The Secret Behind *The Tortoise and the Hare*: Information Design in Contests*

Job Market Paper

Alejandro Melo Ponce†

This version—November, 2018
Click here for current version

Abstract

I analyze the optimal information disclosure problem under commitment of a "contest designer" in a class of binary action contests with incomplete information about the abilities of the players. If the contest designer wants to incentivize the players to play in equilibrium a particular strategy profile, she can design an information disclosure rule, formally a stochastic communication mechanism, to which she will commit and then use to "talk" with the players. The main tool to carry out the analysis is the concept of Bayes Correlated Equilibrium recently introduced in the literature. I find that the optimal information disclosure rules involves private information revelation (manipulation), which is also cost-effective for the designer. Furthermore, the optimal disclosure rule involves asymmetric and in most cases correlated signals that convey only partial information about the abilities of the players.

JEL classification: C72, C79, D44, D82, D83.
Keywords: information design, contests, implementation, incomplete information, Bayes Correlated Equilibrium.

1 Introduction

Have you ever wondered why the hare took a nap? Everybody knows well the fable.[1] But how many of us ever think about who organized such a curious contest. Imagine there was a fox behind it. He decided the where, when and who of the competition. He could also speak with the contestants before the big hour. Maybe he told the tortoise not to give up, the hare is the fastest but weird things could

*This paper was previously presented at various places with the title *Information Design in Contests*

†Department of Economics, Stony Brook University. E-mail: alejandro.meloponce@stonybrook.edu. I am profoundly grateful to my advisor, Pradeep Dubey, for his continuous and invaluable help and encouragement with this project. I also wish to thank Ting Liu, Yair Tauman, Sandro Brusco, Vasiliki Skreta, Marcos Fernandes, Laura Karpuska, Camilo Rubbini, Yijiao Liu, David Ruiz G., Michael Kramm, and seminar audiences at the 28th & 29th Stony Brook International Conferences on Game Theory, the 2017 Econometric Society Summer School, the 2018 Midwest Economic Association Annual Meeting and the 2018 Econometric Society Australasian Meeting for helpful comments and discussions. Special thanks to Brenda Cuellar Marines for her invaluable feedback and support. All remaining errors are my own.

[1] Aesop, *"The Tortoise and the Hare"*, fable 226 in the Perry Index.

happen. Talking to the hare, he may drop that it takes half a day for the tortoise to walk a similar distance. We all know the end of the story: the tortoise won and the fox made a fortune betting against the hare. Let us take a look at how the fox did the magic.

The trick lies in the position of the fox. An agent with privileged information can use it in its favor. We have a lot of examples from daily life. A pollster could influence an election by sharing some information with the parties running in the election. Imagine there are two candidates and candidate A has a slight advantage over candidate B. The pollster can convince candidate B that in the last poll there was a technical draw. So the campaign strategy must adapt if they want to get more votes and be the winners. The pollster can share the information with candidate A or not. If the pollster shares, the candidate can adapt immediately so he can compete in the new scenario. But if the candidate does not get the information, adaption will take time and he could loose valuable votes.

Many examples abound where there is a third party who is in a position to alter the final results of a "competition" by modifying, in first instance, the behavior of the participants. The key is the tools that she posses and the way she uses them. For example, the third party could alter monetary rewards. But instead, she can do the magic by *manipulating the information* that the players observe, without increasing her costs. Without further ado, let us turn to the matter in hand.

I present a model that illustrates how information manipulation can be used to alter the results of a contest. The model is deliberately simple so as to derive its conclusions with minimal fuss while at the same time providing a rich ground for its analysis.

Consider a principal who intends to organize a contest between two players. The principal, which we identify as *she*, henceforth will be referred to as "the contest designer". In the contest there is a prize that is to be awarded to the player, each of which we identify as *he*, with the highest output. The output from each player is determined by his innate ability and the effort that he undertakes. Each player is presumed to always know his ability. The contest designer acts as a third-party to the contest between the players in the sense that she does not participate in the strategic interaction between the players. We assume that the designer cannot manipulate the structure of the contest but can only manipulate the information that the players observe by communicating it to them.

The contest, whose basic structure follows the one in Dubey (2013), belongs to a class of games with binary actions, namely to shirk or work, and with incomplete information about the abilities of the players, which are also assumed to be binary, either weak or strong. The class of contests is parameterized by the value of a common prize, the cost of exerting effort, the private first-order beliefs that the players hold about their rival's ability and the value to the designer, in terms of output produced, of the effort profile chosen by the players.

The designer wants to manipulate the beliefs of the players so that they play in equilibrium a particular effort profile. In order to carry out this manipulation, the contest designer can design an *information disclosure rule*, which formally is a stochastic communication mechanism, to which she will *commit* and then use to "talk" to the players. This structure endows the contest designer with more commitment power in the sense that it will allow her to commit to send any distribution of messages that she desires as a function of the realized ability vector *before learning it*. This communication structure will become publicly known to the players. After this stage, the agents will observe the realized messages according to the information disclosure rule and will update their beliefs about the ability of their rival, i.e. their *first-order beliefs*, their beliefs about the first-order beliefs of their rival, i.e. their *second-order beliefs*, and so on.

The main tool to carry out the analysis and search for the optimal information disclosure rule follows the general methodology introduced in Taneva (2015). The goal of the designer is to design an information disclosure rule for each contest which will induce an effort profile as a Bayes Nash Equilibrium (BNE) with the property that such profile maximizes the designer's objective in expectation. Under this view, the designer would have to first characterize the set of all BNE *under all* possible information structures. The novelty here is that in contrast to Taneva (2015), we have an environment in which the players already possess private information about their abilities, so the set of information structures that we consider need to respect this restriction. Although performing this characterization program seems like an insurmountable task, we follow Taneva's approach in which we appeal to the notion of Bayes Correlated Equilibrium, introduced by Bergemann and Morris (2016), which will allow us to characterize the set of all Bayes Nash Equilibria associated with each contest and with the prior private information of the players. Given the assumption that the players already have some prior private information about the ability of their rival, we need to use the full power of Bergemann and Morris's result[2] It turns out that in our contest environment in which the players are *informed of their own ability*, it comes at almost no cost extending the characterization of Taneva since the players have *distributed knowledge*[3] of the ability vector. Thus, when the designer learns the true ability vector before communicating with the players, she does not know anything more that the players already don't know when they pool their information together.

We find that optimal information disclosure rules in contests involve private information revelation (manipulation). The optimal disclosure rule involves asymmetric and, in a robust set of parameters, correlated messages to each player. The messages involved in the optimal information disclosure rule convey only partial information about the abilities of the players. The intuition behind this result can be most clearly understood in the two player contest environment that we describe in this paper. When both players have similar abilities, in an *ex post* assessment of the contest the players find that the competition is evenly poised and each player will find it worthwhile to put effort since they have an equal shot at obtaining the prize. On the other hand, when the players have disparate abilities, an ex post assessment of the contest would lead the players to shirk with high probability, particularly for the strong player. Thus if it were possible, it would be in the interest of the designer to fully and publicly inform the players when they are similar and tell them nothing when they are different. However, no information disclosure rule can implement the previous state of affairs, since giving full and public information when the players are similar immediately makes it common knowledge not

[2]Theorem 1 in Bergemann and Morris (2016) which provides an epistemic relationship between the set of Bayes Correlated Equilibria under some initial information structure and the set of Bayes Nash Equilibria under all information structures that *expand* the first one.

[3]The two players, by pooling their knowledge together, can deduce the full ability vector. See Fagin, Halpern, Moses, and Vardi (2004, p. 23). Bergemann and Morris (2013) call this property *distributed certainty* in a language that is more standard in the game theory literature.

only this fact but also the corresponding event when the players hear nothing when they are different. Therefore, it will be in the interest of the designer to only partially reveal information in such a way that the true ability vector never becomes common knowledge. This is one of the reasons that makes private information necessary. It turns out that it is the beliefs of the weak player that drive the incentive to put effort in the contest, since it is his actions and beliefs when competing against a strong player that actually motivate the strong to work. Hence, the optimal revelation scheme alters not only the first-order beliefs of the players but also the higher-order hierarchies in a non-trivial way. Precisely for the previous reasons, public revelation of information is not optimal, since it generates symmetric hierarchies of beliefs even when the players are different in their abilities. In particular, for a robust set of parameters, the optimal information scheme follows the rule of *informing the underdog*: at some messages, the weak player will become more informed, with respect to his prior private information, about the ability vector; whereas at other messages, the weak player will become fully informed of the ability vector; however, *neither these two facts will be common knowledge* between the two players.

The explicit characterization of the optimal information disclosure rule for every contest allows us to perform an important comparative statics analysis. Suppose that the prize the designer awards consists of a fraction of the total output produced by the contestants. In this scenario, the designer can engage in information design while at the same time altering the value of the prize. While the "revenue" side of information design is intuitively well understood, we are know attaching a "cost" to actually carrying out the information manipulation. We find necessary and sufficient conditions on the parameters of the game to ensure that a private, asymmetric and partial information revelation scheme is optimal for the designer and I also provide conditions for when it is the case that giving no information is optimal. The main message is that we find that under a robust set of parameters for which manipulating information while giving a relatively small prize is *doubly optimal*: it does not only provide incentives for the players to work but it also does it in the most cost-efficient way possible.

The rest of the section is devoted to a survey of the related literature.

1.1 Literature Review

This paper belongs to a very recent and active literature on information design. This strand of the literature about communication in games is different from the cheap talk literature as established by Crawford and Sobel (1982) because the assumption that the information designer can *credibly commit* to an information transmission strategy before learning the true state of the world. The one-agent version of the problem has been extensively studied in the literature since the seminal contribution of Kamenica and Gentzkow (2011) on Bayesian Persuasion, which is preceded by the works of Aumann, Maschler, and Stearns (1995), Brocas and Carrillo (2007) and Benoît and Dubra (2011). Kamenica and Gentzkow provide a characterization of the optimal information design problem for the case of a single sender and receiver using techniques from convex analysis. Their elegant results allows for a clear characterization of optimality of an information structure in the single-agent case. Since then, the one agent-version has been the subject of a productive effort by many authors in many different areas and applications (e.g. Gentzkow and Kamenica (2014); Ely, Frankel, and Kamenica (2015); Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017); Lipnowski and Mathevet (2018), to name a few).

On the other hand, the theory of information design in games is still at an early stage. Nevertheless, optimal solutions have been derived in specific environments, for example as in Vives (1988); Morris and Shin (2002) and Angeletos and Pavan (2007). The most closely related papers to this one in terms of techniques are Taneva (2015) and Bergemann and Morris (2016), as both provide the systematic approach, based on a revelation-principle style methodology, to approach information design problems. As mentioned before, the method is based on the notion of Bayes Correlated Equilibrium that characterizes all Bayes Nash Equilibrium outcomes under all possible information structures. We take advantage of this formulation to simplify and fully characterize the information design problem in contests discussed in this paper. Another closely related paper is Mathevet, Perego, and Taneva (2016) in which they push forward the theory to provide a characterization of the solution to the information design problem in terms of belief-hierarchy distribution instead of information structures. This allows them to provide an expression of the optimal solution in games in terms of an optimal private and public component, where the later part comes from concavification, effectively extending Kamenica and Gentzkow's characterization. Their results assume no prior private information from the players. In particular, they discuss some qualitative properties of belief-hierarchy distributions and information structures which we adapt for our problem of a contest with prior private information from the players.

For a recent an in-depth survey of the literature on information design, the reader is encouraged to consult Bergemann and Morris (2018) and Kamenica (2018).

The literature on asymmetric information and information disclosure in contests motivates this paper. These issues have been the focus of recent work (Fey, 2008; Lim and Matros, 2009; Münster, 2009; Morath and Münster, 2013; Epstein and Mealem, 2013; Gürtler, Münster, and Nieken, 2013; Fu, Gürtler, and Münster, 2013; Dubey, 2013; Denter, Morgan, and Sisak, 2014; Fu, Lu, and Zhang, 2016; Einy, Moreno, and Shitovitz, 2017). Some of these papers study issues related to how an agent should disclose information about his private attributes. Other papers in which a designer is present, study how a designer should disclose "performance evaluations" or study disclosure policies in a one-sided asymmetric information environment. In particular, Lim and Matros (2009) and Fu et al. (2013, 2016) consider the problem of how to reveal information about contestants entries when these are stochastic. Denter et al. (2014) analyze the incentives of a privately informed contestant to disclose his information to his opponent and the incentives for transparency of the designer. Dubey (2013), from which we take the basic environment, analyzes the impact of null versus complete information in a contest in terms of expected output from the players. These last strand of articles focus on comparing the cases of no disclosure versus full disclosure. In this paper, we extend the discussion towards analyzing partial information disclosure rules and contribute towards a classification of their qualitative richness and their impact in manipulating the behavior of the players. We also find that focusing only on no disclosure versus full disclosure *is with loss of generality*, since the optimal information disclosure rules require in general partial revelation of information.

A closely related paper is Zhang and Zhou (2016), in which they analyze the Bayesian persuasion problem in a Tullock contest with one-sided asymmetric information. They assume that one contestant has imperfect information about the cost function of his opponent while the other one is perfectly informed. The contest designer decides how to disclose this information to the imperfectly informed contestant using general disclosure rules with can span the whole spectrum between null disclosure to full disclosure. However, in their analysis, they focus on public disclosure rules, which buys them a great deal of technical convenience, since they can apply the insights from Kamenica and Gentzkow to find an optimal solution in their problem. Compared to Zhang and Zhou, our paper extends the analysis of optimal information disclosure in contests in two directions: we assume all contestants to have incomplete information about while holding private information about the a payoff relevant parameter of the contest, and we allow for fully general information disclosure rules, since the distinction between public and private information and full versus partial disclosure becomes crucial in contests.

Finally, another very closed related paper is Kramm (2018), who focuses on a multi-task Tullock contest in which there is incomplete information from all players about how the success in the contest will depend on the effort mixture put on different tasks. Kramm also considers a methodology inspired by Bergemann and Morris (2016) to solve for the optimal information policy. A feature in Kramm's contest environment is that the players do not hold any prior private information. Nevertheless he derives similar results to ours: he also finds that there is an important distinction between private and public information and that in order for the information disclosure policy to benefit the designer, private information provides the right informational advantage for the players to behave in the way that the designer intends. In his environment, he finds that optimal information disclosure involves sometimes disclosing information to a weak player in a particular task while in other scenarios it is optimal to inform only contestants who are strong in a particular task. Compare this to our results that say that in general both players should be partially informed while in some cases the weak player becomes fully informed. However, the nature of the information disclosure rule in this paper is such that the event that a player becomes fully informed does not become common knowledge when it happens. Also the information disclosure policy sometimes leaves first-order beliefs untouched while operating on the second and higher-order hierarchy.

The rest of the paper follows the following structure. In section 2 we present the description of the model: the contest designer and the players; the role of the designer in manipulating information; and the extended game that is induced by the designer's choice of information disclosure rule. Section 3 describes the role of the Bayes Correlated Equilibrium notion in simplifying the optimal information disclosure problem in the contest. Section 4 presents the characterization of the equilibrium behavior for two particular information disclosure rules; these results will be ancillary to establishing and comparing the main results of the paper. Section 5 presents the main results of the paper, namely the full characterization of the optimal information disclosure rules for the family of contest that we are considering and the cost-benefit analysis of designing information in the contest. Section 6 concludes and describes some extensions and avenues for future research.

2 Model

A principal, which we will refer from now on as the contest designer, intends to run a contest between two players. The contest designer acts as an external agent to the contest between the players and her only role is to disclose information to the players but other than that she does not participate in the strategic interaction. In particular, we assume that the contest designer *can only* manipulate information, but not the structure of the contest.

The contest between the two players follows the structure of the basic model in Dubey (2013), from which we take the particular specification of the contest.

