$. Then

$$\tfrac{I(\theta) - I(\theta^*)}{\theta - \theta^*} > \tfrac{I^*(\theta) - \bar{I}(\theta^*)}{\theta - \theta^*} = \lambda.$$

But then, I being convex, $I(\omega) > I(\theta) + \lambda(\omega - \theta) > I^*(\theta) + \lambda(\omega - \theta) = I(\omega)$, a contradiction. Thus $I^* - I$ is nonnegative on $[0, \theta^*]$ as desired.

All that remains is to show that $I^* \in \mathcal{I}(\bar{I})$. Letting $\underline{I} : \mathbb{R}_+ \to \mathbb{R}_+$ be given by $\underline{I}(\theta) := \max\{\bar{I}(1) - \bar{I}'(1)(1-\theta),\ 0\}$, we need to check that $\underline{I} \leqslant I^* \leqslant \bar{I}$ and I is convex.

On $[0, \theta^*]$, we have $I^* = \bar{I} \geqslant \underline{I}$. On $[\theta^*, \omega]$, we have shown that $I^* \geqslant I \geqslant \underline{I}$, and we know $I^* \leqslant \bar{I}$ by the definition of λ. On $[\omega, \infty)$, we have shown that $I^* \leqslant I \leqslant \bar{I}$, and we have $I^* \geqslant \underline{I}$ by definition. So $\underline{I} \leqslant I^* \leqslant \bar{I}$ globally.

Finally, we verify convexity. Because the two affine functions coincide at $\theta^{**} \geqslant \theta^*$, we know that $I^*(\theta) = \max\{I(\omega) + \lambda(\theta - \omega),\ I(1) - (1-\theta)I'(1)\}$ for $\theta \in [\theta^*, \infty)$. A maximum of two affine functions, $I^*|_{[\theta^*, \infty)}$ is convex. Moreover, $I^*|_{[0, \theta^*]}$ is convex. Globally convexity then follows if I^* is subdifferentiable at θ^*. But λ is a subdifferential of $\bar{I} \geqslant I^*$ at θ^*, and the two functions coincide at θ^*. It is therefore a subdifferential for I^* at the same, as required. □

Lemma A.3.5. *Suppose $\tilde{H} : \Theta \to \mathbb{R}$ has $\tilde{H}(\cdot) = \tilde{H}(0) + \int_0^{(\cdot)} \tilde{h}(\theta)\,\mathrm{d}\theta$ for some $\tilde{h}$ of bounded variation. Then, for any $I, \widehat{I} \in \mathcal{I}$ such that $I(1) - \widehat{I}(1) = I'(1) - \widehat{I}'(1) = 0$, we have*

$$\left[\tilde{H}(0)\widehat{I}'(0) + \int_0^1 \tilde{H}\,\mathrm{d}\widehat{I}'\right] - \left[\tilde{H}(0)I'(0) + \int_0^1 \tilde{H}\,\mathrm{d}I'\right] = \int_0^1 (\widehat{I} - I)\,\mathrm{d}\tilde{h}.$$

Proof.

$$\begin{aligned}
&\left[\tilde{H}(0)\widehat{I}'(0)+\int_0^1 \tilde{H}\,\mathrm{d}\widehat{I}'\right]-\left[\tilde{H}(0)I'(0)+\int_0^1 \tilde{H}\,\mathrm{d}I'\right]\\
&=\tilde{H}(0)(\widehat{I}-I)'(0)+\int_0^1 \tilde{H}\,\mathrm{d}(\widehat{I}-I)'\\
&=\tilde{H}(0)(\widehat{I}-I)'(0)+\left[(\widehat{I}-I)'\tilde{H}\right]_0^1-\int_0^1 (\widehat{I}-I)'\,\mathrm{d}\tilde{H}\\
&=-\int_0^1 (\widehat{I}-I)'(\theta)\tilde{h}(\theta)\,\mathrm{d}\theta\\
&=-\left[(\widehat{I}-I)\tilde{h}\right]_0^1+\int_0^1 (\widehat{I}-I)\,\mathrm{d}\tilde{h}\\
&=\int_0^1 (\widehat{I}-I)\,\mathrm{d}\tilde{h}.
\end{aligned}$$

□

We now complete our elementary proof that upper censorship is an optimal persuasion rule for convex-concave objectives. Recall, for $\theta^* \in [0,1]$ and $\mu \in \Delta\Theta$, a θ^* upper censorship of μ is

$$\mathbf{1}_{[0,\theta^*)}\mu+\mu[\theta^*,1]\delta_{\frac{1}{\mu[\theta^*,1]}\int_{[\theta^*,1]}\theta\,\mathrm{d}\mu(\theta)} \in \Delta\Theta$$

if $\mu[\theta^*,1]>0$, and simply μ if $\mu[\theta^*,1]=0$.

Lemma A.3.6. *Suppose $\tilde{H}:\Theta\to\mathbb{R}$ is continuous, and $\omega\in[0,1]$ is such that $\tilde{H}$ is (strictly) convex on $[0,\omega]$ and (strictly) concave on $[\omega,1]$. Then, if $\bar{\mu}\in\Delta\Theta$ has no atoms $<\max\operatorname{supp}(\bar{\mu})$, some (every) solution to $\max_{\mu\in\Delta\Theta:\ \mu\preccurlyeq\bar{\mu}}\int\tilde{H}\,\mathrm{d}\mu$ is a θ^* upper censorship of $\bar{\mu}$ for some $\theta^*\in[0,\omega]$. Moreover, this θ^* upper censorship puts probability 1 on $[0,\theta^*]\cup[\omega,1]$.*

Proof. Let μ be a solution to the given program. Taking $\bar{I}:=I_{\bar{\mu}}$ and $I:=I_\mu$, note that the conditions of Lemma A.3.4 are satisfied. Let $I^*\in\mathcal{I}$, $\theta^*\in[0,\omega]$, and $\theta^{**}\in[\omega,1]$ be as delivered by Lemma A.3.4 and $\mu^*\in\Delta\Theta$ be such that $I^*=I_{\mu^*}$. Then, by Lemma A.3.5

(letting $\tilde{h} := \tilde{H}'$),

$$\begin{aligned}\int_0^1 \tilde{H}\,\mathrm{d}\mu^* - \int_0^1 \tilde{H}\,\mathrm{d}\mu &= \tilde{H}(0)(I^* - I)'(0) + \int_0^1 \tilde{H}\,\mathrm{d}(I^* - I)' \\ &= \int_0^\omega (I^* - I)\,\mathrm{d}\tilde{h} + \int_\omega^1 (I - I^*)\,\mathrm{d}(-\tilde{h}).\end{aligned}$$

As $\tilde{h}$ is (strictly) increasing on $[0,\omega)$ and (strictly) decreasing on $[\omega,1]$, it follows from the definition of I^* that $\int_0^1 \tilde{H}\,\mathrm{d}\mu^* \geqslant \int_0^1 \tilde{H}\,\mathrm{d}\mu$, (strictly so, given continuity of $I^* - I$, unless $I = I^*$). Optimality of μ then tells us that μ^* is optimal (and equal to μ).

By construction, $\mu^*[0,\theta] = \bar{\mu}[0,\theta]$ for every $\theta \in [0,\theta^*)$, and (since, by hypothesis, $\bar{\mu}\{\theta^*\} = 0$ if $\bar{\mu}(\theta^*,1] > 0$) we have $|[\theta^*,1] \cap \operatorname{supp}(\mu^*)| = 1$. But these properties—which will clearly also be satisfied by a θ^* upper censorship of $\bar{\mu}$—characterize a unique distribution of any given mean. Therefore, μ^* is a θ^* upper censorship of $\bar{\mu}$.

Finally, the "moreover" point follows from $\theta^{**} \geqslant \omega$, as guaranteed by Lemma A.3.4.

□

Lemma A.3.7. *There is a unique $\bar{\theta}_\chi \in [0,1]$ such that*[12] $\int_0^{\bar{\theta}} \chi\mu_0[0,\theta]\,\mathrm{d}\theta$ *is*

$$\begin{cases} > \bar{\theta} - \theta_0 & \text{for } \bar{\theta} \in [0,\bar{\theta}_\chi) \\ = \bar{\theta} - \theta_0 & \text{for } \bar{\theta} = \bar{\theta}_\chi \\ < \bar{\theta} - \theta_0 & \text{for } \bar{\theta} \in (\bar{\theta}_\chi, 1]. \end{cases}$$

Moreover, $\bar{\theta}_\chi \geqslant \theta_0$ and, if credibility is imperfect, $\bar{\theta}_\chi < 1$.

Proof. Let $\varphi(\bar{\theta}) := (\bar{\theta} - \theta_0) - \int_0^{\bar{\theta}} \chi\mu_0[0,\theta]\,\mathrm{d}\theta = \int_0^{\bar{\theta}} \big(1 - \chi\mu_0[0,\theta]\big)\,\mathrm{d}\theta - \theta_0$ for $\bar{\theta} \in \Theta$. Clearly, φ is continuous and strictly increasing. Next, observe that $\varphi(\theta_0) = -\int_0^{\theta_0} \chi\mu_0[0,\theta]\,\mathrm{d}\theta \leqslant 0$, and

$$\varphi(1) = (1 - \theta_0) - \int_0^1 \chi\mu_0[0,\theta]\,\mathrm{d}\theta = I_{\mu_0}(1) - I_{\chi\mu_0}(1) = I_{(\mathbf{1}-\chi)\mu_0}(1) \geqslant 0,$$

[12] Integration by parts shows that this definition of $\bar{\theta}_\chi$ is equivalent to that in Equation θ^*-IC.

with the last inequality being strict if $\chi\mu_0 \neq \mu_0$. The result then follows from the intermediate value theorem. □

In what follows, recall the mean distribution $\bar{\mu}_\chi$ as defined in Section 2.5.

Lemma A.3.8. *For any $\theta \in [0,1]$, we have*

$$I_{\bar{\mu}_\chi}(\theta) = \max\{I_{\chi\mu_0}(\theta),\ \theta - \theta_0\} = \begin{cases} I_{\chi\mu_0}(\theta) & :\ \theta \leqslant \bar{\theta}_\chi \\ \theta - \theta_0 & :\ \theta \geqslant \bar{\theta}_\chi. \end{cases}$$

Moreover, $E\bar{\mu}_\chi = \theta_0$.

Proof. That $I_{\bar{\mu}_\chi}$ coincides with $I_{\chi\mu_0}$ on $[0,\bar{\theta}_\chi]$ and has derivative 1 on $(\bar{\theta}_\chi, 1]$ follows directly from the definition of $\bar{\mu}_\chi$. Noting that $I_{\chi\mu_0}(\bar{\theta}_\chi) = \bar{\theta}_\chi - \theta_0$ by Lemma A.3.7, it follows that $I_{\bar{\mu}_\chi}(\theta) = \theta - \theta_0$ for $\theta \in [\bar{\theta}_\chi, 1]$.

Next, recall that $I_{\chi\mu_0}(\theta) - (\theta - \theta_0)$ is nonnegative for $\theta \in [0,\bar{\theta}_\chi]$ and nonpositive for $\theta \in [\bar{\theta}_\chi, 1]$ by Lemma A.3.7. Consequently, $I_{\bar{\mu}_\chi}(\theta) = \max\{I_{\chi\mu_0}(\theta),\ \theta - \theta_0\}$ for every $\theta \in [0,1]$.

Finally, $E\bar{\mu}_\chi = 1 - I_{\bar{\mu}_\chi}(1) = \theta_0$. □

We now prove Claim 2.1.

Proof. First, we show that $\widehat{v}(\bar{\mu}_\chi) = \max_{\theta^* \in [0,\bar{\theta}_\chi]} \int H \,\mathrm{d}\mu_{\chi,\theta^*}$, and that the maximum on the RHS is attained. By Lemma A.3.6, there is some $\theta^* \in [0,1]$ such that $\widehat{v}(\bar{\mu}_\chi) = \int H \,\mathrm{d}\mu_{\chi,\theta^*}$. As $\bar{\mu}_\chi[0,\bar{\theta}_\chi] = 1$, we have $\mu_{\chi,\theta} = \mu_{\chi,\bar{\theta}_\chi}$ for every $\theta \in [\bar{\theta}_\chi, 1]$; so we may without loss take $\theta^* \leqslant \bar{\theta}_\chi$. Furthermore, since

$$\int H \,\mathrm{d}\mu_{\chi,\theta^*} = \widehat{v}(\bar{\mu}_\chi) = \max_{\mu \preccurlyeq \bar{\mu}_\chi} \int H \,\mathrm{d}\mu \geqslant \int H \,\mathrm{d}\mu_{\chi,\theta}$$

for every $\theta \in [0,\bar{\theta}_\chi]$, the maximum is attained.

Next, given $\theta^* \in [0, \bar{\theta}_\chi]$, we exhibit an equilibrium in which S communicates via a θ^*-upper-censorship pair, and observe that this induces S value $\int H \, \mathrm{d}\mu_{\chi,\theta^*}$—in particular showing $\int H \, \mathrm{d}\mu_{\chi,\theta^*} \leqslant v^*_\chi(\mu_0)$. To that end, define the belief map $\pi : M \to \Delta\Theta$ via

$$\pi(m) = \begin{cases} \delta_m & : \ m \in [0, \theta^*) \\ \gamma & : \ \text{otherwise,} \end{cases}$$

where $\gamma := \frac{[1-\chi \mathbf{1}_{[0,\theta^*)}]\mu_0}{1-\chi\mu_0[0,\theta^*)}$ (with $\gamma := \delta_1$ if $\chi\mu_0[0,\theta^*) = 1$). Then let R behavior be given by $\alpha := H \circ E \circ \pi$. The Bayesian property is now straightforward, and the R incentive condition holds by construction. To verify that this is a χ-equilibrium, then, we need only check that S behavior is optimal under influenced reporting. As the set of interim own-payoffs S can induce with some message is $\{H(\theta) : \ \theta \in [0, \theta^*) \text{ or } \theta = E\gamma\}$, and H is strictly increasing on $[0,1]$, it remains to show that $E\gamma \geqslant \theta^*$. This holds vacuously if $\gamma = \delta_1$, so focus on the alternative case in which $\bar{\mu}_\chi[\theta^*, \bar{\theta}_\chi] > 0$. In this case,

$$\begin{aligned}
\bar{\mu}_\chi[\theta^*, \bar{\theta}_\chi] \ \left(E\gamma - \theta^*\right) &= \int_{[\theta^*, \bar{\theta}_\chi]} (\theta - \theta^*) \, \mathrm{d}\bar{\mu}_\chi(\theta) \\
&= -\int_{[\theta^*, 1]} (\theta^* - \theta) \, \mathrm{d}\bar{\mu}_\chi(\theta) \\
&= \int_{[0,\theta^*)} (\theta^* - \theta) \, \mathrm{d}\bar{\mu}_\chi(\theta) - (\theta^* - \theta_0) \ \text{(by Lemma A.3.8)} \\
&= \left[(\theta^* - \theta)\bar{\mu}_\chi[0,\theta]\right]_0^{(\theta^*)^-} - \int_{[0,\theta^*)} (-1)\bar{\mu}_\chi[0,\theta] \, \mathrm{d}\theta - (\theta^* - \theta_0) \\
&= [0 - 0] + I_{\chi\mu_0}(\theta^*) - (\theta^* - \theta_0) \\
&\geqslant 0 \text{ by Lemma A.3.8.}
\end{aligned}$$

S incentive-compatibility follows. To show this equilibrium generates the required payoff, it suffices to show that the induced distribution μ of posterior means is equal to

μ_{χ,θ^*}. For any $\theta \in [0,\theta^*)$, notice that

$$\mu[0,\theta) = \int_0^\theta \chi \,\mathrm{d}\mu_0 = \bar{\mu}_\chi[0,\theta) = \mu_{\chi,\theta^*}[0,\theta).$$

Moreover, $|[\theta^*,1] \cap \operatorname{supp}(\mu)| = 1 = |[\theta^*,1] \cap \operatorname{supp}(\mu_{\chi,\theta^*})|$. Equality then follows from equality of their means (Lemma A.3.8).