The two players are assumed to be risk neutral and ex-ante symmetric. Each player $i \in \{1, 2\}$ can have one of two abilities, $a_i \in \{\alpha, \beta\} = A_i \subset \mathbb{R}$, where $\alpha < \beta$. Thus, we identify α with a *weak player* and β with a *strong player*. Each player can choose from two effort levels $e_i \in \{0, 1\} = \{\mathit{Shirk}, \mathit{Work}\} = E_i$. As usual, $a \in A = A_1 \times A_2$ will denote the ability vector and $e \in E = E_1 \times E_2$ the effort vector.

Each player i, given his ability and effort chosen, produces *output* according to the production function $f : E_i \times A_i \to \mathbb{R}$ given by

$$f(e_i, a_i) = \begin{cases} a_i & \text{if } e_i = 0 \\ k(a_i)a_i & \text{if } e_i = 1 \end{cases}$$

where $k : A_i \to \mathbb{R}$ is a function such such that $k(a_i) > 1$ for every $a_i \in A_i$ and every i. For both players, the marginal cost of putting effort is $\varkappa > 0$ and they both put a *common value* $\upsilon > 0$ on a *prize*, which is awarded in full to the player with the higher output and randomized equally in case of a tie. Thus, each player i has a payoff function $u_i : E \times A \to \mathbb{R}$ given by

$$u_i(e, a) = \begin{cases} \upsilon - \varkappa e_i & \text{if } f(e_i, a_i) > f(e_j, a_j) \\ \frac{\upsilon}{2} - \varkappa e_i & \text{if } f(e_i, a_i) = f(e_j, a_j) \\ -\varkappa e_i & \text{if } f(e_i, a_i) < f(e_j, a_j) \end{cases} \tag{2.1}$$

For reasons that will become clear later, it will be convenient to perform a normalization of the payoffs in (2.1). Formally, let $F : \mathbb{R}_+ \to \mathbb{R}_+$, be the positive linear transformation given by $f(u) = u/\upsilon$. Let $\phi = \varkappa/\upsilon > 0$ denote the *normalized marginal cost of working*. The normalized payoffs are given by

$$\hat{u}_i(e,a) = f(u_i(e,a)) = \begin{cases} 1-\phi e_i & \text{if } f(e_i,a_i) > f(e_j,a_j), \\ \frac{1}{2}-\phi e_i & \text{if } f(e_i,a_i) = f(e_j,a_j), \\ -\phi e_i & \text{if } f(e_i,a_i) < f(e_j,a_j). \end{cases} \tag{2.2}$$

We will impose some assumptions on the parameters of the model to make the analysis tractable and comparable with the results in Dubey (2013).

A1. *(Minimum valuation).* $\phi < 1$, or equivalently $\varkappa < \upsilon$. This assumption precludes that the contest does not become trivial by making shirking strictly dominant under any information structure. This assumption enables us to focus on the failure to work caused by strategic competition. Notice that when $\phi < 1/2$ or equivalently $2\varkappa < \upsilon$, the prize is large enough to guarantee even if it is split equally in the case of a tie, the players still find it worthwile to put effort.

A2. *(Monotonicity of Output).* $k(a_i)a_i$ is strictly monotonic for each $a_i \in A_i$ and every i.

A3. *(Ordering of Output).* $\beta < k(\alpha)\alpha$. This assumption is concerned with the ordering of $k(\alpha)\alpha$ and β, i.e. the output when the *weak* player works and the one from the *strong* player when he shirks respectively. Intuitively, this assumption says that if a weak player works, then he can beat a strong player who shirks.

Let $K \subset \mathbb{R}^4_{++}$ denote the set of productivities that satisfy assumptions A.2 and A.3,

$$K = \left\{(\alpha, \beta, k(\alpha)\alpha, k(\beta)\beta) \in \mathbb{R}^4_{++} | \alpha < \beta < k(\alpha)\alpha < k(\beta)\beta\right\}.$$

An arbitrary 4-tuple from this set will be denoted by k.

We will assume that the players are privately and independently informed of their own ability and that their beliefs[4] about their rival's ability after being informed is constant across abilities, strictly positive and symmetric between players. These restrictions imply the existence of a symmetric, statistically independent *common prior* from which the posterior beliefs are derived. We can collect the previous observations into the following assumption.

A4. *(Common prior & Constant beliefs).* For each player $i = 1,2$, the posterior beliefs about the vector of abilities $a \in A$ are symmetric between players and constant across abilities, i.e.

$$prob(\alpha|\alpha) = prob(\alpha|\beta) = \psi \in (0,1).$$

These beliefs are induced by a common prior P_ψ that satisfies statistical independence, as follows:

$$\mathsf{P}_\psi(\alpha,\alpha) = \psi^2,\ \mathsf{P}_\psi(\alpha,\beta) = \mathsf{P}_\psi(\beta,\alpha) = \psi(1-\psi),\ \mathsf{P}_\psi(\beta,\beta) = (1-\psi)^2,$$

therefore[5] $\mathsf{P}_\psi \in \mathsf{int}\big(\Delta(A)\big)$. Abusing notation, we will identify P_ψ with ψ.

[4] Formally, these are the first order beliefs of the players about the state $a \in A$

[5] For any set X, $\Delta(X)$ denotes the set of probability measures.

The Contest designer's preferences over states and outcomes can also be represented by a utility function $v : E \times A \to \mathbb{R}$. In this paper, we make the assumption the designer has additive preferences over the output of the players, specifically, he cares about total output:

$$v(e,a) = \sum_{i \in \{1,2\}} f(e_i, a_i).$$

We also assume that the designer shares the prior ψ with the players.

The collection $C_{\psi,\phi} = ((A_i, E_i, \hat{u}_i)_{i=1,2}, \psi)$ defines a contest for each (ψ, ϕ). Assumptions A.1–A.4 constrain the set of possible values of (ψ, ϕ) to the open rectangle $(0,1) \times (0,1) \subset \mathbb{R}^2$. Any contest corresponds to a point in this rectangle, so let $C = (0,1)^2$ and by an abuse of notation we identify each $C_{\psi,\phi}$ with the point $(\psi, \phi) \in C$. Each (ψ, ϕ) has four *payoff relevant states*, which can be represented by four bimatrix games. Under assumptions A.1–A.4, the game (ψ, ϕ) is as in Figure 1 where we use the normalized payoffs as defined in equation (2.2).

We define a *contest environment* to be the set of all contests $(\psi, \phi) \in C$ as parameterized by $(\psi, \phi, k) \in C \times K$. All of our analysis will be confined to this set of contest environments.

$\alpha\alpha$	W	S
W	$\frac{1}{2}-\phi, \frac{1}{2}-\phi$	$1-\phi, 0$
S	$0, 1-\phi$	$\frac{1}{2}, \frac{1}{2}$

Prob. ψ^2

$\alpha\beta$	W	S
W	$-\phi, 1-\phi$	$1-\phi, 0$
S	$0, 1-\phi$	$0, 1$

Prob. $\psi(1-\psi)$

$\beta\alpha$	W	S
W	$1-\phi, -\phi$	$1-\phi, 0$
S	$0, 1-\phi$	$1, 0$

Prob. $\psi(1-\psi)$

$\beta\beta$	W	S
W	$\frac{1}{2}-\phi, \frac{1}{2}-\phi$	$1-\phi, 0$
S	$0, 1-\phi$	$\frac{1}{2}, \frac{1}{2}$

Prob. $(1-\psi)^2$

Figure 1: The contest (ψ, ϕ) under assumptions A.1–A.4: normalized payoffs.

2.1 Information disclosure rules

The desiger has the ability to manipulate information and moreover to bring new information of his own to the table. The designer, with her knowledge of the priorψ and before learning the state $a \in A$, *commits to an information disclosure rule.*

Definition 2.1. *An* ***information disclosure*** *rule is a set of finite messages M_i, one for each player, and a family of conditional probability distributions $\pi : A \to \Delta(M)$, where $M = M_1 \times M_2$.*

For each $a \in A$, the notation $\pi(\cdot|a)$ denotes the probability distribution over M conditional on the state a. Since we are assuming that M is finite, we can represent π as a family of vectors:

$$\big((\pi(m|a))_{m \in M}\big)_{a \in A}.$$

We denote the information disclosure rule by $\mathcal{D} = ((M_i)_{i=1,2}, \pi)$.

The next example illustrates three particular disclosure rules.

Example 2.1. The *null informatation rule*, denoted $\mathcal{N}$, provides now new information to the players. Formally, $\mathcal{N} = ((M_i)_{i=1,2}, \pi)$ where $M_i = \{\varnothing\}$ for each $i = 1, 2$ and $\pi(\cdot|a) = [(\varnothing, \varnothing)]$ for[6] all $a \in A$.

The *complete information rule*, denotes $\mathcal{C}$, publicly reveals the private information of the players, i.e. fully reveals their abilities. Formally, $\mathcal{C} = ((M_i)_{i=1,2}, \pi)$, $M_i = \{\mathsf{S}, \mathsf{D}\}$, $i = 1, 2$, $\pi(\cdot|a) = [(\mathsf{S}, \mathsf{S})]$ when $a \in \{\alpha, \beta\}$ and $\pi(\cdot|a) = [(\mathsf{D}, \mathsf{D})]$ when $a \in \{\alpha\beta, \beta\alpha\}$.

A middle ground between the previous two is the *p-q disclosure rule*, denoted by $\mathcal{D}_{p,q}$. Formally, $\mathcal{D}_{p,q} = ((M_i)_{i=1,2}, \pi)$, $M_i = \{\mathsf{S}, \mathsf{D}, \varnothing\}$, $i = 1, 2$ and $\pi(\cdot|a) = p[(\mathsf{S}, \mathsf{S})] + (1 - p)[(\varnothing, \varnothing)]$ when $a \in \{\alpha, \beta\}$ and $\pi(\cdot|a) = q[(\mathsf{D}, \mathsf{D})] + (1 - q)[(\varnothing, \varnothing)]$ when $a \in \{\alpha\beta, \beta\alpha\}$. Intuitively, this rule behaves as if the designer had two weighted coins. The designer uses the coin with weight p in the event that the players are similar, in which he publicly reveals the state with probability p and says nothing with probability $1 - p$; and uses the coin with weight q in the event that the players are different, in which he publicly reveals the state with probability q and says nothing with probability $1 - q$. ◀

After choosing $\mathcal{D}$, its structure is publicly announced to the players, i.e. it is made common knowledge. This means that the players will become aware of how the designer will communicate with them in terms of messages and the probability of hearing a particular message. After the information disclosure rule $\mathcal{D}$ becomes common knowledge, the designer learns $a \in A$ and sends privately a message to each player according to $\pi : A \to \Delta(M)$. Notice that the disclosure rule $\mathcal{D}$ induces a Bayesian Game over each $(\psi, \phi) \in C$, which we denote $\Gamma_{\mathcal{D},\psi,\phi}$.

In the game $\Gamma_{\mathcal{D},\psi,\phi}$, players will choose a *behavioral stragegy* $\sigma_i : A_i \times M_i \to \Delta(E_i)$. The following definition[7] is standard.

Definition 2.2. *Let $\sigma = (\sigma_i)_{i=1,2}$ be a profile of behavioral strategies in $\Gamma_{\mathcal{D},\psi,\phi}$. The profile σ is a* ***Bayes Nash Equilibrium*** *(BNE) if for every player $i \in \{1, 2\}$, for every $a_i \in A_i$, for every $m_i \in M_i$ we have that*

$$\operatorname{supp}\big(\sigma_i(\cdot|a_i, m_i)\big) \subseteq \arg\max_{e_i' \in E_i} \left\{ \sum_{e_j, m_j, a_j} prob(a_j|a_i)\pi(m_i, m_j|a)\sigma_j(e_j|a_j, m_j)\hat{u}_i(e_i', e_j, a) \right\}.$$

Let $\mathfrak{E}(\mathcal{D}, \psi, \phi)$ denote the set of Bayes Nash Equilibria of $\Gamma_{\mathcal{D},\psi,\phi}$.

Depending on the context, we sometimes suppress the dependence on (ψ, ϕ) on the induced game $\Gamma_{\mathcal{D}}$. and its equilibrium set $\mathfrak{E}(\mathcal{D})$.

[6]The notation $[x]$ denotes the probability measure that puts probability one on the point x.

[7]For a probability measure $p \in \Delta(X)$, where X is a discrete space, the support p is the set of points with strictly positive probability, i.e. $\operatorname{supp}(p) = \{x \in X | p(x) > 0\}$.

2.2 The contest design (information) problem

The expected payoff for the designer of information disclosure rule $\mathcal{D}$ and any strategy profile σ from the players is

$$V(\mathcal{D},\sigma;\psi,\phi,k) = \sum_{a,e} \mathsf{P}(a)\left(\sum_{m\in M}\pi(m|a)\prod_{i=1,2}\sigma_i(e_i|a_i,m_i)\right)v(e,a). \tag{2.3}$$

The information design problem for the contest designer is given by

$$\bar{V}(\psi,\phi,k) = \max_{\mathcal{D}} \max_{\sigma\in\mathfrak{E}(\psi,\phi,\mathcal{D})} V(\mathcal{D},\sigma;\psi,\phi,k). \tag{2.4}$$

Notice that in problem (2.4) we are assuming a selection criteria for equilibria[8] which is the one that benefits the designer, in case there is multiplicity of equilibria in $\Gamma_{\mathcal{D},\psi,\phi}$.

If we can find an optimal information disclosure rule $\mathcal{D}^*$ that solves problem (2.4), it is not guaranteed that such a rule will induce a unique equilibrium in the extended game. The only thing that we can guarantee is that in the game induced by $\mathcal{D}^*$ there will be an equilibrium σ^* that will achieve the best effort profile for the designer. However, it may be the case that the optimal information disclosure rule also induces other equilibria different from σ^*. Thus, an important question is what equilibrium effort profiles such equilibria engender and how these compare to the ones generated by the profile σ^*. In order to answer this question we will need to characterize the whole equilibrium set $\mathfrak{E}(\mathcal{D}^*,\psi,\phi)$ of the extended game generated by the optimal information disclosure rule.

The next subsection introduces adequate notation and some definitions that will help us to characterize the equilibrium set.

2.3 The game and beliefs induced by an information disclosure rule

2.3.1 The induced game and the extended type space

Recall that, as described in subsection 2.1, once the information disclosure rule is chosen by the designer, he *commits* to it, in the sense that its probabilistic structure is disclosed publicly to the players, i.e. it is made *common knowledge*. Afterwards, the information disclosure rule is used to create the messages than then will be communicated privately and truthfully to the players.

The previous discussion implies that we can think about the prior information that the players already posses, i.e. knowledge of their own abilities, together with the message that they hear from the designer as their *type* in the incomplete information game $\Gamma_{\mathcal{D}}$. Formally the type space of each player is $T_i = A_i \times M_i$ for $i = 1,2$, where each *type* $t_i \in T_i$ denotes the vector (a_i, m_i), i.e. player i's a_i is her *ability type* and m_i is its *message type*.

Given the information disclosure rule $\mathcal{D} = ((M_i)_{i=1,2},\pi)$, the probability that the type vector

[8] As we will see below, for some special information disclosure rules and some values of the parameters ϕ and ψ, the set $\mathfrak{E}(\mathcal{D},\psi,\phi)$ is a singleton.

$t = (t_1, t_2) = \big((a_1, m_1), (a_2, m_2)\big) \in T$ is realized, denoted by $\mathsf{P}_T(t)$, can be computed by[9]

$$\mathsf{P}_T(t) = \mathsf{P}(a)\pi(m|a). \tag{2.5}$$

Of course, we will have that[10] $\mathsf{marg}_A \mathsf{P}_T = \mathsf{P}$, i.e. the marginal with respect to the ability types A of the joint measure P_T must equal the original prior over A.

After player i learns his ability type a_i and hears from the designer his message type m_i, i.e. the type $t_i = (a_i, m_i)$, he uses this information to form a posterior belief $\hat{p}_i : T_i \to \Delta(T_j)$[11] over the types of his rival $t_j \in T_j$ that he deems possible:

$$\hat{p}_i(t_j|t_i) = prob(t_j|t_i) = \begin{cases} \dfrac{\mathsf{P}\big(a_i,a_j\big)\pi\big((m_i,m_j)\big|(a_i,a_j)\big)}{\sum_{a_i,m_j}\mathsf{P}\big(a_i,a'_j\big)\pi\big((m_i,m'_j)\big|(a_i,a'_j)\big)} & \text{if } t_i \in \mathsf{supp}(\mathsf{marg}_{T_i}\mathsf{P}_T) \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

After an information disclosure rule $\mathcal{D}$ is put in place, the analysis of the extended game $\Gamma_{\mathcal{D}}$ can be carried out in the standard way by extending the type space to T, with P_T being the common prior over it and each player i will calculate their respective posterior about the realized type vector t after receiving their own type t_i.

However, there are situations in which we can further simplify the analysis by considering a smaller type space than T. Notice from equations (2.5) and (2.6), we could possibly have that all players assign probability 1 to some particular subsets of T or believe that types in another subset are no longer possible after they receive their information. These ideas can be expressed rigorously with the help of the conditional probabilities as calculated by equation (2.6) by defining the notion of a *belief closed subsets* of the type space.[12]

Definition 2.3. *A subset $W = W_1 \times W_2$ of T is called* ***belief closed*** *if for every player $i = 1, 2$, $W_i \subset T_i$ and for every $t_i \in W_i$, the posterior probability $prob(\cdot|t_i)$ assigns probability one to the set W_j, $j \neq i$.*

Thus, if the profile of players types $t = (t_1, t_2)$ is in the belief closed subset W, under a common prior, this fact can be made *common knowledge* among the two players.

Consider now the equilibrium set $\mathfrak{E}(\mathcal{D})$ and let $\sigma : T \to \Delta(E_1) \times \Delta(E_2)$ be a strategy profile belonging to the equilibrium set. Then, let $\sigma^{\Delta}(\cdot|t) \in \Delta(E)$ be the induced product measure over E

[9] Rigorously, if we define $h_A : A \times M \to A$ as $h_A(a, m) = a$ and $h_M : A \times M \to M$ as $h(a, m) = m$, if $t = (a, m)$ then $\mathsf{P}_T(t) = \mathsf{P}\big(h_A(a, m)\big)\pi\big(h_M(a, m)|h_A(a, m)\big)$.

[10] For a joint probability measure $p \in \Delta(X \times Y)$, $\mathsf{marg}_X\, p \in \Delta(X)$ denotes the marginal distribution over X induced by p.

[11] We depart form the usual notational convention in the literature that denotes the posterior belief of player i at his type t_i about the rival's type t_j as $\hat{p}_i(t_i)[t_j]$ and instead write this belief as $\hat{p}_i(t_j|t_i)$.

[12] The definition (Myerson, 1991, p. 81) of belief closed subsets is usually written in terms of the *Universal Belief Space* (Mertens and Zamir, 1985; Brandenburger and Dekel, 1993). However, the finite and consistent (common prior) type-space that we are using in this paper, a *Harsanyi Model*, can be embedded into the Universal Belief Space. Another name in the literature for belief closed subsets is *belief subspaces* (Zamir, 2009). Whichever the definition or name that we use, the notion that they describe is similar: a subset that contains all the states for the world which are relevant to the situation we are analyzing. Thus, a belief-closed subset describes events that become effectively common knowledge between the players when they happen.

for each $t \in \operatorname{supp}(\mathsf{P}_T)$, i.e. $\sigma^{\Delta}(e|t) = \sigma_1(e_1|t_1)\sigma_2(e_2|t_2)$ for each $e \in E$ and $t \in \operatorname{supp}(\mathsf{P}_T)$. The type action measure induced by the profile σ will be denoted by $\chi_\sigma \in \Delta(T \times E)$ and is given by $\chi_\sigma(t,e) = \mathsf{P}_T(t)\sigma^{\Delta}(e|t)$ for each $t \in \operatorname{supp}(\mathsf{P}_T)$ and each $e \in E$. Notice that the expectation of the utility of the designer in equation (2.3) is taken with respect to χ_σ. In particular, the expression in the inner parenthesis of (2.3) is a posterior probability:

$$\frac{\operatorname{marg}_{A\times E}[\chi_\sigma](a,e)}{\operatorname{marg}_A[\chi_\sigma](a)} \quad \text{for each } a \in A \text{ and each } e \in E,$$

and then each of these posteriors is averaged over $(e,a) \in E \times A$ using as weights the original prior probability $\mathsf{P}(a)$.

2.3.2 Hierarchies of beliefs and properties of information disclosure rules

Consider the extended Bayesian game $\Gamma_{\mathcal{D}}$ induced by the information disclosure rule $\mathcal{D}$ and the extended type space associated with it, $T = A \times M$. We have described previously in the previous subsection how to calculate the posterior beliefs for the players about the type of their opponent for each type they may end up having: their private information about their respective abilities and the messages received from the designer. However, recall that the only payoff-relevant information from the point of view of the players is their ability types, i.e. the vector $a \in A$. So it is important to understand how the information disclosure rule affect the beliefs of each player about a, i.e. their *first-order beliefs*. Moreover, given the uncertainty the players have about the full vector of abilities a, and since the decisions the other player takes are relevant, then so are their beliefs about what beliefs about a the opponent holds, i.e. their *second-order beliefs*. Similarly, since the second-order beliefs are relevant and unknown to the players, then they must also hold beliefs about the second-order beliefs, i.e. their *third-order beliefs* and so on. Thus, the notion of the *infinite hierarchies of beliefs* pops up naturally in our context.

Although Harsanyi's (Harsanyi, 1967, 1968a,b) notion of *type* allows us to bypass explicitly considering the infinite hierarchies of beliefs it is nevertheless instructive, for the purposes of this paper, to analyze how an information disclosure rule $\mathcal{D}$ impacts those hierarchies. In particular, the explicit construction of the hierarchy of beliefs induced by an optimal information disclosure rule will allows us to describe what is its role in inducing the players to behave as intended by the designer. Furthermore, we will attempt to classify the optimal information disclosure rules by how they affect the hierarchy of beliefs. Finally, this classification will depend on some properties of the information disclosure rules that depend on how they affect some or all levels of the hierarchy.

Because of the previous discussion we now present a discussion of how to extract the hierarchies of beliefs from the extended type-space T and the posterior probabilities $\hat{p}_i : T_i \to \Delta(T_j)$ from each player. The construction that we present is standard (Battigalli, 2018; Maschler, Solan, and Zamir, 2013), which we adapt to the current model in the paper.

Extracting the infinite hierarchy Recall, that since the game has a common prior over A and the designer commits to the information disclosure rule $\mathcal{D}$ and publicly announces its structure, we can take the structure of the induced game $\Gamma_{\mathcal{D}}$ and its extended type-space $T = A \times M$ to be common knowledge. Therefore, the posterior probability functions $\hat{p}_i : T_i \to \Delta(T_j)$ are also common knowledge.

Now, define for each player the function $\vartheta_i : T_i \to A_i$ as the projection of $T_i = A_i \times M_i$ into the set of ability types A_i i.e. $\vartheta_i(a_i, m_i) = a_i$. The *first-order beliefs* of player i about $a \in A$ (since player i knows his own ability, i.e. his component a_i of $a \in A$, then his beliefs about a are completely defined by his beliefs about A_j):

$$\forall a_j \in A_j,\ h_i^1(a_j|t_i) = \hat{p}_i(\vartheta_i^{-1}(a_j)|t_i) = (\mathsf{marg}_{A_j}\ \hat{p}_i(\cdot|t_i))(a_j),$$

where $\vartheta_j^{-1}(a_j) = \{t_j \in T_j : \vartheta_j(t_j) = a_j\}$. The functions $t_j \mapsto h_j^1(\cdot|t_j) \in \Delta(A_i)$ for $j = 1, 2$ are also common knowledge. Thus, we can also calculate the *second-order beliefs* about $a \in A$, which are the joint beliefs of a player about a and about his opponent's first-order beliefs about a:

$$\forall(\bar{a}_j, \bar{h}_j^1) \in A_j \times \Delta(A_i),\ h_i^2(\bar{a}_j, \bar{h}_j^1|t_i) = \sum_{\substack{t_j : \vartheta_j(t_j) = \bar{a}_j,\\ h_j^1(\cdot|t_j) = \bar{h}_j^1}} \hat{p}_i(t_j|t_i) = \hat{p}_i\left((\vartheta_j, h_j^1)^{-1}(\bar{a}_j, \bar{h}_j^1)|t_i\right),$$

where $(\vartheta_j, h_j^1)^{-1}(\bar{a}_j, \bar{h}_j^1) = \{t_j \in T_j : (\vartheta_j, h_j^1)(t_j) = (\bar{a}_j, \bar{h}_j^1)\}$. Notice that $h_i^1(\cdot|t_i) = \mathsf{marg}_{A_j}\ h_i^2(\cdot|t_i)$. Intuitively, this means that the first-order beliefs of $h_i^1(\cdot|t_i) \in \Delta(A_j)$ of player i can be obtained as the marginal distribution over A_j of the joint distribution $h_i^2(\cdot|t_i) \in \Delta(A_j \times \Delta(A_i))$[13]. We can iterate the construction to compute for each type, the corresponding *third-order beliefs* about a, *fourth-order beliefs* about a, and so on. Therefore, the beliefs of all orders of player i about $a \in A$ are determined by his type t_i according to the function

$$t_i \mapsto (\vartheta_i(t_i), h_i^1(\cdot|t_i), h_i^2(\cdot|t_i), h_i^3(\cdot|t_i), \ldots).$$

The *infinite hierarchy of beliefs* of player i at type t_i is thus $h_i(\cdot|t_i) = (h_i^k(\cdot|t_i))_{k=1}^{\infty}$. As we defined before, the ability-type of a player $a_i = \vartheta_i(t_i)$ is only one component of his overall type, which also includes information about the beliefs about all the relevant parameters $a_j, h_j^1, h_j^2, h_j^3, \ldots$, of the game $\Gamma_{\mathcal{D}}$.

For the original game, Γ in which the type space $T = A$, with the statistically independent prior P_ψ, note that the hierarchies of beliefs are quite simple, since they are identical across players and all ability types: $\underline{h}^1 = \psi[\alpha] + (1-\psi)[\beta]$, $\underline{h}^2 = \psi[(\alpha, \underline{h}^1)] + (1-\psi)[(\beta, \underline{h}^1)]$, $\underline{h}^3 = \psi[(\alpha, \underline{h}^1, \underline{h}^2)] + (1-\psi)[(\beta, \underline{h}^1, \underline{h}^2)]$ and so on. We denote the infinite hierarchy for the original game $\underline{h} = (\underline{h}_1, \underline{h}_2, \ldots)$.

Properties of Information disclosure rules With the notion of the posterior probability and hierarchies of beliefs induced by the messages received by the disclosure rule $\mathcal{D}$ we can now state some

[13] This is the notion of *coherency.*

descriptive properties, in terms of their information content, of information disclosure rules. Some of this properties are adapted from Mathevet et al. (2016) but adapted by taking into consideration that the players possess private information because they know their own abilities.

Definition 2.4. *We say that an information disclosure rule* $\mathcal{D} = \big((M_i)_{i=1,2}, \pi\big)$ *satisfies* ***public disclosure*** *if both of the following two conditions hold:*

1. *For all* $m \in \bigcup_{a\in A} \mathrm{supp}(\pi(\cdot|a))$ *and for every* $a \in A$, *we have that* $h_i^1(a_j|t_i) = h_j^1(a_i|t_j)$ *for all* $i, j, i \neq j$ *and all* $t_i = (a_i, m_i), t_j = (a_j, m_j)$.

2. $\mathrm{marg}_{M_j}\ \hat{p}_i(\cdot|t_i) = [m_j]$ *for all* t_i *and for all* i *and* $j, i \neq j$.

Any information rule that is not public is said to satisfy ***private disclosure****.*

For symmetric message spaces, we can consider a simplified version of the previous definition directly in terms of the family of conditional distributions π from the disclosure rule.

Definition 2.5. *The information disclosure rule* $\mathcal{D} = \big((M_i)_{i=1,2}, \pi\big)$ *is said to satisfy* ***public disclosure*** *if for all* $i, j = 1, 2$ $i \neq j$ *we have that* $M_i = M_j = M_p$ *and* π *satisfies for all* $a \in A$

$$\pi\big(\{(m_i, m_j) \in M_p^2 : m_i = m_j\}|a\big) = 1,$$
$$\pi\big(\{(m_i, m_j) \in M_p^2 : m_i \neq m_j\}|a\big) = 0.$$

The set $\{(m_i, m_j) \in M_p^2 : m_i = m_j\}$ *will be called the* diagonal *of* M_p^2, $\mathrm{diag}\,(M_p^2)$. *Any information disclosure rule which is not public is said to satisfy* ***private disclosure****.*

The next properties that we discuss have to do with the *informativeness* of the information disclosure rule.

Definition 2.6. *We say that an information disclosure rule* $\mathcal{D} = \big((M_i)_{i=1,2}, \pi\big)$ *is*

1. ***Uninformative*** *if* $h_i(\cdot|t_i) = \underline{h}(\cdot)$ *for all* $t_i = (a_i, m_i) \in T_i$ *and* i.

2. ***Informative*** *if it is not uninformative.*

3. ***Certain for player*** i ***at type*** t_i *if there is an ability* $a_i \in A_i$ *and a message* $m_i' \in M_i$ *such that* $t_i = (a_i, m_i')$ *and* $h_i^1(\cdot|t_i) = [a_j]$ *for some* $a_j \in A_j$.

4. ***Completely certain*** *if it is public and certain for all players and all types* $t \in T = A \times M$.

5. ***Correlated*** *if exists* $a \in A$ *such that* $\big(\mathrm{marg}_{M_1}\ \pi(\cdot|a)\big)\big(\mathrm{marg}_{M_2}\ \pi(\cdot|a)\big) \neq \pi(\cdot|a)$.

6. ***Uncorrelated*** *if for all* $a \in A$, $\big(\mathrm{marg}_{M_1}\ \pi(\cdot|a)\big)\big(\mathrm{marg}_{M_2}\ \pi(\cdot|a)\big) = \pi(\cdot|a)$.

Intuitively, an uninformative information disclosure rule leaves the players with the same beliefs as they had before receiving the message from the designer, while an informative rules alters the infinite hierarchy of beliefs non-trivially. Under a certain information disclosure rule, some player with some ability type might come to belief with certainty that the true ability vector after hearing a particular message from the designer. However, this may not hold true for other players or messages. Finally, if a rule is certain for all players and at all messages while at the same time being public, then it is completely certain. This means that not only all players come to believe with certainty the true ability vector, but this fact also becomes *common knowledge.*

3 Simplifying the designer's problem

The goal of this section is to describe the approach that we will take to obtain the solution of problem (2.4). In the definition of this problem, notice that the space from which we take the "outside" maximization is the space of all finite-message information disclosure rules. This set is an infinite dimensional space. Consider two natural numbers n_1 and n_2 and let $N = n_1 n_2$. Consider the message space $M = M_1 \times M_2$ in which the individual message spaces contain respectively n_1 and n_2 messages, so that the M contains N possible joint messages. Then for each $a \in A$, the $\pi(\cdot|a) \in \Delta(M)$ is a point in the $(N-1)$-dimensional simplex, i.e.

$$\pi(\cdot|a) \in \Delta_{N-1} = \left\{ x \in \mathbb{R}^N : \sum_{j=1}^{N} x_j = 1, x_j \geq 0 \right\}$$

Thus, the space of all finite-message information structures is given by

$$\mathcal{I} = \bigcup_{\substack{(n_1,n_2)\in\mathbb{N}^2 \\ n_1 n_2 = N}} \{M : M = M_1 \times M_2, |M_1| = n_1, \ |M_2| = n_2\} \times \left(\Delta_{N-1}\right)^4.$$

Therefore, as it is right now, finding the optimal disclosure rule in problem (2.4) is potentially very hard, since the set $\mathcal{I}$ of decision variables is infinite-dimensional.