Finally, we show that $v^*_\chi(\mu_0) \leqslant \widehat{v}(\bar{\mu}_\chi)$. To that end, let (β,γ,k) solve the program of Theorem 2.1 – and, without loss, say $\beta = \mu_0$ if $k = 0$. Let $\omega := \omega^* \wedge E\gamma$, and see that $H(E\gamma) \wedge H$ is continuous, convex on $[0,\omega]$, and concave on $[\omega,1]$. Therefore, by Lemma A.3.6, there is some $\theta^* \in [0,\omega]$ such that the θ^* upper censorship of β belongs to $\operatorname{argmax}_{\widehat{\beta} \preccurlyeq \beta} \int H(E\gamma) \wedge H \,\mathrm{d}\widehat{\beta}$. Let $\lambda := \beta[0,\theta^*) \in [0,1]$, $\eta := \frac{\mathbf{1}_{[\theta^*,1]}\beta}{1-\lambda} \in \Delta\Theta$, $\widehat{\gamma} := \frac{(1-k)\gamma + (1-\lambda)k\eta}{1-\lambda k} \in \Delta\Theta$, and $\widehat{\beta} := \frac{\mathbf{1}_{[0,\theta^*)}\beta}{\lambda} \in \Delta\Theta$.[13] Two observations will enable us to bound S payoffs across all equilibria. First, as a monotone transformation of an affine functional, $v = H \circ E$ is quasiconcave, implying $\bar{v} = v$. Second, Lemma A.3.6 tells us $E\eta \geqslant \omega$, so that $H(E\gamma) \wedge H$ is concave on $\operatorname{co}\{E\gamma, E\eta\}$. Now, observe that

$$\begin{aligned}
v^*_\chi(\mu_0) &= k\widehat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) \\
&= k\int H(E\gamma) \wedge H \,\mathrm{d}\left[\mathbf{1}_{[0,\theta^*)}\beta + (1-\lambda)\delta_{E\eta}\right] + (1-k)H(E\gamma) \\
&= k\left[\lambda\int H \,\mathrm{d}\widehat{\beta} + (1-\lambda)H(E\gamma) \wedge H(E\eta)\right] + (1-k)H(E\gamma) \wedge H(E\gamma) \\
&\leqslant k\lambda\int H \,\mathrm{d}\widehat{\beta} + (1-k\lambda)H(E\gamma) \wedge H(E\widehat{\gamma}) \\
&\leqslant \int H \,\mathrm{d}\left[k\lambda\widehat{\beta} + (1-\lambda k)\delta_{E\widehat{\gamma}}\right] \\
&\leqslant \widehat{v}\left(k\lambda\widehat{\beta} + (1-\lambda k)\delta_{E\widehat{\gamma}}\right).
\end{aligned}$$

[13] In case any of the described objects is defined by an expression with a zero denominator, we define it as follows: $\eta := \delta_1$ if $\lambda = 1$, $\widehat{\gamma} := \delta_1$ if $\lambda k = 1$, and $\widehat{\beta} := \delta_0$ if $\lambda = 0$.

Letting $\widehat{\mu} := k\lambda\widehat{\beta} + (1-\lambda k)\delta_{E\widehat{\gamma}}$, the payoff ranking (and so too the claim) will follow if we show that $\widehat{\mu} \preccurlyeq \bar{\mu}_\chi$. As (appealing to Lemma A.3.8) $E\bar{\mu}_\chi = \theta_0 = E\widehat{\mu}$, it suffices to show that $I_{\widehat{\mu}} \leqslant I_{\bar{\mu}_\chi}$.

For $\theta \in [0, E\widehat{\gamma})$, we have $\delta_{E\widehat{\gamma}}[0,\theta] = 0$. Therefore, over the interval $[0, E\widehat{\gamma}]$, we have

$$I_{\widehat{\mu}} = I_{\lambda k\widehat{\beta}} + (1-\lambda k) I_{\delta_{E\widehat{\gamma}}} = I_{\lambda k\widehat{\beta}} \leqslant I_{k\beta} = I_{\mu_0} - I_{(1-k)\gamma} \leqslant I_{\mu_0} - I_{(\mathbf{1}-\chi)\mu_0} = I_{\chi\mu_0}.$$

Now, as $I_{\widehat{\mu}}(1) = 1-\theta_0$ and (since $E\widehat{\gamma} \geqslant \theta_0$) we have $I'_{\widehat{\mu}}|_{(E\widehat{\gamma},1)} = 1$, we know $I_{\widehat{\mu}}(\theta) = \theta - \theta_0$ for $\theta \in [E\widehat{\gamma}, 1]$. In particular, we learn that $I_{\widehat{\mu}}(\theta) \leqslant \max\{I_{\chi\mu_0}(\theta), \theta - \theta_0\}$ for $\theta \in [0, E\widehat{\gamma}] \cup [E\widehat{\gamma}, 1]$. Lemma A.3.8 then tells us that $I_{\widehat{\mu}} \leqslant I_{\bar{\mu}_\chi}$. □

Comparative Statics

Now, we prove Claim 2.2. In fact, because the proof applies without change, we prove a slightly stronger result, providing comparative statics results in the credibility function *and the prior*, holding the prior mean fixed. Specifically, given two pairs of parameters $\langle \mu_0, \chi\rangle$ and $\langle \tilde{\mu}_0, \tilde{\chi}\rangle$ such that $E\mu_0 = E\tilde{\mu}_0 = \theta_0$, we show that $v^*_\chi(\mu_0) \geqslant v^*_{\tilde{\chi}}(\tilde{\mu}_0)$ if and only if $\bar{\mu}_\chi \succcurlyeq \bar{\tilde{\mu}}_{\tilde{\chi}}$.

Proof. Appealing to Claim 2.1 and Lemma A.3.5,

$$\begin{aligned} v^*_\chi(\mu_0) - v^*_{\tilde{\chi}}(\mu_0) &= \widehat{v}(\bar{\mu}_\chi) - \widehat{v}(\bar{\tilde{\mu}}_{\tilde{\chi}}) \\ &= \max_{I \in \mathcal{I}(I_{\bar{\mu}_\chi})} \left[H(0) I'(0) + \int_0^1 H \,\mathrm{d}I' \right] - \max_{\tilde{I} \in \mathcal{I}(I_{\bar{\tilde{\mu}}_{\tilde{\chi}}})} \left[H(0) \tilde{I}'(0) + \int_0^1 H \,\mathrm{d}\tilde{I}' \right] \\ &= \max_{I \in \mathcal{I}(I_{\bar{\mu}_\chi})} \int_0^1 H \,\mathrm{d}I' - \max_{\tilde{I} \in \mathcal{I}(I_{\bar{\tilde{\mu}}_{\tilde{\chi}}})} \int_0^1 H \,\mathrm{d}\tilde{I}' \\ &= \max_{I \in \mathcal{I}(I_{\bar{\mu}_\chi})} \int_0^1 I \,\mathrm{d}h - \max_{\tilde{I} \in \mathcal{I}(I_{\bar{\tilde{\mu}}_{\tilde{\chi}}})} \int_0^1 \tilde{I} \,\mathrm{d}h. \end{aligned}$$

Let $I_* := I_{\bar{\mu}_\chi}$ and $\tilde{I}_* := I_{\bar{\tilde{\mu}}_{\tilde{\chi}}}$. We now need to show that $\max_{I \in \mathcal{I}(I_*)} \int_0^1 I \,\mathrm{d}h \geqslant \max_{\tilde{I} \in \mathcal{I}(\tilde{I}_*)} \int_0^1 \tilde{I} \,\mathrm{d}h$ for every continuous, strictly quasiconcave $h : [0,1] \to \mathbb{R}$ if and only if $I_* \geqslant \tilde{I}_*$.

First, if $I_* \leqslant \tilde{I}_*$ then $\mathcal{I}(I_*) \subseteq \mathcal{I}(\tilde{I}_*)$, delivering the payoff ranking.