However, we can use a generalization of Aumann's correlated equilibrium (Aumann, 1987, 1974) to games of incomplete information due to Bergemann and Morris (2016) to simplify the problem.

Definition 3.1. *Let $\lambda : A \to \Delta(E)$ be a* decision rule*, i.e. a family of conditional probability distributions over E indexed by the states $a \in A$. Then we say that λ is a* ***Bayes Correlated Equilibrium*** *(BCE) of $(\psi, \phi) inC$ if for each $i = 1, 2$, $a \in A$ and $e_i \in E_i$ we have that*

$$\sum_{e_j,a_j} prob(a_j|a_i)\lambda(e_i, e_j|a)\hat{u}_i(e_i, e_j, a) \geq \sum_{e_j,a_j} prob(a_j|a_i)\lambda(e_i, e_j|a)\hat{u}_i(e_i', e_j, a) \ \forall e_i' \in E_i \tag{3.1}$$

As we mentioned in the introduction, Taneva (2015) was the first one to to use the notion of a BCE to provide the general finte approach to derive the optimal information structure of the designer. Her approach, which is the one we follow in this paper, is based on the following theorem from Bergemann and Morris (2016), which provides the cornerstone of the analysis.

Theorem 3.2. *(Bergemann and Morris, 2016, Thm. 1, p. 495). Decision rule λ is a BCE of $(\psi, \phi) \in C$ if and only if there exists and information disclosure rule $\mathcal{D}$ and a BNE of $\Gamma_{\mathcal{D},\psi,\phi}$ which induces λ. Strategy profile b of $\Gamma_{\mathcal{D}}$ induces λ as follows:*

$$\lambda(e|a) = \sum_{m \in M} \pi(m|a) \left(\prod_{i=1,2} b_i(e_i|a_i, m_i) \right) \quad \forall (a, e) \in A \times E. \tag{3.2}$$

If λ is a BCE of $\psi, \phi \in C$ then the payoff for the designer from λ is, when the productivities are $k \in K$

$$V(\lambda; \psi, \phi, k) = \sum_{a,e} \mathsf{P}(a)\lambda(e|a)v(e, a). \tag{3.3}$$

Therefore, what the theorem claims is that

$$\bar{V}(\psi, \phi, k) = \max_{\mathcal{D}} \max_{\sigma \in \mathfrak{E}(\psi,\phi,\mathcal{D})} V(\mathcal{D}, \sigma; \psi, \phi, k) = \max_{\lambda \in BCE(\psi,\phi)} V(\lambda; \psi, \phi, k). \tag{3.4}$$

The inner maximization in the middle part of equation (3.4) implies that we are using an equilibrium selection criterion. In the theorem, the quantifier *"there exists"* is equivalent to this inner maximization. Thus, by using theorem 3.2 we cannot escape from the selection criterion. Changing the quantifier in the statement of the theorem to *for all* implies changing the inner maximization for a *minimization*. With this qualification, then the information design problem becomes of finding the best information rule assuming that the agents will play the worst equilibrium. In this case we can no longer apply theorem 3.2. Recent contributions (Mathevet et al., 2016; Carroll, 2016) are attempts to push the analysis for this case, which the literature calls *adversarial information design* (Bergemann and Morris, 2018). However, we will show that in the model we are considering, the best equilibrium selection issue is diminished since in the equilibrium set induced by the optimal disclosure rule, the best equilibrium turns out to be the *unique pure strategy symmetric equilibrium* generically.

After finding an *optimal Bayes Correlated Equilibrium*, it is straightforward to come up with an optimal information structure. The next proposition, which is a corollary of theorem 3.2, explains how to do it.

Proposition 3.3. *Let $\lambda^*(\psi, \phi, k) \in \arg\max_{\lambda \in BCE(\psi,\phi)} V(\lambda; \psi, \phi, k)$. An optimal information structure $\mathcal{D}^* = (M^*, \pi^*)$ can be constructed as follows:*

- *We set $M^* = M_1^* \times M_2^*$ where $M_i^* = E_i$ for each $i = 1, 2$. The previous message space is* canonical *in the sense that it will provide an action recommendation. Alternatively, any space M' which isomorphic to M^* also works, i.e. if we can establish a bijection $M_i^* \leftrightarrow M_i'$ for the individual message spaces for each $i = 1, 2$.*

- *We set $\pi^*(m|a) = \lambda^*(e|a)$ for all $e \in E$, $m \in M^*$ and $a \in A$. In here M^* stands for either the canonical message space or any alternate message space that is isomorphnic to it.*

- *Any disclosure rule $\mathcal{D}^*$ which uses canonical message spaces or an isomorphic message space will be called* canonical.

Let σ^* be the best BNE for the designer under the disclosure rule $\mathcal{D}^*$, i.e.

$$\sigma^* \in \underset{\sigma \in \mathfrak{E}(\psi, \phi, \mathcal{D}^*)}{\arg\max} V(\mathcal{D}^*, \sigma; \psi, \phi, k).$$

Then $\bar{V}(\psi, \phi, k) = V(\mathcal{D}^*, \sigma^*; \psi, \phi, k)$. Furthermore, if $\mathcal{S}$ is any other optimal information disclosure rule, not necessarily canonical, then there exists a canonical disclosure rule $\mathcal{D}^*$ such that

$$\bar{V}(\psi, \phi, k) = \max_{\sigma \in \mathfrak{E}(\psi, \phi, \mathcal{D}^*)} V(\mathcal{D}^*, \sigma; \psi, \phi, k) = \max_{\sigma \in \mathfrak{E}(\psi, \phi, \mathcal{S})} V(\mathcal{S}, \sigma; \psi, \phi, k),$$

that is, we can always find for any optiimal information disclosure rule, a canonical rule that is also optimal.

The previous proposition follows the same line of reasoning as proposition 2 in Taneva (2015). Intuitively, its main implication is that it is without loss to work with an information disclosure rule that supplies to the players an *action recommendation.* This is similar in spirit to the revelation principle (Myerson, 1979, 1991). According to proposition 3.3, if σ^* is a BNE under the information disclosure rule π^* and the private information already held by the players, then by theorem 3.2, the decision rule that it generates is a BCE. Thus, when the designer uses an information disclosure rule that mimics exactly this disclosure rule as stated in the proposition, then by the obedience constraints in the definition of a Bayes Correlated Equilibrium, it will be in the best interest of the players to actually play the equilibrium profile σ^* that is implicitly recommended by that information disclosure rule. Thus, a designer that wishes to design the optimal information disclosure rule, can simplify the problem by first looking at the optimal set of BCE distributions, pick the best one and then use it to construct the optimal information disclosure rule.

Another important remark about the statement of 3.3 is that we took the time to describe the optimal information disclosure rule not only in terms of the canonical message space that gives action recommendations but also in terms of any message space that is equivalent to it. Although the notion of action recommendations is suitable to perform the analysis and to pin down the optimal information disclosure rule, it is much easier to interpret the informational content of the messages using an equivalent message space.

Although using the BCE notion makes the analysis tractable, it is still an arduous task to compute the optimal BCE distribution. The set of BCE distributions consists of a family of four 2×2 conditional distributions that need to satisfy the obedience constraints in the BCE definition together with a set of constraints that make each member of the family a valid probability distribution. Furthermore, we need to do this for each possible game $(\psi, \phi, k) \in C \times K$ that the designer considers. Appendix A provides the detailed steps to find an optimal BCE.

4 Characterization of the equilibrium sets of the null and complete disclosure rules

In this section, we fully characterize for all games $(\psi, \phi) \in C$ the equilibrium sets of the null and complete information disclosure rules. The purpose of this characterization is threefold: to illustrate the richness of behavior that arises from two simple and intuitive disclosure rules, to expand and complement the main results in Dubey (2013) and to provide a benchmark for the main results in the next section.

The full characterization of the set of Bayes Nash Equilibria for disclosure rules $\mathcal{N}$ and $\mathcal{C}$ across all contests in $C = (0,1)^2$ is given in propositions 4.1 and 4.3. In order to present a clean statement of these propositions, we need some preparations. Consider the following collection of subsets $R_s \subset C$, $s = 1, \ldots, 5$ given by

$$\begin{aligned}
R_1 =&\{(\psi, \phi) \in C \,|\, 2\phi \leq \psi\}, \\
R_2 =&\{(\psi, \phi) \in C \,|\, 2\phi \geq 2 - \psi\}, \\
R_3 =&\{(\psi, \phi) \in C \,|\, \psi \leq 2\phi \leq 1 - \psi\}, \\
R_4 =&\{(\psi, \phi) \in C \,|\, 1 + \psi \leq 2\phi \leq 2 - \psi\}, \\
R_5 =&\{(\psi, \phi) \in C \,|\, \max(1 - \psi, \psi) \leq 2\phi \leq \min(1 + \psi, 2 - \psi)\}.
\end{aligned} \tag{4.1}$$

For reference and visualization of the regions in (4.1), see figure 2. We are now ready to state the propositions, but we relegate the proofs to the appendix.

Proposition 4.1. *Consider the null information disclosure rule $\mathcal{N}$. For each $(\psi, \phi) \in C$ the equilibrium set $\mathfrak{E}(\psi, \phi, \mathcal{N})$ is as follows.*

1. *If $(\psi, \phi) \in \text{int}(R_1)$ then the* unique *equilibrium*[14] *is for players all of abilities to work and is in dominant strategies, that is*

$$\sigma_i^{\mathcal{N}}(\cdot|a_i, \varnothing) = [W], \; \forall a_i \in A_i, \; i = 1, 2. \tag{4.2}$$

2. *If $(\psi, \phi) \in \text{int}(R_2)$ then the* unique *equilibrium is for players all of abilities to shirk and is in dominant strategies, that is*

$$\sigma_i^{\mathcal{N}}(\cdot|a_i, \varnothing) = [S], \; \forall a_i \in A_i, \; i = 1, 2. \tag{4.3}$$

3. *If $(\psi, \phi) \in \text{int}(R_3)$ then the* unique *equilibrium $\sigma^{\mathcal{N}}$ is for the high ability player to always work and the low ability player to always shirk, i.e.*

$$\sigma_i^{\mathcal{N}}(\cdot|\alpha, \varnothing) = [S], \quad \sigma_i^{\mathcal{N}}(\cdot|\beta, \varnothing) = [W], \quad i = 1, 2. \tag{4.4}$$

[14] The notation $\text{int}(X)$ denotes the interior of the set X.

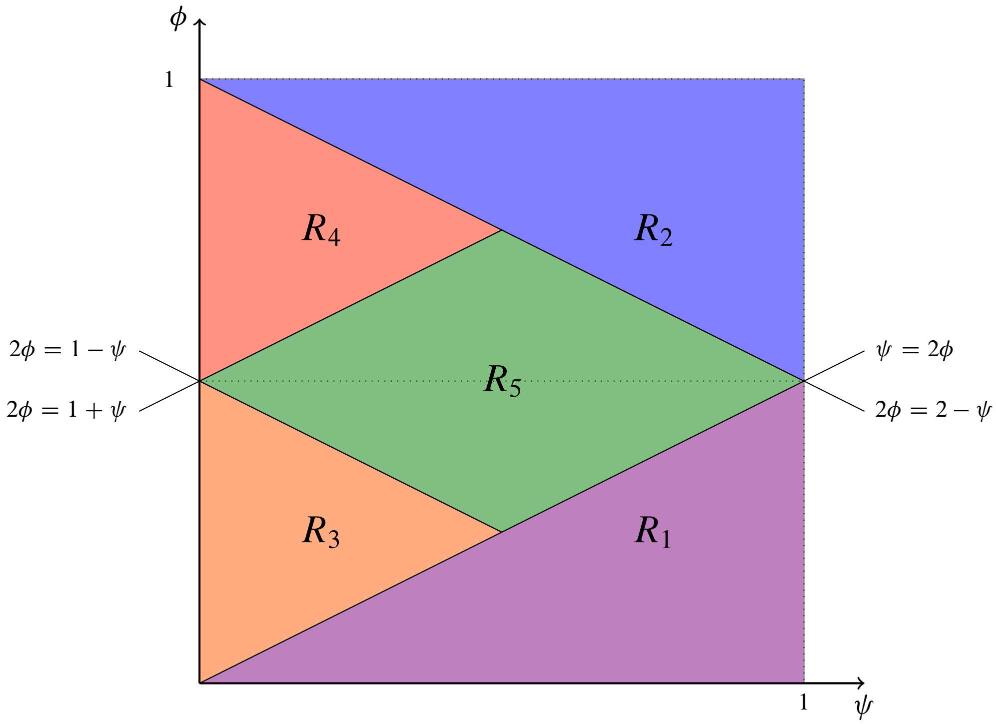


Figure 2: Regions described by (4.1)

4. *If* $(\psi, \phi) \in \mathrm{int}(R_4)$ *then the* unique *equilibrium* $\sigma^{\mathcal{N}}$ *is for the high ability player to always shirk and the low ability player to always work, i.e.*

$$\sigma_i^{\mathcal{N}}(\cdot|\alpha, \varnothing) = [W], \quad \sigma_i^{\mathcal{N}}(\cdot|\beta, \varnothing) = [S], \quad i = 1, 2. \tag{4.5}$$

5. *If* $(\psi, \phi) \in \mathrm{int}(R_5)$*, then the* unique *equilibrium* $\sigma^{\mathcal{N}}$ *is in* completely mixed *strategies:*

$$\sigma_i^{\mathcal{N}}(W|\alpha, \varnothing) = \frac{2\phi + \psi - 1}{2\psi}, \quad \sigma_i^{\mathcal{N}}(W|\beta, \varnothing) = \frac{2 - 2\phi - \psi}{2(1 - \psi)}, \quad i = 1, 2. \tag{4.6}$$

6. *If* $(\psi, \phi) \in R_1 \cap R_5$*, then there is a continuum of equilibrium strategies, the high ability player always work but the low ability player plays any mixed strategy in* $[0, 1]$ *, i.e.:*

$$\sigma_i^{\mathcal{N}}(W|\alpha, \varnothing) \in [0, 1], \quad \sigma_i^{\mathcal{N}}(\cdot|\beta, \varnothing) = [W], \quad i = 1, 2. \tag{4.7}$$

7. *If* $(\psi, \phi) \in \mathrm{ri}(R_1 \cap R_3)$*, the high ability player* [15] *always work and the low ability player plays any mixed strategy as long as it belongs to a particular proper subset of* $[0, 1]$ *, i.e.:*

$$\sigma_i^{\mathcal{N}}(W|\alpha, \varnothing) \in \left[\frac{2\psi - 1}{2\psi}, 1\right], \quad \sigma_i^{\mathcal{N}}(\cdot|\beta, \varnothing) = [W], \quad i = 1, 2. \tag{4.8}$$

[15] The notation $\mathrm{ri}(X)$ denotes the *relative interior* of the set X, i.e. $\mathrm{ri}(X) = \{x \in X | \exists \epsilon > 0, B_\epsilon(x) \cap \mathrm{aff}(X) \subseteq X\}$, where $\mathrm{aff}(X)$ is the affine hull of X and $B_\epsilon(x)$ is a ball of radius ϵ centered at x.

8. *If* $(\psi, \phi) \in \mathrm{ri}(R_3 \cap R_5)$*, then there is a continuum of equilibrium strategies, the low ability player always shirks but the high ability player plays any mixed strategy as long as it belongs to a particular proper subset of* $[0, 1]$ *, i.e.:*

$$\sigma_i^{\mathcal{N}}(\cdot|\alpha, \varnothing) = [S], \quad \sigma_i^{\mathcal{N}}(W|\beta, \varnothing) \in \left[\frac{1}{2(1-\psi)}, 1\right], \quad i = 1, 2. \tag{4.9}$$

9. *If* $(\psi, \phi) \in R_2 \cap R_4$*, then there is a continuum of equilibrium strategies, the high ability player always shirks but the low ability player plays any mixed strategy in* $[0, 1]$ *, i.e.:*

$$\sigma_i^{\mathcal{N}}(W|\alpha, \varnothing) \in [0, 1], \quad \sigma_i^{\mathcal{N}}(\cdot|\beta, \varnothing) = [S], \quad i = 1, 2. \tag{4.10}$$

10. *If* $(\psi, \phi) \in \mathrm{ri}(R_2 \cap R_5)$*, then there is a continuum of equilibrium strategies, the high ability player always shirks but the low ability player plays any mixed strategy as long as it belongs to a particular proper subset of* $[0, 1]$ *, i.e.:*

$$\sigma_i^{\mathcal{N}}(\cdot|\beta, \varnothing) = [S], \quad \sigma_i^{\mathcal{N}}(W|\alpha, \varnothing) \in \left[0, \frac{1}{2\psi}\right], \quad i = 1, 2. \tag{4.11}$$

11. *If* $(\psi, \phi) \in \mathrm{ri}(R_4 \cap R_5)$*, then there is a continuum of equilibrium strategies, the low ability player always works but the high ability player plays any mixed strategy as long as it belongs to a particular proper subset of* $[0, 1]$ *, i.e.:*

$$\sigma_i^{\mathcal{N}}(\cdot|\alpha, \varnothing) = [W], \quad \sigma_i^{\mathcal{N}}(W|\beta, \varnothing) \in \left[0, \frac{1-2\psi}{2-2\psi}\right], \quad i = 1, 2. \tag{4.12}$$

With the characterization given by proposition 4.1 we can calculate the payoff to the designer of the null information disclosure rule, which we state as the next corollary.

Corollary 4.2. *Let* $V(\mathcal{N}; \psi, \phi, k) = \max_{\sigma \in \mathfrak{E}(\mathcal{N})} V(\mathcal{N}, \sigma; \psi, \phi, k)$ *denote the inner maximization in expression* (2.4)*. Then we have that*

1. *If* $(\psi, \phi) \in R_1$, $V(\mathcal{N}; \psi, \phi, k) = 2\big(\psi k(\alpha)\alpha + (1-\psi)k(\beta)\beta\big)$.
2. *If* $(\psi, \phi) \in R_2$, $V(\mathcal{N}; \psi, \phi, k) = 2\big((1-\psi)\alpha + \psi\beta\big)$.
3. *If* $(\psi, \phi) \in R_3 \setminus R_1$, $V(\mathcal{N}; \psi, \phi, k) = 2\big(\psi\alpha + (1-\psi)k(\beta)\beta\big)$.
4. *If* $(\psi, \phi) \in R_4 \setminus R_5$, $V(\mathcal{N}; \psi, \phi, k) = 2\big((1-\psi)\beta + \psi k(\alpha)\alpha\big)$.
5. *If* $(\psi, \phi) \in \mathrm{int}(R_5)$, $V(\mathcal{N}; \psi, \phi, k) = (1-2\phi+\psi)\alpha + (2\phi-\psi)\beta + (2\phi+\psi-1)k(\alpha)\alpha + (2-2\phi-\psi)k(\beta)\beta$.

The next proposition describes the equilibrium behaviour under the complete information disclosure rule.

Proposition 4.3. *Consider the complete information disclosure rule $\mathcal{C}$. For each $(\psi, \phi) \in C$, the equilibrium set $\mathfrak{E}(\psi, \phi, \mathcal{C})$ is as follows.*

1. *If $(\psi, \phi) \in \{(\psi, \phi) \in C | \phi < 1/2\}$, then the equilibrium is* unique*:*

$$\begin{aligned} \sigma_i^{\mathcal{C}}(\cdot|a_i, S) = [W], \quad i = 1, 2,\ a_i \in A_i \\ \sigma_i^{\mathcal{C}}(S|\alpha, D) = 1 - \phi, \quad \sigma_i^{\mathcal{C}}(S|\beta, D) = \phi, \quad i = 1, 2. \end{aligned} \tag{4.13}$$

2. *If $(\psi, \phi) \in \{(\psi, \phi) \in C | \phi > 1/2\}$, then the equilibrium is* unique*:*

$$\begin{aligned} \sigma_i^{\mathcal{C}}(\cdot|a_i, \mathrm{S}) = [S], \quad i = 1, 2,\ a_i \in A_i \\ \sigma_i^{\mathcal{C}}(S|\alpha, \mathrm{D}) = 1 - \phi, \quad \sigma_i^{\mathcal{C}}(S|\beta, \mathrm{D}) = \phi, \quad i = 1, 2. \end{aligned} \tag{4.14}$$

3. *If $(\psi, \phi) \in \{(\psi, \phi) \in C | \phi = 1/2\}$, then there is a continuum of equilibria:*

$$\begin{aligned} \sigma_i^{\mathcal{C}}(\cdot|a_i, \mathrm{S}) = [0, 1], \quad i = 1, 2,\ a_i \in A_i \\ \sigma_i^{\mathcal{C}}(S|\alpha, \mathrm{D}) = 1 - \phi, \quad \sigma_i^{\mathcal{C}}(S|\beta, \mathrm{D}) = \phi, \quad i = 1, 2. \end{aligned} \tag{4.15}$$

Similarly as we did before, with the characterization given by proposition 4.3 we can calculate the payoff to the designer of the complete information disclosure rule, which we state as the next corollary.

Corollary 4.4. *Let $V(\mathcal{C}) = \max_{\sigma \in \mathfrak{E}(\mathcal{C})} V(\mathcal{C}, \sigma)$ denote the inner maximization in expression* (2.4)*. Then we have that*

1. *If $\phi < 1/2$ then $V(\mathcal{C}, \psi, \phi, k) = 2\Big(\psi(1-\psi)\big(\alpha(1-\phi) + \beta\phi\big) + k(\alpha)\alpha\psi(\phi + \psi(1-\phi)) + k(\beta)\beta(1-\psi)(1-\psi\phi)\Big)$.*
2. *If $\phi = 1/2$ then $V(\mathcal{C}, \psi, \phi, k) = \psi\big((1-\psi)(\alpha+\beta) + k(\alpha)\alpha(1+\psi)\big) + k(\beta)\beta(2 - 3\psi + \psi^2)$.*
3. *If $\phi > 1/2$ then*

$$V(\mathcal{C}, \psi, \phi, k) = 2\Big(\psi\alpha\big(1-\phi(1-\psi)\big) + (1-\psi)\beta\big(1-(1-\phi)\psi\big) + \psi(1-\psi)\big(\phi k(\alpha)\alpha + (1-\phi)k(\beta)\beta\big)\Big).$$

In terms of the properties introduced in definition 2.6, the null information disclosure rule $\mathcal{N}$ is non informative and (trivially) public and the complete information rule is completely certain, since it is public and certain for all players at all messages. The null rule doesn't alter the beliefs of the players, while the complete information rule not only informs all the players, but it makes it such that this fact becomes commonly known.

However, the null and complete disclosure rules are in most cases not the optimal rules for the designer. According to proposition 4.1, the null disclosure rule is optimal for the case in which $2\phi \leq \psi$ or equivalently when $2\varkappa \leq \psi\upsilon$. Intuitively, in this case the value of the prize is so high when compared to the cost of putting effort that both players find strictly dominant to put effort without the need to receive more information from the part of the designer. On the other hand, the results in Dubey (2013) find that the complete information rule $\mathcal{C}$ performs, in general, better than the null rule $\mathcal{N}$ when $2\phi > \psi$ and $\phi < 1/2$, or equivalently, $2\varkappa > \psi\upsilon$ and $2\varkappa < \upsilon$. Propositions 4.1 and 4.3 extend the results in Dubey (2013) by extending the parameter space to all $\phi \in (0, 1)$. Thefore, a trivial calculation shows that the complete information disclosure rule, in general, also performs better than the null rule when $1/2 \leq \phi < 1$ or equivalently when $\upsilon \leq 2\varkappa$. However, the analysis the null and complete information rules is not sufficient to pin down the optimum for the contest information design problem. It is possible to construct a rule that outperforms the complete information rule $\mathcal{C}$ for all cases in which the null information rule $\mathcal{N}$ is not optimal. In the next section, we show how to construct the globally optimal information disclosure rule for the information design problem.

5 Optimal information Disclosure: Main results

In this section we discuss the main results of the paper. The first subsection describes the full characterization of the optimal information disclosure rule for each contest in $C \times K$. After presenting these set of results, we illustrate them by means of a numerical example that is meant to showcase the main features of the characterization. Finally in the last subsection, by fully taking advantage of our characterization, we perform a comparative statics exercise in which we allow the designer to alter the value of the prize simultaneously while engaging in information design. The results of this exercise deliver the necessary and sufficient conditions at which the optimal information structure not only achieves the goals of the designer but it also does at the *cheapest* possible way. All proofs are relegated to appendix A.

5.1 Characterization of the optimal information disclosure rule

We begin by describing the main features of the optimal information disclosure rule. In order to describe it in a succinct way, we need to introduce some new terminology and notation that will be useful.

Definition 5.1. *The productivity differential by ability type is the difference in output between working and shirking for each of the ability types, weak and strong. We denote this differentials as*

$$\mathrm{d}_\alpha = f(W,\alpha) - f(S,\alpha) = k(\alpha)\alpha - \alpha$$
$$\mathrm{d}_\beta = f(W,\beta) - f(S,\beta) = k(\beta)\beta - \beta$$

We say that the productivity differential is ***regular*** *if* $\mathrm{d}_\beta > \mathrm{d}_\alpha$. *Otherwise we say that it is* ***non-regular****, i.e.* $\mathrm{d}_\beta < \mathrm{d}_\alpha$.

For every contest environment $(\psi, \phi, k) \in C \times K$, the class of optimal canonical information disclosure rules uses the same message space and the probabilistic structure of the rules share some global features. The next proposition states these facts formally.

Proposition 5.2 (Optimal Information Disclosure Rules). *The class of optimal canonical information disclosure rules*

$$\mathfrak{D} = \left\{ \mathcal{D} : \left((M_i)_{i=1,2}, \pi\right) : \begin{array}{l} \text{there exists } (\psi, \phi, k) \in C \times K \text{ such that} \\ (1)\bar{V}(\psi, \phi, k) = \max_{\sigma \in \mathfrak{E}(\psi,\phi,\mathcal{D})} V(\mathcal{D}, \sigma; \psi, \phi, k), \\ (2)\mathcal{D} \text{ is equivalent to a canonical rule.} \end{array} \right\}$$

has symmetric message spaces given by

$$M_i = \{\textit{Hard-Fought}, \neg\textit{Hard-Fought}\} = \{m^H, \neg m^H\} \quad \textit{for } i = 1, 2,$$

and $\pi : A \to \Delta(M)$ *is given by*

$\pi(\cdot\|\alpha\alpha)$	m^H	$\neg m^H$
m^H	ξ_α	0
$\neg m^H$	0	$1 - \xi_\alpha$

$\pi(\cdot\|\alpha\beta)$	m^H	$\neg m^H$
m^H	μ	δ
$\neg m^H$	ζ	$1-\mu-\delta-\zeta$

$\pi(\cdot\|\beta\alpha)$	m^H	$\neg m^H$
m^H	μ	ζ
$\neg m^H$	δ	$1-\mu-\delta-\zeta$

$\pi(\cdot\|\beta\beta)$	m^H	$\neg m^H$
m^H	ξ_β	0
$\neg m^H$	0	$1 - \xi_\beta$

where

- ξ_a *for* $a = \{\alpha\alpha, \beta\beta\}$ *is the conditional probability that both players receive the same message* m^H *at the states in which they are similar;*
- μ *corresponds to the conditional probability that both players receive the same message* m^H *at states in which they are different;*
- δ *is the conditional probability of the weak player* α *receiving the message* m^H *while the strong player* β *received the message* $\neg m^H$ *at states in which they are different;*
- ζ *is the conditional probability that the weak player* α *received the message* $\neg m^H$ *while the strong player* β *received the message* m^H *at states in which they are different.*

The optimal choices of the previous probabilities for each contest environment are given by the mapping $\Lambda : (\psi, \phi, k) \mapsto (\xi_\alpha, \xi_\beta, \mu, \delta, \zeta)$, *whose dependence on* k *is only through the ratio of productivity differentials* $\mathrm{d}_\beta/\mathrm{d}_\alpha$.

Some remarks about proposition 5.2 are in order. Notice that the message space from the optimal class of information disclosure rules is described in terms of an equivalent space to the canonical message space that gives action recommendation. The choice to present the proposition in that way is due to the fact that it is easier and more intuitive to interpret the information that the designer is giving to the players in terms of a more "amiable" message space. The designer is telling the players whether the contest is gonna be hard-fought or not with different probabilities depending on the state. This of course gives information about the actual state that the players are. Notice that when the players are similar, which is the case in which the competition is more "fierce", the designer tells the same message to both of them. In the event that the players are different, in which there is a higher risk of both of them shirking with higher probability, it is crucial that the designer puts positive probability on giving them different messages in order to curb this behavior.

Another remarkable feature is that the probability mapping Λ depends on the vector of productivities k only though the productivity differential ratio. We can think about this productivity differential ratio as defining, when the players are different, what action profile is more valuable to the designer: a strong player working and a weak player shirking or vice versa, a strong player shirking and a weak player working. The ordering of output between these two cases, which is what the designer is interested in, is defined by the productivity differential ratio. In particular, when productivity is *regular*, in the sense of definition 5.1, the former case is better for the designer; whereas when productivity is *non-regular* the later case is better. When the normalized cost of putting effort ϕ is less than half, or equivalently the value of the prize is greater than two times the cost of putting effort, $\upsilon > 2\varkappa$, productivity ratios that are regular or non-regular become the only important cases that the designer has to consider in terms of k. However, for contest environments in which the the normalized cost of putting effort is more than half, i.e. $\phi > 1/2$, or equivalently $2\varkappa > \upsilon$, the value of the prize is so low that the competition when the players are similar becomes lackluster. Because of this, it is more difficult for the designer to incentivize the players to put effort when the players are similar. Thus the designer needs to make a choice on exactly which state to focus more: when the players are similar and weak, or when the players are similar and strong. This trade-off also has implications on the incentives that the designer is able to give when the players are different. The previous reasons imply that although the probabilities will still depend on k only through the productivity differential ratio, the designer will need to consider cases in which the ratio is regular and very high and cases in which the ratio is non-regular and very low.

On the other hand, the remaining probabilities will also depend on (ψ, ϕ). Recall that each of these points represents a particular contest in C. At each contest, the BCE notion determines what can of behavior can arise as a BNE for some information disclosure rule. This behavior is constrained of course by the obedience constraints embodied in the BCE concept. Furthermore, any information disclosure rule needs to satisfy the probability constraints to make it valid system of conditional probability distributions. At each particular contest (ψ, ϕ), the designer will pick the best BCE distribution from the set of feasible BCE distributions. The best BCE distribution is chosen, of course, in consideration of its value to the designer, which is determined by the productivity vector k through the ratio of the productivity differentials $\mathrm{d}_\beta/\mathrm{d}_\alpha$, ask explained in the last paragraph. Thus, the choice probabilities $(\xi_\alpha, \xi_\beta, \mu, \delta, \zeta)$ depends on the trade-off between what is feasible for the designer at a particular contest and achieving her goals. More precisely, what is feasible at the (ψ, ϕ) depends on the trade-offs that the players face between obtaining the prize and the cost of putting effort, whereas the goals of the designer depend on the value of effort from the players in terms of productivity.

The previous discussion summarizes the intuition about the behavior of the mapping Λ, which gives the optimal probabilities at each contest environment. The characterization of this mapping is given in ample detail in appendix A and is quite involved. However, the next proposition summarizes some qualitative features of Λ.