Conversely, suppose $I_* \not\leqslant \tilde{I}_*$. Then, elements of $\mathcal{I}$ being continuous, there are some $\theta_1, \theta_2 \in \Theta$ such that $\theta_1 < \theta_2$ and $I_* > \tilde{I}_*$ on (θ_1, θ_2). If h is increasing, then

$$v^*_\chi(\mu_0) - v^*_{\tilde{\chi}}(\mu_0) = \int_0^1 I_* \, \mathrm{d}h - \int \tilde{I}_* \, \mathrm{d}h = \int_0^1 (I_* - \tilde{I}_*) \, \mathrm{d}h.$$

As $(I_* - \tilde{I}_*)$ is strictly positive over (θ_1, θ_2), globally bounded, and globally continuous, there is $\epsilon > 0$ small enough that $\left(\epsilon \int_0^{\theta_1} + \int_{\theta_1}^{\theta_2} + \epsilon \int_{\theta_2}^1\right) \left[I_*(\theta) - \tilde{I}_*(\theta)\right] \mathrm{d}\theta > 0$. It is then straightforward to construct a shock distribution whose continuous density h satisfies $h'|_{(0,\theta_1)\cup(\theta_2,1)} = \epsilon\zeta$ and $h'|_{(\theta_1,\theta_2)} = \zeta$ for some $\zeta > 0$. Such a shock distribution witnesses a failure of $v^*_\chi(\mu_0) \geqslant v^*_{\tilde{\chi}}(\tilde{\mu}_0)$.

□

A.3.6 Proofs from Section 2.6: Investing in Credibility

In this section, we prove the following formal claim concerning the public persuasion application with costly endogenous credibility.

Claim A.1. There exists an optimal credibility choice. Moreover, any optimal choice (along with S-optimal equilibrium) is a cutoff credibility choice, and entails full revelation by the official reporting protocol.

Toward the proof, we first establish the following lemma.

Lemma A.3.9. *For any non-cutoff credibility choice (i.e. any χ such that there is no $\theta^* \in [0,1]$ with $\chi = \mathbf{1}_{[0,\theta^*)}$ μ_0-a.s.), there is some cutoff credibility choice that yields S a strictly higher best equilibrium payoff net of costs.*

Proof. Consider any credibility choice χ not of the desired form. In particular, this implies that χ is not μ_0-a.s. equal to 1, so that $\chi\mu_0(\Theta) < 1$.

As μ_0 is atomless, there is some $\theta^* \in [0,1)$ such that $\mu_0[0,\theta^*) = \chi\mu_0(\Theta)$. That $\mathbf{1}_{[0,\theta^*)}\mu_0 \neq \chi\mu_0$ but the two have the same total measure implies that $\operatorname{supp}[(\mathbf{1} - \chi)\mu_0]$

intersects $[0,\theta^*)$. For each $\theta_* \in [0,\theta^*]$, define the function $\eta_{\theta_*} := I_{\mathbf{1}_{[0,\theta_*)}\mu_0} - I_{\chi\mu_0} : \mathbb{R}_+ \to \mathbb{R}$. By construction, its right-hand-side derivative at any θ is given by $\eta'_{\theta_*}(\theta) = \int_0^\theta (\mathbf{1}_{[0,\theta_*)} - \chi)\,\mathrm{d}\mu_0$. In particular, this implies (since $\chi\mu_0$ strictly first-order stochastically dominates $\mathbf{1}_{[0,\theta^*)}$) that η'_{θ^*} is globally nonnegative, weakly quasiconcave with peak at θ^*, and not globally zero. In particular, $\eta_{\theta^*}(0) = 0$ yields $\eta_{\theta^*} \geqslant 0$ and $\epsilon := \frac{1}{2}\eta_{\theta^*}(\theta^*) > 0$. Now, with the prior being atomless and η_{θ^*} continuous, there is some $\theta_* \in [0,\theta^*)$ close enough to θ^* to ensure that $\eta_{\theta_*}(\theta_*) \geqslant \epsilon$ and $\mu_0(\theta_*,\theta^*] \leqslant \epsilon$. Let $\eta := \eta_{\theta_*}$.

As η' is weakly quasiconcave on $[0,1]$ (with peak at θ_*), we have $\inf\eta'[0,1] = \min\{\eta'(0),\eta'(1)\} = \min\{0,\eta'(1)\}$. But

$$\eta'(1) = \int_0^{\theta_*} 1\,\mathrm{d}\mu_0 - \int_0^1 \chi\,\mathrm{d}\mu_0 = \mu_0[0,\theta_*] - \mu_0[0,\theta^*] \geqslant -\epsilon,$$

so that $\eta'|_{[0,1]} \geqslant -\epsilon$.

Let us now observe that η is nonnegative over $[0,1]$. First, any $\theta \in [0,\theta_*]$ has $\eta(\theta) = \eta_{\theta^*}(\theta) \geqslant 0$. Next, any $\theta \in [\theta_*,1]$ has

$$\eta(\theta) = \eta(\theta_*) + \int_{\theta_*}^{\theta} \eta'(\tilde{\theta})\,\mathrm{d}\tilde{\theta} \geqslant \epsilon + (1-\theta_*)(-\epsilon) = \theta_*\epsilon > 0.$$

So $I_{\mathbf{1}_{[0,\theta_*)}\mu_0} \geqslant I_{\chi\mu_0}$ globally. Lemma A.3.8 then implies that $\bar{\mu}_{\mathbf{1}_{[0,\theta_*)}} \succcurlyeq \bar{\mu}_\chi$. Finally, Claim 2.2 tells us that $v^*_{\mathbf{1}_{[0,\theta_*)}}(\mu_0) \geqslant v^*_\chi(\mu_0)$. Meanwhile, the cost of credibility $\mathbf{1}_{[0,\theta_*)}$, is strictly below that of credibility χ. □

Now, we prove Claim A.1

Proof. Consider any credibility choice χ and accompanying χ-equilibrium. Lemma A.3.9 shows that χ is a cutoff credibility choice with cutoff $\theta_* \in [0,1]$, or can be replaced with one for a strict improvement to the objective. Our analysis of public persuasion says that the χ-equilibrium entails influenced θ^* upper censorship for some

cutoff $\theta^* \in [0,1]$, or can be replaced with it for a strict improvement to the objective. Our main-text observation on the endogenous credibility problem (that no gratuitous credibility should be purchased) tells us that $\theta_* \leqslant \theta^*$, or else θ_* can be lowered to θ^* for a strict gain to the objective. But then, since $\chi|_{[\theta_*,1]} = 0$, it is purely a normalization to set $\theta^* = \theta_*$.

The above observations tell us that we may as well restrict to the case that there is some cutoff $\theta^* \in [0,1]$ such that S invests in cutoff credibility choice with cutoff θ^*, official reporting always reveals the state, and influenced reporting reveals itself but provides no further information.

Thus, S solves (where the argument for H on the right is taken to be 1 when $\theta^* = 1$)

$$\max_{\theta^* \in [0,1]} \int_0^{\theta^*} H \, \mathrm{d}\mu_0 - c\left(\mu_0[0,\theta^*)\right) + H\left(\frac{\int_{\theta^*}^1 \theta \, \mathrm{d}\mu_0(\theta)}{\mu_0[\theta^*,1]}\right).$$

This program is continuous with compact domain, so that an optimum exists. □

A.4 Appendix for Chapter 3: Proofs

A.4.1 Preliminaries

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 A.4.1 (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*

$$u_{\bar{\theta}}(y) \leqslant u_{\bar{\theta}}(x) \implies u_{\underline{\theta}}(y) \leqslant u_{\underline{\theta}}(x),$$
$$u_{\underline{\theta}}(y) \geqslant u_{\underline{\theta}}(x) \implies u_{\bar{\theta}}(y) \geqslant u_{\bar{\theta}}(x).$$

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

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

The proof of the first implication is analogous. □

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

Lemma A.4.2. *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*

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

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

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

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

$$\bar{p} - \underline{p} = \phi(q) = v_{\bar{\theta}_{f'}}(q) - v_{\underline{\theta}_f}(q) \Leftrightarrow u_{\underline{\theta}_f}(\underline{p}, q) = u_{\bar{\theta}_f}(\bar{p}, q).$$

□

Lemma A.4.3. *The efficient quality for payoff type θ_f defined as*

$$\widehat{q}_{\theta_f} := \operatorname*{argmax}_{q \geqslant 0} v_{\theta_f}(q) - \kappa(q)$$

exists, is unique and increasing in θ_f.