Proposition 5.3. *Consider the mapping $\Lambda : (\psi, \phi, k) \mapsto (\xi_\alpha, \xi_\beta, \mu, \delta, \zeta)$ which gives the optimal probabilities for the optimal information disclosure rule at each contest environment. Then there are:*

1. *a set of regions of the set of productivities $\{K_t\}_{t=1,\dots,5}$ in which the boundaries of each region are determined by the following cut-off values on the productivity differential ratio:*

$$\left\{\frac{2\phi-1}{2\phi}, 1, 2, \frac{2\phi}{2\phi-1} : 1/2 < \phi < 1\right\};$$

2. *a set of regions of the set of contests $\{C_s\}_{s=1,\dots,15}$ and a family of functions*

$$\{\Lambda_s : (\psi, \phi) \mapsto (\xi_\alpha, \xi_\beta, \mu, \delta, \zeta)\}_{l\in\mathcal{L}}$$

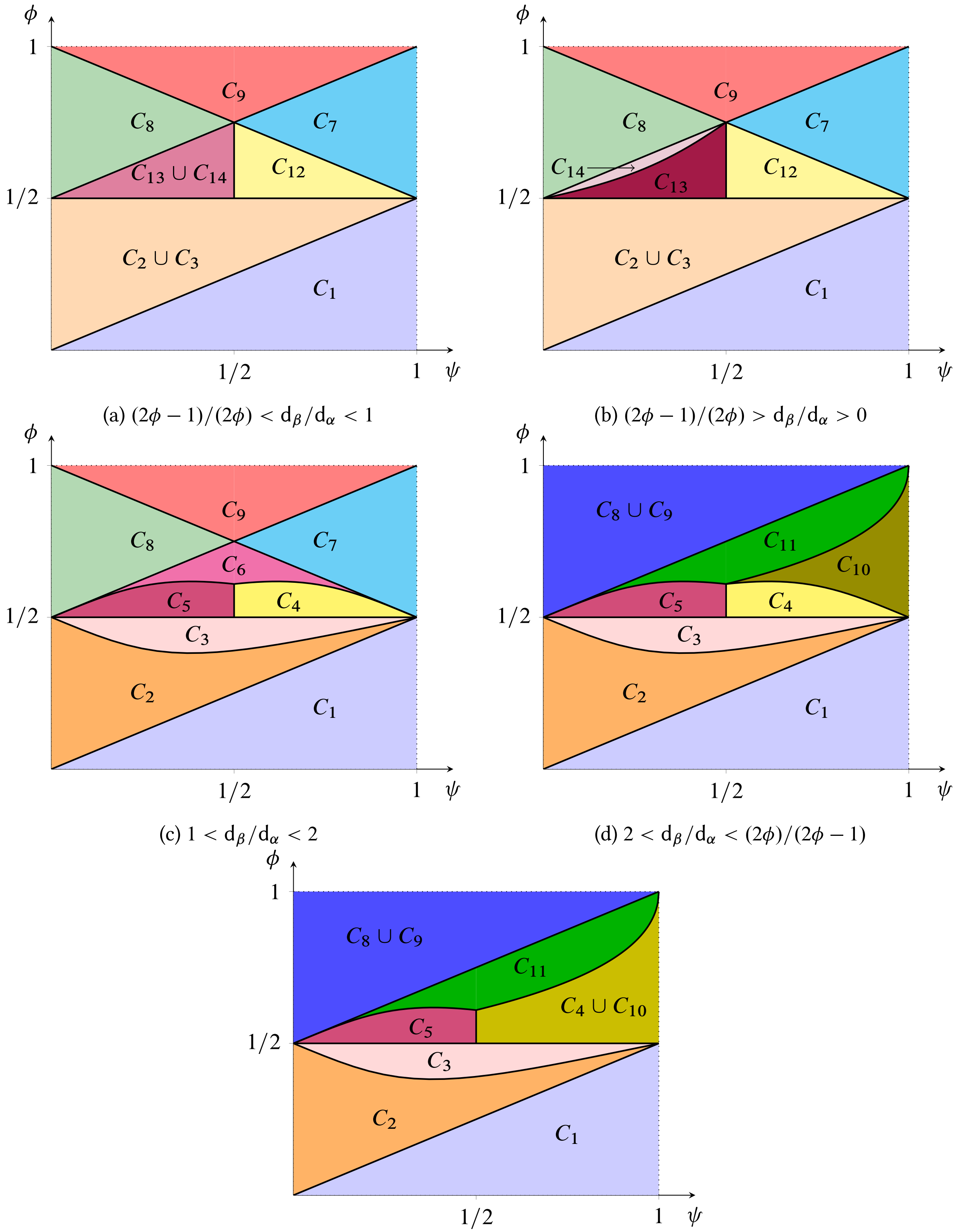
that do not depend on k, such that

for each C_s , there exists a region $K_{t'}$ and a function $\Lambda_{l'}$ such that $\Lambda|_{C_s\times K_{t'}} = \Lambda_{l'}$, i.e. the restriction of Λ to $C_s \times K_{t'}$ is equal to the function $\Lambda_{l'}$ which is constant (does not depend) with respect to K. This means that Λ is a piece-wise function whose parts are defined by $\{C_s\}_{s=1,\dots,15}$ an at each region, the implied probabilities are independent from k.

The contents of proposition 5.3 are illustrated by figure 3, which denotes the regions $\{K_t\}_{t=1,\dots,5}$, and by figure 4, which illustrates the regions $\{C_s\}_{s=1,\dots,15}$

Figure 3: Ratio of productivity differentials—relevant regions

As mentioned before, the full characterization of Λ is rather involved. However, there is a subset of the collection of regions at which the description of Λ is relatively simple. Furthermore, these regions will play a role in the subsequent comparative statics exercise that we will carry out in subsection 5.3. We present these regions, the optimal value of the probabilities and the value to the designer at each of them in the next proposition.



(a) $(2\phi - 1)/(2\phi) < \mathrm{d}_\beta/\mathrm{d}_\alpha < 1$

(b) $(2\phi - 1)/(2\phi) > \mathrm{d}_\beta/\mathrm{d}_\alpha > 0$

(c) $1 < \mathrm{d}_\beta/\mathrm{d}_\alpha < 2$

(d) $2 < \mathrm{d}_\beta/\mathrm{d}_\alpha < (2\phi)/(2\phi - 1)$

(e) $(2\phi)/(2\phi - 1) < \mathrm{d}_\beta/\mathrm{d}_\alpha$

Figure 4: Structure of $\Lambda(\psi, \phi, k)$

Proposition 5.4. *Consider the following regions:*

$$
\begin{aligned}
C_1 &= \{(\psi,\phi)\in C : 2\phi \le \psi\},\\
C_2 &= \left\{(\psi,\phi)\in C : \psi \le 2\phi \le \frac{1-2\psi-\psi^2+\sqrt{1-4\psi+10\psi^2-4\psi^3+\psi^4}}{2(1-\psi)}\right\},\\
C_3 &= \left\{(\psi,\phi)\in C : \frac{1-2\psi-\psi^2+\sqrt{1-4\psi+10\psi^2-4\psi^3+\psi^4}}{2(1-\psi)} \le 2\phi \le 1\right\},\\
C_{15} &= \{(\psi,\phi)\in C : 2\phi = 1\}.
\end{aligned}
$$

Then the optimal probabilities, as given by $\Lambda : (\psi,\phi,k)\mapsto(\xi_\alpha,\xi_\beta,\mu,\delta,\zeta)$ are

- *For all $(\psi,\phi,k)\in C_1\times K$, $\Lambda(\psi,\phi,k)$ is constant and is given by*

$$\Lambda(\psi,\phi,k) = (1,1,1,0,0).$$

  Furthermore, Λ gives an optimal value of

$$\bar{V}(\psi,\phi,k) = 2\big(\psi(\mathrm{d}_\alpha+\alpha)+(1-\psi)(\mathrm{d}_\beta+\beta)\big) = 2\big(\psi k(\alpha)\alpha+(1-\psi)k(\beta)\beta\big)$$

- *If productivity is regular and $(\psi,\phi)\in C_2$ then $\Lambda(\psi,\phi,k)$ is given by*

$$\Lambda(\psi,\phi,k) = \left(1,1,\tfrac{\psi(1-2\phi)}{2\phi(1-\psi)},0,\tfrac{2\phi-\psi}{2\phi(1-\psi)}\right).$$

  Furthermore, Λ implies that the optimal value for the designer is

$$\bar{V}(\psi,\phi,k) = 2\mathrm{d}_\beta(1-\psi)+\mathrm{d}_\alpha\frac{\psi^2}{\phi}+2(\alpha\psi+\beta(1-\psi)).$$

- *If productivity is regular and $(\psi,\phi)\in C_3$ then $\Lambda(\psi,\phi,k)$ is given by*

$$\Lambda(\psi,\phi,k) = \begin{pmatrix} 1\\ 1\\ \frac{2\psi+\phi(3-4\psi-\psi^2)-2\phi^2(1-\psi)-1}{2\psi(1-\psi)}\\ \frac{2\phi^2(1-\psi)-\psi^2+\phi\left(1-2\psi-\psi^2\right)}{2(1-\psi)\psi}\\ \frac{2\phi^3(1-\psi)+\psi^2+\phi\left(2-2\psi-3\psi^2\right)+\phi^2\left(6\psi+\psi^2-5\right)}{2\phi\psi(1-\psi)} \end{pmatrix}.$$

  Furthermore, Λ implies that the optimal value for the designer is

$$
\begin{aligned}
\bar{V}(\psi,\phi,k) = {}& \mathrm{d}_\alpha\left(2\phi(1-\psi)+2\psi+\psi^2-1\right)\\
&+\mathrm{d}_\beta\left(3-2\phi(1-\psi)-4\psi-\psi^2+\frac{\psi^2}{\phi}\right)+2(\alpha\psi+\beta(1-\psi)).
\end{aligned}
$$

- *If productivity is non-regular and* $(\psi, \phi) \in C_2 \cup C_3$ *then* $\Lambda(\psi, \phi, k)$ *is given by*

$$\Lambda(\psi, \phi, k) = \left(1, 1, \tfrac{2\phi^2 - 2\phi^3 + \psi - 3\phi\psi + \phi^2\psi}{2\phi(1-\psi)}, \tfrac{\phi(2\phi-\psi)}{2(1-\psi)}, \tfrac{(1-\phi)^2(2\phi-\psi)}{2\phi(1-\psi)}.\right)$$

  Furthermore, Λ *implies that the optimal value for the designer is*

$$\bar{V}(\psi, \phi, k) = \mathrm{d}_\beta \left(2 - 2(1+\phi)\psi + \psi^2\right) + \mathrm{d}_\alpha \left(2\phi\psi - \psi^2 + \frac{\psi^2}{\phi}\right) + 2(\alpha\psi + \beta(1-\psi)).$$

- *If* $(\psi, \phi) \in C_{15}$*, i.e.* $(\psi, \phi) = (0, 1) \times \{1/2\}$ *and for all* $k \in K$*, we have that* $\Lambda(\psi, \phi, k)$ *is constant and is given by*

$$\Lambda(\psi, \phi, k) = (1, 1, 1/4, 1/4, 1/4).$$

  Furthermore, Λ *gives, for all* $\psi \in (0, 1)$ *and* $k \in K$*, an optimal value of*

$$\bar{V}(\psi, 1/2, k) = 2(\alpha\psi + \beta(1-\psi)) + \mathrm{d}_\alpha\psi(1+\psi) + \mathrm{d}_\beta(1-\psi)(2-\psi)$$

An important remark about proposition 5.4 is that if $0 < \phi \leq 1/2$ then when the players are similar, they will hear with probability one the message "hard-fought", i.e. $(\xi_\alpha, \xi_\beta) = (1, 1)$. Also notice that in region C_1, all probabilities are independent of the contest environment. This is just a restatement of the result from proposition 4.1 when the value of the prize is sufficiently large. In this case, it will be optimal to give a trivially uninformative public signal, or more easily, not give any information at all. On the other hand, region C_{15} is also special because all the probabilities in there are also independent of the contest environment. This is due to the fact that when $\phi = 1/2$, the set of BCE shrinks to a lower-dimensional subspace, which constraints our choice of the optimal information disclosure rule to put $(\mu, \delta, \zeta) = (1/4, 1/4, 1/4)$. On the other hand, we will refer to environments in union of the regions $C_2 \cup C_3$ as environments with *medium prizes.*

Now at last we are ready to state the main theorem of this section, which is a description of the global properties of the optimal information disclosure rules.

Theorem 5.5 (Global properties of the set of optimal information disclosure rules)**.** *The class of optimal canonical information disclosure rules* $\mathfrak{D}$ *satisfies the following properties:*

- *Every* $\mathcal{D}$ *is* ***informative*** *except when* $(\psi, \phi) \in C_1$*, i.e. except when the value of the prize is sufficiently large,* $\upsilon \geq \frac{2\varkappa}{\psi}$*.*

- *Every* $\mathcal{D}$ *is* ***never completely certain****, but for some regions it may be* ***certain for some players and some ability-message pairs*** (a_i, m_i)*, i.e. some players may become certain about the ability vector but this fact will not be common knowledge.*

- *Every* $\mathcal{D}$ *is* ***asymmetric****:* $\delta \neq \zeta$ *except on region* C_1*, i.e. when the prize is sufficiently large and on region* C_{15}*, i.e. when* $\upsilon = 2\varkappa$*.*

- *Every* $\mathcal{D}$ *is* ***private****: either* δ *or* ζ *(or both) are different from zero.*

- *Every $\mathcal{D}$ is **correlated**, except when $(\psi, \phi) \in C_1 \cup C_2 \cup C_{15}$.*
- *Every $\mathcal{D}$ induces a non-trivial hierarchy of beliefs, except when $(\psi, \phi) \in C_1$.*

In the next subsection, we illustrate, with the help of a particularly simple example, some of the features of the characterization presented in this section.

5.2 Example with medium prizes

In this subsection we present the full characterization of a numerical example with particularly nice features. We fix the contest at the parameter values of $\psi = \frac{1}{2}$ and $\phi = \frac{1}{3}$ but allow $k \in K$ to vary freely. This situation corresponds to $(\psi, \phi) = (\frac{1}{2}, \frac{1}{3}) \in C_2$ according to the regions described in proposition 5.4. Although the characterization can be obtained from proposition 5.4, whose proof is in appendix A, we nevertheless provide a direct proof in appendix B by solving the actual linear program implied by the set of Bayes Correlated Equilibria at $(\psi, \phi) = (\frac{1}{2}, \frac{1}{3})$ by explicitly going through the iterations in the simplex algoritihm .

The next result characterizes the optimal information disclosure rule for the designer and its expected payoff.

Proposition 5.6 (Optimal Information structure for the example)**.** *Let $(\psi, \phi) = (\frac{1}{2}, \frac{1}{3})$. Define the family of information disclosure rules $\{\mathcal{S}(\theta) : \theta \in [0, 1]\}$ in which*

$$M_i = \{\textit{Hard-Fought}, \neg\textit{Hard-Fought}\} = \{m^H, \neg m^H\} \quad \textit{for } i = 1, 2,$$

and π is given by

$\pi(\cdot\|\alpha\alpha)$	m^H	$\neg m^H$
m^H	1	0
$\neg m^H$	0	0

$\pi(\cdot\|\alpha\beta)$	m^H	$\neg m^H$
m^H	$\frac{1}{2} + \frac{\theta}{9}$	$\frac{\theta}{18}$
$\neg m^H$	$\frac{1}{2} - \frac{5\theta}{18}$	$\frac{\theta}{9}$

$\pi(\cdot\|\beta\alpha)$	m^H	$\neg m^H$
m^H	$\frac{1}{2} + \frac{\theta}{9}$	$\frac{1}{2} - \frac{5\theta}{18}$
$\neg m^H$	$\frac{\theta}{18}$	$\frac{\theta}{9}$

$\pi(\cdot\|\beta\beta)$	m^H	$\neg m^H$
m^H	1	0
$\neg m^H$	0	0

for any $\theta \in [0, 1]$. Then, there exist $\theta \in [0, 1]$ such that $\mathcal{S}(\theta)$ solves the information design problem:

$$\bar{V} = \max_{\mathcal{D}} \max_{\sigma \in \mathfrak{E}(\mathcal{D})} V(\mathcal{D}, \sigma) = \max_{\sigma \in \mathfrak{E}(\mathcal{S}(\theta))} V(\mathcal{S}(\theta), \sigma).$$

Specifically, we have that

1. *If $\mathrm{d}_\beta > \mathrm{d}_\alpha$ then $\mathcal{S}(0)$ is optimal and gives an expected payoff of*

$$\bar{V} = \alpha + \beta + \frac{3}{4}\mathrm{d}_\alpha + \mathrm{d}_\beta \tag{5.1}$$

2. *If* $\mathrm{d}_\beta < \mathrm{d}_\alpha$ *then* $\mathcal{S}(1)$ *is optimal and gives an expected payoff of*

$$\bar{V} = \alpha + \beta + \frac{5}{6}\mathrm{d}_\alpha + \frac{11}{12}\mathrm{d}_\beta. \tag{5.2}$$

3. *If* $\mathrm{d}_\beta = \mathrm{d}_\alpha$, *then for any* $\theta \in [0, 1]$, *the rule* $\mathcal{S}(\theta)$ *is optimal and gives an expected payoff of*

$$\bar{V} = \alpha + \beta + \frac{7}{4}\mathrm{d}_\beta. \tag{5.3}$$

Example 5.1. We will illustrate proposition (5.6) by specifying some values for the productivities of the players for each case.