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

$$\zeta_{\theta_f}(q) := v_{\theta_f}(q) - \kappa(q) \tag{A.1}$$

and note that it is continuous, twice differentiable and satisfies

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

$$\zeta'_{\theta_f}(0) > 0 \tag{A.3}$$

$$\lim_{q\to\infty} \zeta'_{\theta_f}(q) < 0 \tag{A.4}$$

$$\zeta''_{\theta_f} > 0, \tag{A.5}$$

$$\frac{\partial \zeta'_{\theta_f}(\widehat{q}_{\theta_f})}{\partial \theta_f} > 0. \tag{A.6}$$

Properties (A.3) and (A.4) together with the Mean Value Theorem ensure that there exists a unique global efficient quantify $\widehat{q}_{\theta_f}$ defined as

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

Note that (A.5) implies

$$\operatorname{sgn}(\zeta'_{\theta_f}(q)) = \operatorname{sgn}(\widehat{q}_{\theta_f} - q). \tag{A.7}$$

In addition, (A.6) implies that $\widehat{q}_{\theta_f}$ is increasing in θ_f. □

A.4.2 Proofs

Proof of Observation 3.1 on page 75: 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

$\{(p'_i, q_i)\}_{i\in\{1,\dots,n\}}$ set in frame h with

$$\begin{aligned}
p'_i &= p_i + \Delta_i \\
\Delta_1 &= v\left(\theta_h^1, q_1\right) - v\left(\theta_f^1, q_1\right) \\
\Delta_i &= \Delta_{i-1} + v(\theta_h^i, q_i) - v(\theta_f^i, q_i) - \left[v(\theta_h^i, q_{i-1}) - v(\theta_f^i, q_{i-1})\right]
\end{aligned}$$

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

$$\begin{aligned}
v(\theta_h^i, q_i) - p'_i &= v(\theta_f^i, q_i) + v(\theta_h^i, q_i) - v(\theta_f^i, q_i) - p_i - \Delta_i \\
&= v(\theta_f^i, q_{i-1}) + v(\theta_h^i, q_i) - v(\theta_f^i, q_i) - p_{i-1} - \Delta_i \\
&= v(\theta_f^i, q_{i-1}) - p_{i-1} - \Delta_{i-1} + \left[v(\theta_h^i, q_{i-1}) - v(\theta_f^i, q_{i-1})\right] \\
&= v(\theta_h^i, q_{i-1}) - p_{i-1} - \Delta_{i-1} \\
&= v(\theta_h^i, q_{i-1}) - p'_{i-1}
\end{aligned}$$

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

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

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

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

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 3.1 relies on many arguments that are required for the proofs of the following results as well. To avoid repetition, we prove Theorem 3.2 and Theorem 3.2 together.

Proof of Proposition 3.1 on page 80: 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 A.4.4. *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. $\mathbf{0} \succcurlyeq_j^l d_j$ *(decoys are undesirable in l).*

Proof. Let $d_0 = x$ for brevity. We construct the decoys $d_i = (p_i, q_i)$ recursively.[14] For $i \in \{1, \ldots, n\}$, pick

$$d_i \sim_{\theta_l^i} \mathbf{0} \tag{A.8}$$

$$d_i \sim_{\theta_h^i} d_{i-1} \tag{A.9}$$

or equivalently

$$v(\theta_l^i, q_i) - p_i = 0$$
$$v(\theta_h^i, q_i) - p_i = v(\theta_h^i, q_{i-i}) - p_{i-i}$$

Existence of such a d_i follows from Lemma A.4.2. To verify this construction, we proceed through a series of claims.

Claim A.2. Decoy quantities are increasing: $q_i \geqslant q_{i-1}$.

Proof of Claim A.2: By the two defining relations

$$v(\theta_l^i, q_i) = p_i$$
$$v(\theta_h^i, q_i) - p_i = v(\theta_h^i, q_{i-1}) - p_{i-1}$$

Hence

$$v(\theta_h^i, q_i) - v(\theta_l^i, q_i) = v(\theta_h^i, q_{i-1}) - v(\theta_{i-1}^l, q_{i-1}) > v(\theta_h^i, q_{i-1}) - v(\theta_l^i, q_{i-1})$$

[14]The present proof can be extended to the case without ordered types (but maintaining ordered frames).

and $q_i > q_{i-1}$ is established as it is implied by single crossing from

$$\int_{\theta_l^i}^{\theta_h^i} v_\theta(\theta, q_i)\mathrm{d}\theta > \int_{\theta_l^i}^{\theta_h^i} v_\theta(\theta, q_{i-1})\mathrm{d}\theta$$

△

This also shows that all (p, q) are positive, as $q_i \geqslant q_0 = q_x$.

Claim A.3. 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 A.3: 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

$$d_k \sim_{\theta_h^k} d_{k-1}$$

and $q_k > q_{k-1}$. This implies

$$d_k \succcurlyeq_{\theta_h^k} d_{k-1}$$

and by since $\theta_h^i \geqslant \theta_h^k$

$$d_k \succcurlyeq_{\theta_h^i} d_{k-1}$$

The desired result follows by transitivity.

Second, suppose $j > i$. Again for all $k \in [i, j]$,

$$d_k \sim_{\theta_h^k} d_{k-1}$$

and $q_k > q_{k-1}$. But then

$$d_k \preccurlyeq_{\theta_h^i} d_{k-1}$$

for every decoy since $\theta_h^i \geqslant \theta_h^k$. The desired result again follows by transitivity. △

□

Proof of Lemma 3.1 on page 81: A continuation problem for type $\theta \in \Theta$ with contract c_θ satisfying all three properties is given by $e_\theta = (\{\mathbf{0}, (\{\mathbf{0}, c_\theta\} \cup \{d_{\theta'}\}_{\theta' > \theta}, h)\}, l)$, where the contracts $\{d_{\theta'}\}_{\theta' > \theta}$ are constructed in Lemma A.4.4 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 3.2 on page 81: Suppose that $\mathbf{c}$ satisfies $\{\mathrm{P}^h_\theta\}_{\theta \in \Theta_R}$, $\{\mathrm{P}^l_\theta\}_{\theta \in \Theta_C}$, $\{\mathrm{IC}^h_{\theta\theta'}\}_{\theta < \theta'}$, $\{\mathrm{IC}^h_{\theta\theta'}\}_{\theta \in \Theta, \theta' \in \Theta_R}$ for some partition $\{\Theta_C, \Theta_R\}$ of Θ. Then let e_θ for each $\theta \in \Theta_C$ be constructed as in Lemma 3.1 and consider a standard EDP

$$e^* = \Big(\{e_\theta\}_{\theta \in \Theta_C} \cup \{c_\theta\}_{\theta \in \Theta_R} \cup \{\mathbf{0}\}, h\Big).$$

Notice that from Lemma 3.1 it follows for each $\theta \in \Theta_C$ that there exist an outcome σ^θ of e_θ, such that

$$\begin{aligned} &\sigma^\theta(\theta') \in \{\mathbf{0}, c_\theta\}, \forall \theta' \in \Theta, \\ &\sigma^\theta(\theta') = \mathbf{0}, \forall \theta' > \theta, \\ &\sigma^\theta(\theta) = c_\theta. \end{aligned}$$

Now let σ be such that $\sigma(\theta) = c_\theta$. To show that σ is an outcome of e^*, notice that constraints $\{\mathrm{IC}^h_{\theta\theta'}\}_{\theta \in \Theta, \theta' \in \Theta_R}$ and $\{\mathrm{P}^h_\theta\}_{\theta \in \Theta_R}$ imply that $\forall \theta \in \Theta$,

$$\sigma(\theta) = c_\theta \in \operatorname*{argmax}_{\{c_{\theta'}\}_{\theta' \in \Theta_R}} u_{\theta_h}.$$

Similarly, constraints $\{\text{IC}^h_{\theta\theta'}\}_{\theta<\theta'}$ and $\{\text{P}^h_\theta\}_{\theta\in\Theta_R}$ imply that $\forall\theta\in\Theta$,

$$\sigma(\theta) = c_\theta \in \underset{\{c_{\theta'}\}_{\theta\leqslant\theta'}}{\operatorname{argmax}} u_{\theta_h}.$$

Therefore, σ satisfies

$$\sigma(\theta) \in \underset{\{\sigma^{\theta'}(\theta)\}_{e_\theta}\cup\{c_{\theta'}\}_{\theta'\in\Theta_R}}{\operatorname{argmax}} u_{\theta_h},$$

which means that it is an outcome of e^*. □

Proof of Theorem 3.1 on page 77 and Theorem 3.2 on page 83: 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 Lemma A.4.4 and Lemma 3.1. 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. $\Sigma^{e^*}(\theta)\cap\{c'_\theta\}_{\theta'\in\Theta_R\setminus\theta}\neq\varnothing$ implies that $c_\theta\in\Sigma^{e^*}(\theta)$. From Lemma 3.1, 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

$$\eta(\theta) := \max\{\theta'\in\Theta_R|\theta'<\theta\}$$

the closest revealed type below a given type θ, and

$$\chi(\theta) := \min\{\theta'\in\Theta_R|\theta'>\theta\}$$

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.

$$\max_{\Theta_C \subseteq \Theta} \max_{\{(p_\theta, q_\theta)\}_{\theta \in \Theta}} \sum_{\theta \in \Theta} \mu_\theta \left(p_\theta - \kappa(q_\theta) \right) \tag{A.10}$$

$$\text{s.t. } v_{\theta_h}(q_\theta) - p_\theta \geqslant 0 \quad \forall \theta \in \Theta_R$$

$$v_{\theta_l}(q_\theta) - p_\theta \geqslant 0 \quad \forall \theta \in \Theta_C$$

$$v_{\theta_h}(q_\theta) - p_\theta \geqslant v_{\theta_h}(q_{\eta(\theta)}) - p_{\eta(\theta)} \quad \forall \theta \in \Theta \tag{A.11}$$

$$v_{\theta_h}(q_\theta) - p_\theta \geqslant v_{\theta_h}(q_{\chi(\theta)}) - p_{\chi(\theta)} \quad \forall \theta \in \Theta \tag{A.12}$$

We have the following

Lemma A.4.5. *The solution to the doubly relaxed problem satisfies R-monotonicity*

$$\theta, \theta' \in \Theta_R, \theta > \theta' \implies q_\theta \geqslant q_{\theta'} \tag{A.13}$$

and solves the relaxed problem.

Proof. Consider $\theta \in \Theta_R$, $\eta := \eta(\theta) < \theta$. Then $\theta = \chi(\eta)$ and we have

$$v_{\theta_h}(q_\theta) - p_\theta \geqslant v_{\theta_h}(q_\eta) - p_\eta$$

$$v(\eta_h, q_\eta) - p_\eta \geqslant v(\eta_h, q_\theta) - p_\theta$$

and hence

$$v_{\theta_h}(q_\theta) - v(\eta_h, q_\theta) \geqslant v_{\theta_h}(q_\eta) - v(\eta_h, q_\eta)$$

$$\int_{\eta_h}^{\theta_h} v_\theta(t, q_\theta) \mathrm{d}t \geqslant \int_{\eta_h}^{\theta_h} v_\theta(t, q_\eta) \mathrm{d}t$$

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 $\text{IC}^h_{\theta\theta'}$ 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

$$\begin{aligned}
v_{\theta_h}(q_\theta) - p_\theta &\geqslant v_{\theta_h}(q_{\eta(\theta)}) - p_{\eta(\theta)} \\
&= \left(v_{\theta_h}(q_{\eta(\theta)}) - v(\eta(\theta)_h, q_{\eta(\theta)})\right) + v(\eta(\theta)_h, q_{\eta(\theta)}) - p_{\eta(\theta)} \\
&\geqslant \left(v_{\theta_h}(q_{\eta(\theta)}) - v(\eta(\theta)_h, q_{\eta(\theta)})\right) + v(\eta(\theta)_h, q_{\theta'}) - p_{\theta'} \\
&\geqslant \left(v_{\theta_h}(q_{\theta'}) - v(\eta(\theta)_h, q_{\theta'})\right) + v(\eta(\theta)_h, q_{\theta'}) - p_{\theta'} \\
&= v_{\theta_h}(q_{\theta'}) - p_{\theta'}
\end{aligned}$$

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 A.4.6. *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 A.4.7. *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 3.4 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. △

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

$$c_{\theta^i} \sim_{\theta^i_l} \mathbf{0} \implies c_{\theta^i} \prec_{\theta^m_l} \mathbf{0}, \tag{A.14}$$

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

$$c_\eta \sim_{\theta^i_h} c_{\theta^i} \implies c_\eta \succ_{\theta^m_h} c_{\theta^i} \tag{A.15}$$

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 $c_{\eta(\underline{\theta})} \succcurlyeq_{\theta_h^m} c_{\underline{\theta}} \succ_{\theta_h^m} c_{\theta^m}$, 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 3.3 on page 84: Let $\mathbf{c} = (c_\theta)_\theta = ((p_\theta, q_\theta))_\theta$ be an optimal vector of contracts implemented by some EDP. By Theorem 3.1, 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 3.4. We proved that $q_\theta < \widehat{q_{\theta_h}}$ as Lemma A.4.7. Therefore there is only one case left to consider. 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 (A.1), note that since $q_\theta < \bar{q}_\theta$

$$\zeta_{\theta_h}(q_\theta) < \zeta_{\theta_l}(\widehat{q}_{\theta_l})$$

and consequently

$$\zeta_{\theta_h}(q_\theta) - \Delta < \zeta_{\theta_l}(\widehat{q}_{\theta_l}) - \Delta$$
$$p_\theta - \kappa(q_\theta) < \widehat{p_{\theta_l}} - \Delta - \kappa(\widehat{q_{\theta_l}})$$

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 3.4 on page 85: 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

$$\begin{aligned} \max_{(p,q)} \; & p - \kappa(q) \\ \text{s.t. } & v_{\theta_l}(q) - p \geqslant 0 \\ & v_{\theta_h}(q) - p \geqslant \Delta \end{aligned}$$

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

$$\Delta \leqslant \left[v_{\theta_h}(\widehat{q}_{\theta_l}) - v_{\theta_l}(\widehat{q}_{\theta_l}) \right]. \tag{A.16}$$

Similarly, note that the optimal contract dropping the first constraint is $\left(v_{\theta_h}(\widehat{q}_{\theta_h}) - \Delta, \widehat{q}_{\theta_h} \right)$, 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

$$\Delta \geqslant v_{\theta_h}(\widehat{q}_{\theta_h}) - v_{\theta_l}(\widehat{q}_{\theta_h}). \tag{A.17}$$

In the intermediate case, both constraints are binding,

$$v_{\theta_l}(q^\star) = p$$
$$v_{\theta_h}(q^\star) - v_{\theta_l}(q^\star) = \Delta$$

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 3.5 on page 87: 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 (A.1), we can bound those values as

$$\Pi^R_{R;\mu} \geqslant \mu_\theta \zeta_{\theta_h}(\widehat{q}_{\theta_h}) \tag{A.18}$$
$$\Pi^R_{C;\mu} \leqslant \mu_\theta \zeta_{\theta_l}(\widehat{q}_{\theta_l}) + \sum_{\theta' \neq \theta} \mu_{\theta'} \zeta_{\theta'_h}(\widehat{q}_{\theta'_h}) \leqslant \mu_\theta \zeta_{\theta_l}(\widehat{q}_{\theta_l}) + (1-\mu_\theta)\zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h}), \tag{A.19}$$

where $\tilde{\theta} := \max\{\Theta \setminus \theta\}$.

Note that Lemma A.4.3 implies that

$$\zeta_{\theta_h}(\widehat{q}_{\theta_h}) > \zeta_{\theta_h}(\widehat{q}_{\theta_l}) > \zeta_{\theta_l}(\widehat{q}_{\theta_l}), \tag{A.20}$$

and define

$$\bar{\mu}_\theta := \frac{\zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h})}{\zeta_{\theta_h}(\widehat{q}_{\theta_h}) - \zeta_{\theta_l}(\widehat{q}_{\theta_l}) + \zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h})} \in (0,1).$$

Finally, combining (A.18), (A.19) and (A.20) yields

$$\begin{aligned}
\Pi^R_{R;\mu} - \Pi^R_{C;\mu} &\geqslant \mu_\theta \zeta_{\theta_h}(\widehat{q}_{\theta_h}) - \mu_\theta \zeta_{\theta_l}(\widehat{q}_{\theta_l}) - (1-\mu_\theta)\zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h}) \\
&= \mu_\theta \left[\zeta_{\theta_h}(\widehat{q}_{\theta_h}) - \zeta_{\theta_l}(\widehat{q}_{\theta_l}) + \zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h})\right] - \zeta_{\tilde{\theta}_h}(\widehat{q}_{\tilde{\theta}_h}) \\
&\geqslant 0.
\end{aligned}$$

Therefore, for any $\mu_\theta \in [\bar{\mu}_\theta, 1]$, it is optimal to reveal θ.[15]

□

Proof of Proposition 3.6 on page 87: 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. □

Proof of Proposition 3.7 on page 87: Consider the profit from concealing all types except the highest, $\Pi_C(\epsilon) := \sum_{\theta<\bar{\theta}} \mu_\theta v_{\theta_l}(\widehat{q}_{\theta_l}) + \mu_{\bar{\theta}} v_{\bar{\theta}_h}(\widehat{q}_{\bar{\theta}_h})$, 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 3.3 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 $\{\epsilon_\theta : \theta \in \Theta \setminus \bar{\theta}\}$. □

Proof of Theorem 3.3 on page 94: Let e_0 denote an EDP constructed for sophisticated types in Theorem 3.1. 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,

$$e_{n+1} = \left(\bigcup_{i=0}^{n+1} E_i, h \right). \tag{A.21}$$

[15] This bound is typically not tight, as we introduced slack in (A.19).