1. $\alpha = 3, \beta = 4, k(\alpha)\alpha = 5, k(\beta)\beta = 7$. In this case we have that $\frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} = \frac{3}{2}$. Thus

$$\bar{V} = \frac{23}{2} > V(\mathcal{C}) = \frac{65}{6} > V(\mathcal{N}) = \frac{59}{6}.$$

2. $\alpha = 3, \beta = 4, k(\alpha)\alpha = 5, k(\beta)\beta = 5.5$. In this case we have that $\frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha}\frac{3}{4}$. Thus

$$\bar{V} = \frac{241}{24} > V(\mathcal{C}) = \frac{115}{12} > V(\mathcal{N}) = \frac{103}{12}.$$

3. $\alpha = 3, \beta = 4, k(\alpha)\alpha = 5, k(\beta)\beta = 6$. In this case we have that $\frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} = 1$. Thus

$$\bar{V} = \frac{21}{2} > V(\mathcal{C}) = 10 > V(\mathcal{N}) = 9$$

◄

In the information design problem, it is important to remark again that for any information disclosure rule, we are analyzing the equilibrium behavior that it engenders and then choosing the best equilibrium from the point of view of the designer, as pointed out in section 2.1. We can see this in theorem 5.6 by noticing that to compute the expected payoff of the designer for the optimal rule $\mathcal{S}(\theta)$, we are taking the maximum over the equilibrium set $\mathfrak{E}(\mathcal{S}(\theta))$ of the incomplete information game induced by $\mathcal{S}(\theta)$. Therefore, what our results imply is that *at least one* equilibrium of $\mathfrak{E}(\mathcal{S}(\theta))$ will be the optimal from the perspective of the designer. However, up until this point our results are silent about the behavior of other equilibria in the equilibrium set.

5.2.1 Calculating the hiearchies of beliefs and Characterizing equilibrium set

The optimal information structure $\mathcal{S}(\theta)$ induces a Bayesian Game, which we denote $\Gamma_{\mathcal{S}(\theta)}$. The equilibrium set of $\Gamma_{\mathcal{S}(\theta)}$ is denoted as $\mathfrak{E}(\mathcal{S}(\theta))$. In light of the discussion at the end of the previous paragraph, it is important to obtain the full characterization of the equilibrium set to fully understand how the two players will behave at all possible equilibria engendered by the optimal disclosure rule $\mathcal{S}(\theta)$.

When the designer uses the optimal information disclosure rule $\mathcal{S}(\theta)$, with message spaces $M_i = \{m^H \neg m^H\}$ for each $i = 1, 2$ we can enumerate the types t_i that this information rule induces as follows:

$$\begin{aligned} t_i^1 &= (\alpha, m^H), \quad t_i^2 = (\alpha, \neg m^H), \\ t_i^3 &= (\beta, m^H), \quad t_i^4 = (\beta, \neg m^H). \end{aligned} \tag{5.4}$$

Thus, the joint distribution over $T = T_1 \times T_2$, as calculated from equation (2.5), is given by

$\mathrm{P}_T(t_1, t_2)$	t_2^1	t_2^2	t_2^3	t_2^4
t_1^1	$\frac{1}{4}$	0	$\frac{1}{8} + \frac{\theta}{36}$	$\frac{\theta}{72}$
t_1^2	0	0	$\frac{1}{8} - \frac{5\theta}{72}$	$\frac{\theta}{36}$
t_1^3	$\frac{1}{8} + \frac{\theta}{36}$	$\frac{1}{8} - \frac{5\theta}{72}$	$\frac{1}{4}$	0
t_1^4	$\frac{\theta}{72}$	$\frac{\theta}{36}$	0	0

(5.5)

Thus, for each $\theta \in [0, 1]$, the information disclosure rule $\mathcal{S}(\theta)$ engenders a new larger type space T. As pointed out in proposition (5.6), the value of θ that makes $\mathcal{S}(\theta)$ optimal depends on the configuration of the player's productivities by their ability type. In particular, notice that for when $\theta \in \{0, 1\}$, i.e. the boundary of the interval [0, 1], we get a very different distribution P_T over T:

$\mathrm{P}_T(t_1, t_2)$	t_2^1	t_2^2	t_2^3	t_2^4
t_1^1	$\frac{1}{4}$	0	$\frac{1}{8}$	0
t_1^2	0	0	$\frac{1}{8}$	0
t_1^3	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	0
t_1^4	0	0	0	0

for $\theta = 0$, (5.6a)

$\mathrm{P}_T(t_1, t_2)$	t_2^1	t_2^2	t_2^3	t_2^4
t_1^1	$\frac{1}{4}$	0	$\frac{11}{72}$	$\frac{1}{72}$
t_1^2	0	0	$\frac{1}{18}$	$\frac{1}{36}$
t_1^3	$\frac{11}{72}$	$\frac{1}{18}$	$\frac{1}{4}$	0
t_1^4	$\frac{1}{72}$	$\frac{1}{36}$	0	0

for $\theta = 1$. (5.6b)

Notice that in this example, for the type space and distribution given in equation (5.6a) we can eliminate for each player $i = 1, 2$ the type t_i^4 from the analysis because no player believes with positive probability that this type obtains. Formally speaking, we can use the definition of a *belief closed subset* as defined in definition (2.3) to make this claim rigorous. For the type space in equation (5.6a), the set $W = W_1 \times W_2$ where $W_i = \{t_i^1, t_i^2, t_i^3\}$ for $i = 1, 2$ is a belief closed subset of T. In this case, the players will consider that any type vector in W is possible, but vectors in $T \setminus W$, since they occur with probability 0, will not be deemed possible. Thus, we can use W as the reduced type space for this case.

The hierarchies of beliefs The posterior belief functions $\hat{p}_i : T_i \to \Delta(T_j)$ for each player can be readily computed from the prior P_T.

$\hat{p}_i(\cdot\|t_i)$	t_2^1	t_2^2	t_2^3
t_1^1	$\frac{2}{3}$	0	$\frac{1}{3}$
t_1^2	0	0	1
t_1^3	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

for $\theta = 0$, (5.7a)

$\hat{p}_i(\cdot\|t_i)$	t_2^1	t_2^2	t_2^3	t_2^4
t_1^1	$\frac{3}{5}$	0	$\frac{11}{30}$	$\frac{1}{30}$
t_1^2	0	0	$\frac{2}{3}$	$\frac{1}{3}$
t_1^3	$\frac{1}{3}$	$\frac{4}{33}$	$\frac{6}{11}$	0
t_1^4	$\frac{1}{3}$	$\frac{2}{3}$	0	0

for $\theta = 1$. (5.7b)

For each player i the *first-order beliefs*, which are an element of $\Delta(A_j)$, are

$$\begin{aligned} h_i^1(\cdot|t_1^1) &= \tfrac{2}{3}[\alpha] + \tfrac{1}{3}[\beta] \\ h_i^1(\cdot|t_1^2) &= [\beta] \\ h_i^1(\cdot|t_1^3) &= \tfrac{1}{2}[\alpha] + \tfrac{1}{2}[\beta] \end{aligned} \qquad \text{for } \theta = 0, \tag{5.8a}$$

$$\begin{aligned} h_i^1(\cdot|t_1^1) &= \tfrac{3}{5}[\alpha] + \tfrac{2}{5}[\beta] \\ h_i^1(\cdot|t_1^2) &= [\beta] \\ h_i^1(\cdot|t_1^3) &= \tfrac{5}{11}[\alpha] + \tfrac{6}{11}[\beta] \\ h_i^1(\cdot|t_1^4) &= [\alpha] \end{aligned} \qquad \text{for } \theta = 1. \tag{5.8b}$$

Notice that for $\theta = 0$, both players of type t_i^2 become certain of the state ability vector, whereas for $\theta = 1$, both players of types t_i^2 and t_i^4 become certain of the ability vector.

For each player i the *second-order beliefs* are, which are an element of $\Delta(A_j \times \Delta(A_i))$, are

$$\begin{aligned} h_i^2(\cdot|t_1^1) = &\ \tfrac{2}{3}[(\alpha, \tfrac{2}{3}[\alpha] + \tfrac{1}{3}[\beta])] + \tfrac{1}{3}[(\beta, \tfrac{1}{2}[\alpha] + \tfrac{1}{2}[\beta])] \\ h_i^2(\cdot|t_1^2) = &\ [(\beta, \tfrac{1}{2}[\alpha] + \tfrac{1}{2}[\beta])] \\ h_i^2(\cdot|t_1^3) = &\ \tfrac{1}{4}[(\alpha, \tfrac{2}{3}[\alpha] + \tfrac{1}{3}[\beta])] + \tfrac{1}{4}[(\alpha, [\beta])] + \tfrac{1}{2}[(\beta, \tfrac{1}{2}[\alpha] + \tfrac{1}{2}[\beta])] \end{aligned} \quad \text{for } \theta = 0, \tag{5.9a}$$

$$\begin{aligned} h_i^2(\cdot|t_1^1) = &\ \tfrac{3}{5}[(\alpha, [\beta])] + \tfrac{11}{30}[(\beta, \tfrac{5}{11}[\alpha] + \tfrac{6}{11}[\beta])] + \tfrac{1}{30}[(\beta, [\alpha])] \\ h_i^2(\cdot|t_1^2) = &\ \tfrac{2}{3}[(\beta, \tfrac{5}{11}[\alpha] + \tfrac{6}{11}[\beta])] + \tfrac{1}{3}[(\beta, [\alpha])] \\ h_i^2(\cdot|t_1^3) = &\ \tfrac{1}{3}[(\alpha, \tfrac{3}{5}[\alpha] + \tfrac{2}{5}[\beta])] + \tfrac{4}{33}[(\alpha, [\beta])] + \tfrac{6}{11}[(\beta, \tfrac{5}{11}[\alpha] + \tfrac{6}{11}[\beta])] \\ h_i^2(\cdot|t_1^4) = &\ \tfrac{1}{3}[(\alpha, \tfrac{3}{5}[\alpha] + \tfrac{2}{5}[\beta])] + \tfrac{2}{3}[(\alpha, [\beta])] \end{aligned} \quad \text{for } \theta = 1. \tag{5.9b}$$

The *infinite hierarchy* for each player i can be computed in a similar way by iterating the constructions using the posterior beliefs function $\hat{p}_i$ as shown in subsection 2.3.2.

Notice that although both players at some types became certain of the state given their first-order beliefs, this fact didn't became common knowledge, as can be seen from the second-order hierarchies. In particular, it is at states in which the players are different when they have the possibility of becoming certain of this fact. However, this should not be commonly known. It is precisely for this reason that public information disclosure rules, as the complete information rule or p, q rule, are not optimal. In order to incentivize the players when they are different to work, the designer needs to selectively inform them with some probability. However, this needs to be done privately. It is crucial to have private information because in that way the designer can generate some asymmetry at the higher-order hierarchies, which is needed for the players to work with higher probability when they are different.

The equilibrium set We will focus on symmetric equilibria. The following two propositions characterize the symmetric BNE sets for the cases $\theta = \{0, 1\}$.

Proposition 5.7. *Let $\theta = 0$ so that the optimal information disclosure rule is given by $\mathcal{S}(0)$. For each i, let $\sigma_i(t_i) \in [0,1]$ denote the probability putting effort when player i is of type t_i. Let $\sigma_i = (\sigma_i(t_i^k))_{k=1,\dots,3}$ denote the strategy vector of player i. The set of symmetric BNE for this optimal information disclosure rule is given by*

$$\mathfrak{E}(\mathcal{S}(0)) = \{\sigma = (\sigma_1, \sigma_2) : \sigma_i(t_i^1) \in [\frac{1}{3}, 1], \sigma_i(t_i^2) = 0, \sigma_i(t_i^3) = 1, i = 1, 2.\} \tag{5.10}$$

The unique pure-strategy equilibrium profile σ^*, where $\sigma_i^* = (1,0,1)$ for each $i = 1,2$ is the one that achieves the optimal value for the designer, i.e. $\overline{V} = V(\mathcal{S}(0), \sigma^*)$. Furthermore, for each $\sigma \in \mathfrak{E}(\mathcal{S}(0))$, $V(\mathcal{S}(0), \sigma) > V(\mathcal{C})$, i.e. each equilibrium profile in $\mathfrak{E}(\mathcal{S}(0))$ gives a larger utility for the designer than the complete information disclosure rule.

Proposition 5.8. *Let $\theta = 1$ so that the optimal information disclosure rule is given by $\mathcal{S}(1)$. For each i, let $\sigma_i(t_i) \in [0,1]$ denote the probability putting effort when player i is of type t_i. Let $\sigma_i = (\sigma_i(t_i^k))_{k=1,\dots,4}$ denote the strategy vector of player i. Let $Y = \mathsf{conv}\{(0, \frac{1}{2}), (\frac{2}{11}, 0), (1, 0)\}$. Then the set of symmetric BNE for this optimal information disclosure rule is given by*

$$\mathfrak{E}(\mathcal{S}(1)) = \left\{\sigma = (\sigma_1, \sigma_2) : \left(\sigma_i(t_i^1), \sigma_i(t_i^2)\right) \in Y, \sigma_i(t_i^3) = 1, \sigma_i(t_i^4) = 0, i = 1, 2.\right\} \tag{5.11}$$

The unique pure-strategy equilibrium profile σ^*, where $\sigma_i^* = (1, 0, 1, 0)$ for each $i = 1, 2$ is the one that achieves the optimal value for the designer, i.e. $\overline{V} = V(\mathcal{S}(1), \sigma^*)$. Furthermore, if we define $\tilde{Y} =$ $\mathsf{conv}\{(0, \frac{1}{2}), (\frac{2}{11}, 0)\}$, then for each $\sigma \in \mathfrak{E}(\mathcal{S}(1))$ we have that

- *If σ satisfies that $\left(\sigma_i(t_i^1), \sigma_i(t_i^2)\right) \in Y \setminus \tilde{Y}$ for each $i = 1, 2$ then $V(\mathcal{S}(1), \sigma) > V(\mathcal{C})$.*
- *If σ satisfies that $\left(\sigma_i(t_i^1), \sigma_i(t_i^2)\right) \in \tilde{Y}$ for each $i = 1, 2$. then $V(\mathcal{S}(1), \sigma) = V(\mathcal{C})$.*

5.3 Optimal design when the designer can alter the value of the prize

In this section, we use the results from subsection 5.1 to derive a comparative statics result which is quite pertinent in this context. We ask the question of whether information design is useful when the designer can alter the value of the prize υ, or equivalently, alter the normalized cost of putting effort ϕ.

The rationale for asking whether the designer can gain from information design when altering the value of the prize is two-fold. Firstly, we can think that the designer has no control over the beliefs of the players, embodied by ψ, nor on the abilities as defined in k or the cost of putting effort. At the outset of the game we can interpret the model as a situation in which the designer has no control over the previous parameters. However, the designer has control not only over the information structure of the game via the disclosure rules but also on the value of the prize that he wishes to offer. What happens to the optimal information structure when the value of the prize changes is a very relevant comparative statics exercise.