To define E_i, let the most preferred alternative in e_{i-1} for type θ^i be $x_i := \text{argmax}_{\mathcal{C}(e_{i-1})} u_{\theta^i_h}$. 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 $E_i = \left\{\left(\left\{\left(\{N_i, \{d^{\theta'}_{N,i}\}_{\theta'>\theta^i:\theta'\in\Theta_S}, \mathbf{0}\}, h\right), \left(\{\widehat{c}_{\theta^i_h}, \{d^{\theta'}_i\}_{\theta'>\theta^i:\theta'\in\Theta_S}, \mathbf{0}\}, h\right), \mathbf{0}\right\}, l\right)\right\}$.

We now have to specify N_i and the decoys and verify 1-3 above. First, use the mapping from Lemma A.4.4 to set $N_i := \mathcal{D}(x, (\theta_{i-1}, \theta^i))$ so that

$$N_i \sim_{\theta^i_h} x_i \tag{A.22}$$

$$q_{N_i} \geqslant q_{x_i} \tag{A.23}$$

$$N_i \preccurlyeq_{\theta^i_l} \mathbf{0} \tag{A.24}$$

Second, define the decoys as

$$(d^{\theta'}_{N,i})_{\theta'>\theta^i:\theta'\in\Theta_S} := \mathcal{D}(N_i, (\theta^i, \Theta_S^{>\theta^i})),$$
$$(d^{\theta'}_{i})_{\theta'>\theta^i:\theta'\in\Theta_S} := \mathcal{D}(\widehat{c}_{\theta^i_h}, (\theta^i, \Theta_S^{>\theta^i})),$$

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_h^i, 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, $N_i \succcurlyeq_{\theta_h^I} d_{N,i}^{\theta'}$ and $q_{N_i} \leqslant q_{d_{N,i}^{\theta'}}$ and hence $N_i \succcurlyeq_{\theta_l^i} d_{N,i}^{\theta'}$. But $N_i \preccurlyeq_{\theta_l^i} \mathbf{0}$. As the decision problem containing $\widehat{c}_{\theta_h^i}$ also contains the outside option, continuing to this menu is part of a naive outcome. From the menu $\left(\{\widehat{c}_{\theta_h^i}, \{d_i^{\theta'}\}_{\theta' > \theta^i : \theta' \in \Theta_S}, \mathbf{0}\}, h\right)$, the DM chooses $\widehat{c}_{\theta_h^i}$ by the construction of the decoys. This establishes 3.

To see 2, note that all decoys have higher quality than N_i and $\widehat{c}_{\theta_h^i}$, respectively, and are less preferred according to θ_h^i. Hence, they are less preferred by lower naive types θ_h^j by single crossing. Furthermore, $\widehat{c}_{\theta_h^i}$ 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 $N_i \sim_{\theta_h^i} x_i$ and $q_{N_i} \geqslant q_{x_i}$ imply $N_i \preccurlyeq_{\theta_h^j} x_i$ for all $j < i$. By the induction hypothesis, $N_j \succcurlyeq_{\theta_h^j} \operatorname{argmax}_{\mathcal{C}(e_{i-1})} u_{\theta_h^j} \succcurlyeq_{\theta_h^j} x_i \succcurlyeq_{\theta_h^j} N_i$. 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 3.1. 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 (3.7) satisfies Theorem 3.3. From that, the decomposition theorem is immediate.

If there is an ex-post participation constraint, any naive outcome needs to satisfy $v(\theta) \succcurlyeq_{\theta_n} \mathbf{0}$. The revenue maximal vector of contracts satisfying these constraints is $(\widehat{c}_{\theta_n})_{\theta \in \Theta_N}$. It is immediate from the proof of Theorem 3.3 that this set of contracts can be implemented using an analogous construction.

The outside option trivially satisfies the participation constraint n every frame, so continuing at the root is always part of a valid naive strategy in the interim modifica-

tion of any extensive-form decision problem. Hence, $N^e \subseteq N^{\bar{e}}$, which establishes the observation. □

Proof of Observation 3.3 on page 95: 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 $c^s_\theta \succcurlyeq_{\theta_h} \mathbf{0} \sim \widehat{c}_{\theta_h}$ 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 Observation 3.4 on page 97: 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 3.2, 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 Observation 3.5 on page 98: First, consider the ex-post modification. To implement $(\widehat{c}_{\theta_n})_{\theta\in\Theta}$, the principal can the same construction as in Theorem 3.3, 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 $\bar{\nu}$ to $\bar{e}^*$, such that $\bar{\nu}_\theta = \widehat{c}_{\theta_h}$ for all θ. □

Bibliography

Alonso, R., and O. Câmara (2016a): "Persuading voters," *American Economic Review*, 106(11), 3590–3605.

——— (2016b): "Political Disagreement and Information in Elections," *Games and Economic Behavior*, 100, 390–412.

——— (2018): "On the value of persuasion by experts," *Journal of Economic Theory*, 174, 103–123.

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

Argenziano, R., S. Severinov, and F. Squintani (2016): "Strategic Information Acquisition and Transmission," *American Economic Journal: Microeconomics*, 8(3), 119–155.

Aumann, R. J., and M. Maschler (1966): "Game theoretic aspects of gradual disarmament," *Report of the US Arms Control and Disarmament Agency*, 80, 1–55.

——— (1995): *Repeated Games with Incomplete Information, with the Collaboration of Richard E. Stearns*. The MIT Press.

Bardhi, A., and Y. Guo (2018): "Modes of persuasion toward unanimous consent," *Theoretical Economics*, 13(3), 1111–1149.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2018): "Disclosure and Choice," *The Review of Economic Studies*, 85(3), 1471–1501.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2019): "Mechanisms With Evidence: Commitment and Robustness," *Econometrica*, 87(2), 529–566.

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

Bertomeu, J., E. Cheynel, and D. Cianciaruso (2018): "Strategic Withholding and Imprecision in Asset Measurement," *SSRN Electronic Journal*.

Best, J., and D. Quigley (2017): "Persuasion for the Long Run," Working paper.

Bester, H., and R. Strausz (2001): "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case," *Econometrica*, 69(4), 1077–1098.

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24(2), 265–272.

Blackwell, D., and M. A. Girshick (1954): *Theory of Games and Statistical Decisions*, Wiley Publications in Statistics. J. Wiley & Sons, New York.

Blume, A., O. J. Board, and K. Kawamura (2007): "Noisy talk," *Theoretical Economics*, 2(4), 395–440.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013): "Salience and Consumer Choice," *Journal of Political Economy*, 121(5), 803–843.

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

Chan, J., S. Gupta, F. Li, and Y. Wang (2019): "Pivotal persuasion," *Journal of Economic Theory*, 180, 178–202.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1451.

Deb, R., and M. Said (2015): "Dynamic screening with limited commitment," *Journal of Economic Theory*, 159, 891 – 928, Symposium Issue on Dynamic Contracts and Mechanism Design.

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

DeMarzo, P. M., I. Kremer, and A. Skrzypacz (2019): "Test Design and Minimum Standards," *American Economic Review*, 109(6), 2173–2207.

Dworczak, P., and G. Martini (2019): "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, 127(5), 1993–2048.

Dye, R. A. (1985): "Disclosure of Nonproprietary Information," *Journal of Accounting Research*, 23(1), 123–145.

Dziuda, W. (2011): "Strategic Argumentation," *Journal of Economic Theory*, 146(4), 1362–1397.

Edlin, A. S., and C. Shannon (1998): "Strict Monotonicity in Comparative Statics," *Journal of Economic Theory*, 81(1), 201–219.

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

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

Ely, J. C. (2017): "Beeps," *American Economic Review*, 107(1), 31–53.

Escudé, M. (2019): "Communication with Partially Verifiable Endogenous Information," working paper,.

Esteban, S., and E. Miyagawa (2006a): "Optimal Menu of Menus with Self-Control Preferences," *mimeo*.

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

Esteban, S., E. Miyagawa, and M. Shum (2007): "Nonlinear Pricing with Self-Control Preferences," *Journal of Economic Theory*, 135(1), 306–338.

Felgenhauer, M. (2019): "Endogenous Persuasion with Costly Verification," *The Scandinavian Journal of Economics*, 121(3), 1054–1087.

Fréchette, G., A. Lizzeri, and J. Perego (2019): "Rules and Commitment in Communication: An Experimental Analysis," Working paper.

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

Gentzkow, M., and E. Kamenica (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review: Papers & Proceedings*, 106(5), 597–601.

——— (2017): "Disclosure of Endogenous Information," *Economic Theory Bulletin*, 5(1), 47–56.

Glazer, J., and A. Rubinstein (2004): "On Optimal Rules of Persuasion," *Econometrica*, 72(6), 1715–1736.

——— (2006): "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach," *Theoretical Economics*, 1(4), 395–410.

——— (2012): "A Model of Persuasion with Boundedly Rational Agents," *Journal of Political Economy*, 120(6), 1057–1082.

——— (2014): "Complex Questionnaires," *Econometrica*, 82(4), 1529–1541.

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

Green, J. R., and N. L. Stokey (2007): "A two-person game of information transmission," *Journal of Economic Theory*, 135(1), 90–104.

Grossman, S. J. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law & Economics*, 24, 461–484.

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

Guo, Y., and E. Shmaya (2019a): "Costly Miscalibration," Working paper.

——— (2019b): "The interval structure of optimal disclosure," *Econometrica*, 87(2), 653–675.

Hagenbach, J., F. Koessler, and E. Perez-Richet (2014): "Certifiable Pre-Play Communication: Full Disclosure," *Econometrica*, 82(3), 1093–1131.

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

Hart, S., I. Kremer, and M. Perry (2017): "Evidence Games: Truth and Commitment," *American Economic Review*, 107(3), 690–713.

Hedlund, J. (2017): "Bayesian persuasion by a privately informed sender," *Journal of Economic Theory*, 167, 229–268.

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

——— (2017): "Naivete-Based Discrimination," *The Quarterly Journal of Economics*, pp. 1019–1054.

Herrero, M. J., and S. Srivastava (1992): "Implementation via Backward Induction," *Journal of Economic Theory*, 56(1), 70–88.

Ichihashi, S. (2019): "Limiting Sender's information in Bayesian persuasion," *Games and Economic Behavior*, 117, 276–288.

Johnson, E. J., J. Hershey, J. Meszaros, and H. Kunreuther (1993): "Framing, Probability Distortions, and Insurance Decisions," *Journal of Risk and Uncertainty*, 7, 35–51.

Jovanovic, B. (1982): "Truthful Disclosure of Information," *The Bell Journal of Economics*, 13(1), 36–44.

Jung, W.-O., and Y. K. Kwon (1988): "Disclosure When the Market Is Unsure of Information Endowment of Managers," *Journal of Accounting Research*, 26(1), 146–153.

Kamenica, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11(1).

Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

Kartik, N. (2009): "Strategic Communication with Lying Costs," *Review of Economic Studies*, 76(4), 1359–1395.

Kartik, N., F. X. Lee, and W. Suen (2017): "Investment in Concealable Information by Biased Experts," *The RAND Journal of Economics*, 48(1), 24–43.

Kolotilin, A. (2018): "Optimal Information Disclosure: A Linear Programming Approach," *Theoretical Economics*, 13(2), 607–635.

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017): "Persuasion of a Privately Informed Receiver," *Econometrica*, 85(6), 1949–1964.

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

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

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

Leshno, M., H. Levy, and Y. Spector (1997): "A Comment on Rothschild and Stiglitz's "Increasing Risk: I. A Definition"," *Journal of Economic Theory*, 77(1), 223–228.

Lipnowski, E., and D. Ravid (2019): "Cheap Talk with Transparent Motives," Working Paper, Economics Department, University of Chicago.

Liu, Q., K. Mierendorff, X. Shi, and W. Zhong (2019): "Auctions with limited commitment," *American Economic Review*, 109(3), 876–910.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003): "Projection Bias in Predicting Future Utility," *The Quarterly Journal of Economics*, 118(4), 1209–1248.

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

Mathevet, L., D. Pearce, and E. Stacchetti (2019): "Reputation and Information Design," *Working paper*.

Matthews, S., and A. Postlewaite (1985): "Quality Testing and Disclosure," *RAND Journal of Economics*, 16(3), 328–340.

Milgrom, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics*, 12(2), 380–391.

——— (2008): "What the Seller Won't Tell You: Persuasion and Disclosure in Markets," *Journal of Economic Perspectives*, 22(2), 115–131.

Milgrom, P., and J. Roberts (1986): "Relying on the Information of Interested Parties," *The RAND Journal of Economics*, pp. 18–32.

Min, D. (2018): "Bayesian Persuasion under Partial Commitment," Working paper.

Moser, C., and P. Olea de Souza e Silva (2019): "Optimal Paternalistic Savings Policies," working paper.

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18(2), 301–317.

Myerson, R. B. (1983): "Mechanism design by an informed principal," *Econometrica: Journal of the Econometric Society*, pp. 1767–1797.

Nguyen, A., and T. Y. Tan (2019): "Bayesian Persuasion with Costly Messages," *SSRN Electronic Journal*.

North, A. C., D. J. Hargreaves, and J. McKendrick (1997): "In-Store Music Affects Product Choice," *Nature*, 390(6656), 132–132.

North, A. C., L. P. Sheridan, and C. S. Areni (2016): "Music Congruity Effects on Product Memory, Perception, and Choice," *Journal of Retailing*, 92(1), 83–95.

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

Okuno-Fujiwara, M., A. Postlewaite, and K. Suzumura (1990): "Strategic Information Revelation," *The Review of Economic Studies*, 57(1), 25–47.

Pei, H. D. (2015): "Communication with Endogenous Information Acquisition," *Journal of Economic Theory*, 160, 132–149.

Perez-Richet, E. (2014): "Interim bayesian persuasion: First steps," *American Economic Review*, 104(5), 469–474.

Perez-Richet, E., and V. Skreta (2018): "Test Design under Falsification," Working paper.

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

Rappoport, D. (2017): "Evidence and Skepticism in Verifiable Disclosure Games," working paper,, Working paper.

Rothschild, M., and J. E. Stiglitz (1970): "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2(3), 225–243.

Salamanca, A. (2019): "The Value of Mediated Communication," Working paper.

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

Salant, Y., and R. Siegel (2018): "Contracts with Framing," *American Economic Journal: Microeconomics*, 10(3), 315–346.

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

Schwitzgebel, E., and F. Cushman (2015): "Philosophers' Biased Judgments Persist despite Training, Expertise and Reflection," *Cognition*, 141, 127–137.

Sher, I. (2011): "Credibility and Determinism in a Game of Persuasion," *Games and Economic Behavior*, 71(2), 409–419.

Shin, H. S. (1994): "The Burden of Proof in a Game of Persuasion," *Journal of Economic Theory*, 64(1), 253–264.

Skreta, V. (2006): "Sequentially Optimal Mechanisms," *Review of Economic Studies*, 73(4), 1085–1111.

Spiegler, R. (2011): *Bounded rationality and industrial organization*. Oxford University Press.

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

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

Topkis, D. M. (1998): *Supermodularity and Complementarity*, Frontiers of Economic Research. Princeton University Press, Princeton, NJ.

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

Verrecchia, R. E. (1983): "Discretionary Disclosure," *Journal of Accounting and Economics*, 5, 179–194.

Yu, P. C. (2018): "Optimal Retirement Policies with Present-Biased Agents," working paper,.

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