The second reason comes from the fact that in this model, in the absence of information design, a single prize scheme is optimal. This follows from the result in Moldovanu and Sela (2001, Prop. 1, pp. 547). However, the size of the bursar to be distributed via the prize has a strong impact on altering incentives in the model we are considering as was shown in proposition 4.1. If the value of the prize can be altered and this value is interpreted as a cost to the designer then it is important to know what can be gained through the design of an optimal information disclosure rule.

With the previous motivations in mind, we proceed to describe the profit maximization problem that the designer will face when altering the value of the prize. When the value of the prize can be moved, optimal revenues from the designer when designing information can be expressed as the value achieved from the optimal information disclosure rule. Thus, revenues can be expressed with the help of equation (2.4) as

$$\bar{V}(\upsilon; \varkappa, \psi, k) = \max_{\mathcal{D}} \max_{\sigma \in \mathfrak{S}(\psi, \varkappa/\upsilon, \mathcal{D})} V(\mathcal{D}, \sigma; \psi, \varkappa/\upsilon, k), \tag{5.12}$$

where in $\bar{W}$ now the remaining parameters $(\varkappa, \psi, k)$ that represent respectively the cost of putting effort, the beliefs of the players and the designer and the productivity vector should be considered as parameters. On the other hand, the cost of altering the prize is only the size of the bursar that such a prize represents, i.e. υ.

Therefore, profits are revenues minus costs, $\Pi(\upsilon; \varkappa, \psi, k) = \bar{V}(\upsilon; \varkappa, \psi, k) - \upsilon$. Optimal profits are thus

$$\Pi^*(\varkappa, \psi, k) = \sup_{\upsilon \in (\varkappa, \infty)} \{\bar{V}(\upsilon; \varkappa, \psi, k) - \upsilon\}. \tag{5.13}$$

Notice that in equation (5.13), the supremum is taken over the interval $(\varkappa, \infty)$ since this is in concordance with the restriction that $0 < \phi = \varkappa/\upsilon < 1$ discussed in section 2.

The next proposition follows trivially from proposition 4.1

Proposition 5.9. *Maximal revenues are achieved when $\hat{\upsilon} = 2\varkappa/\psi$ and remain constant for any prize greater than $\hat{\upsilon}$ and for all $\varkappa, \psi$ and $k \in K$. Furthermore, the value $\bar{V}(\hat{\upsilon}, \varkappa, \psi, k)$ is achieved by the null information disclosure rule.*

The simple reason for the statement in proposition 5.9 is that at values greater or equal to $\hat{\upsilon}$, we have that it becomes a dominant strategy to work at all possible states under null information. Thus, in terms of revenues, $\bar{V}(\hat{\upsilon}, \varkappa, \psi, k)$ is the upper bound that we can achieve when altering the prize and under the possibility of designing information, since the optimal disclosure rule can involve saying nothing to the players.

However, as $\varkappa$ changes we have that

$$\frac{\partial \hat{\upsilon}}{\partial \varkappa} = \frac{2}{\psi} > 2.$$

Therefore, an increase in $\varkappa$ needs an increase in the prize by more than two-fold to achieve the upper bound in revenues. In particular, when the probability that each player ascribes to their rival being weak, ψ, tends to zero, the value of the prize that is needed grows without bound. Hence, a very large prize may not be optimal since its benefits may not outweigh its costs. It may be the case that a much smaller prize can be better under an optimal information disclosure rule.

On the other hand, we have that as ψ changes

$$\frac{\partial \hat{\upsilon}}{\partial \psi} = -\frac{2\varkappa}{\psi^2} < 0,$$

so we have that when the players believe with high probability that their rival is weak, the value of the prize required to achieve the upper bound in revenue decreases.

From the discussion above it follows that we will need to compare the performance of the null information scheme given by $\Pi(\hat{\upsilon}; \varkappa, \psi, k) = \bar{V}(\hat{\upsilon}; \varkappa, \psi, k) - \hat{\upsilon}$ to any other different scheme that we are considering. In summary, $\Pi(\hat{\upsilon}; \varkappa, \psi, k)$ provides the threshold on υ at which the null information disclosure rule can become optimal if such value is equal to the supremum in equation (5.13). Therefore, we will look for conditions on the remaining parameters $(\varkappa, \psi, k)$ which imply an optimal prize less than $\hat{\upsilon}$.

For a clean statement of the results, define $L : (0, 1) \to \mathbb{R}$ as

$$L(\psi) = 1 - 2\psi + 3\psi^2 + \sqrt{1 - 4\psi + 10\psi^2 - 4\psi^3 + \psi^4}. \tag{5.14}$$

The function $L(\psi)$ represents the boundary of a region, as a function of the beliefs of the players ψ, that the remaining parameters $\varkappa$ and k need to satisfy in order for information design to be optimal at prize values lower than the one implied $\hat{\upsilon}$ where the optimal information rule involves saying nothing to the players.

The results are presented in the next two theorems. Theorem 5.10 deals with the parameter conditions at which the optimal disclosure rule involves non-trivial communication with the players. Theorem 5.11 complements the previous theorem by describing the optimal disclosure rules that achieve the optimal profits.

Theorem 5.10. *If productivity is regular we have that*

1. *If the parameters $(\varkappa, \psi, k)$ satisfy the condition: $0 < \varkappa - \mathrm{d}_\alpha \psi^2 \leq \left(\frac{\mathrm{d}_\beta - \mathrm{d}_\alpha}{4}\right) L(\psi)$, then the optimal prize scheme is*

$$\upsilon^* = \frac{4\varkappa(1-\psi)}{L(\psi) - 4\psi^2} < \hat{\upsilon}. \tag{5.15}$$

2. *If the parameters $(\varkappa, \psi, k)$ satisfy the condition: $\left(\frac{\mathrm{d}_\beta - \mathrm{d}_\alpha}{4}\right) L(\psi) \leq \varkappa - \mathrm{d}_\alpha \psi^2$, then the optimal prize scheme is*

$$\upsilon^* = 2\varkappa < \hat{\upsilon}. \tag{5.16}$$

3. *If the parameters $(\varkappa, \psi, k)$ satisfy the condition: $\varkappa - \mathrm{d}_\alpha \psi^2 \leq 0$, then the optimal prize scheme is $\hat{\upsilon}$.*

If productivity is non-regular, then

4. *If the parameters $(\varkappa, \psi, k)$ satisfy the condition: $0 < \frac{2\varkappa}{\psi^2} < \mathrm{d}_\alpha + \mathrm{d}_\beta$, then the optimal prize scheme is*

$$\upsilon^* = 2\varkappa < \hat{\upsilon}. \tag{5.17}$$

5. *If the parameters $(\varkappa, \psi, k)$ satisfy the condition: $\frac{2\varkappa}{\psi^2} \geq \mathrm{d}_\alpha + \mathrm{d}_\beta$ then the optimal prize scheme is $\hat{\upsilon}$.*

In cases 1,2 and 4 we have that $\Pi^*(\varkappa, \psi, k) = \Pi(\upsilon^*, \varkappa, \psi, k) > \Pi(\hat{\upsilon}; \varkappa, \psi, k)$. On the other hand, in cases 3 and 5 we have that $\Pi^*(\varkappa, \psi, k) = \Pi(\hat{\upsilon}; \varkappa, \psi, k)$.

Thus, as long as $\varkappa - \mathrm{d}_\alpha \psi^2 \geq 0$ for regular productivity or $\frac{2\varkappa}{\psi^2} < \mathrm{d}_\alpha + \mathrm{d}_\beta$ for non-regular productivity, non-trivial information design is optimal.

Theorem 5.11. *If the parameters satisfy the conditions* $\varkappa - \mathrm{d}_\alpha \psi^2 \geq 0$ *for regular productivity or* $\frac{2\varkappa}{\psi^2} < \mathrm{d}_\alpha + \mathrm{d}_\beta$ *for non-regular productivity then the optimal information disclosure rule* $\mathcal{D}^* = ((M_i)_{i=1,2}, \pi^*)$ *has*

$$M_i\{\textit{Aggressive, Not Aggressive}\} = \{m^A, \neg m^A\} \quad \textit{for } i = 1, 2,$$

and $\pi^* : A \to \Delta(M)$ *is as follows:*

- *For the regular productivity case in which* $\left(\frac{\mathrm{d}_\beta - \mathrm{d}_\alpha}{4}\right) L(\psi) \leq \varkappa - \mathrm{d}_\alpha \psi^2$ *or the non-regular productivity case in which* $0 < \frac{2\varkappa}{\psi^2} < \mathrm{d}_\alpha + \mathrm{d}_\beta$ *(cases 2 and 4 in theorem 5.10 above) we have that*

$$\pi^*(\cdot|\alpha\alpha) = \pi^*(\cdot|\beta\beta) = [(m^H, m^H)],$$
$$\pi^*(e|\alpha\beta) = \pi^*(e|\beta\alpha) = 1/4, \ \forall m \in M.$$

- *For the regular productivity case in which* $\left(\frac{\mathrm{d}_\beta - \mathrm{d}_\alpha}{4}\right) L(\psi) \leq \varkappa - \mathrm{d}_\alpha \psi^2$ *(case 1 in theorem 5.10) we have that*

$$\pi^*(\cdot|\alpha\alpha) = \pi^*(\cdot|\beta\beta) = [(m^H, m^H)],$$
$$\pi^*(m^H, m^H|\alpha\beta) = \pi^*(m^H, m^H|\beta\alpha) = \frac{2\psi}{L(\psi) - 4\psi^2} - \frac{\psi}{1-\psi}$$
$$\pi^*(\neg m^H, m^H|\alpha\beta) = \pi^*(m^H, \neg m^H|\beta\alpha) = \frac{1}{1-\psi} - \frac{2\psi}{L(\psi) - 4\psi^2}$$

6 Concluding remarks

This paper analyzes how a designer can use information to manipulate the beliefs in a contest in which there is incomplete information about the abilities of the contestants. We found that the optimal disclosure rules make the crucial distinction between public and private information and in general always use partial revelation of information. In particular, we showed how private information generates asymmetric hierarchies of beliefs which are necessary for the designer to benefit from information design. We also performed a cost-benefit analysis of optimal information design rule by introducing a "price for information" and taking advantage of the full characterization results that we provided. We found that for a robust set of parameters, information design is the most cost-efficient way to manipulate the players to act following the interests of the contest designer.

Throughout the analysis of optimal information disclosure rules, we have assumed that the designer was able to pick an equilibrium t hat is to her liking from the equilibrium set engendered by an information disclosure rule. This equilibrium selection criterion is embedded into the revelation principle-style argument that provides the foundation to use the notion of Bayes Correlated Equilibrium as a first step into computing the optimal information disclosure rule for the designer. Generically, the optimal information disclosure rules for contests identified by this method engender a continuum of equilibria. Although we showed within the context of a particular example that all equilibria engendered by such a rule perform at least as good as the complete information disclosure rule from the point of view of the designer, it remains an important question if we can refine these results. An important question is to analyze the performance of optimal disclosure rules when we relax the *optimistic* equilibrium selection criterion for another one. In particular, we would like to know what kind of optimal rules arise when we replace the optimistic attitude of the designer by a *pessimistic one* in which she now considers that the worst equilibrium is the one that is going to be picked by the players. In this sense, for any disclosure rule that the designer is considering, this pessimistic criteria ensures that any equilibria in the equilibrium set engendered by the disclosure rules perform at least as good as the *worst-case scenario*.

Tied to the previous notion is the issue about uniqueness of equilibrium under a particular information disclosure rule. In the paper we identified two information disclosure rules, the null and complete rules, that have unique Bayes Nash Equilibrium profiles generically. However, in general these rules were far from optimal. If the designer could find an information disclosure rule that performed better than these two and at the same time engendered a unique equilibrium profile, then she would not need to worry about the possibility of other bad equilibrium arising under such a rule.

A general solution the this problem can be obtained by extending the techniques in Mathevet et al. (2016) to the case in which the players hold prior private information, just like they do in the contests considered here. However, concurrent work to this paper shows that when the designer has a pessimistic attitude, the candidate optimal rule that emerges turns out to be public. In particular the optimal public information disclosure rule turns out to engender generically a unique equilibrium. This is in sharp contrast to the results of this paper and points out the value of public information disclosure in contests as a mean to incentivize unique equilibrium profiles. An open question remains about how general this statement is for other classes of games and designer's objectives.

A second extension is the case of differential information between the designer and the players while still satisfying the assumption of distributed knowledge between the players and the designer. When the designer also has private information, any choice of information disclosure rule runs the risk of informing the players about what the designer knows, and thus impacting the power of the designer to guide the players to her desire course of action. Therefore, further restrictions need to be imposed on the choice of the information disclosure rule in order to counteract this effect. This consideration ties in with the problem of the informed principal in the mechanism design literature.

A final extension that is interesting to analyze is the case of prior private information between the players which does not satisfy the distributed knowledge requirement. For example, suppose that the incomplete information in the game is not only described by the ability vector, whose individual components are assumed to be known privately by the players, but the uncertainty in the contest also comes from a random component in the value of the prize of which the contestants only receive a partially correlated signal. A real-life situation that could be modeled by this environment would be Research and Development races, in which the value of the invention is initially unknown to the players. In this case, a designer perhaps can disclose information about the value of the prize, the contestants' ability or both. In this case, although the techniques from Bergemann and Morris (2016) still apply, it is nevertheless too artificial since the designer would need to know information that the players do not jointly possess.

So, this is how the story ends. The hare never understood what happened that day. The fox, as a designer, was the real winner. Now, we can extract a new moral from the famous fable: *if you want others to behave as you wish, don't say more than you need, and just wait until everything is on its stead.*

A Finding an optimal Bayes Correlated Equilibrium for the Designer

A.1 Simplifying the BCE optimization program

In section 3 we described the steps in which the information design problem (2.4) could be simplified to the search of an optimal Bayes Correlated Equilibrium distribution λ in view of theorem 3.2. Equation (3.4) formalizes the previous claim by showing that the optimal value of the information design problem es equivalent to the value given by the optimal BCE decision rule λ.

For any $(\psi, \phi) \in C$, a decision rule $\lambda : A \to \Delta(E)$ as in definition 3.1 is a four-tuple whose components come from the three-dimensional probability simplex simplex Δ_3, i.e. $\lambda \in \prod_{i=1}^{4} \Delta_3 = \Delta_3^4$. Problem (3.4) is thus:

$$\max_{\lambda \in \Delta_3^4} V(\lambda) \tag{A.1}$$

$$\text{subject to:}$$

$$\sum_{e_j, a_j} prob(a_j | a_i) \lambda(e_i, e_j | a) \big(\hat{u}_i(e_i, e_j, a) - \hat{u}_i(e_i', e_j, a) \big) \geq 0, \tag{A.2}$$

$$\forall i, j = 1, 2, i \neq j, \ \forall a_i \in A_i, \forall e_i, e_i' \in E_i, e_i \neq e_i'.$$

There are some features that we wish to point out about the previous program. We have that $\Delta_3^4 \subset \mathbb{R}^{16}$. If $\lambda \in \Delta_3^4$, then it needs to satisfy four *equality constraints* that ensure that for each $a \in A$, $\sum_{e \in E} \lambda(e|a) = 1$. Each one of the equality constraints can be represented by two paired inequality constraints, for a total of 8. Furthermore, there are 16 non-negativity constraints, i.e. $\lambda(e|a) \geq 0$ for each $a \in A$ and $e \in E$. Additionally the set of inequalities (A.2) describe the *obedience* constraints implied by the BCE concept, which is also linear in λ. Furthermore we have $|E_i| \times |A_i| = 4$ constraints per player for a total of 8. Therefore the program has 32 inequality constraints in 16 variables. The objective (A.1) is linear in λ. Using these observations, we conclude that the previous program is a *linear programming problem.*

The objective of this appendix is to provide the details on how to solve this program.

First of all, notice that we can reduce the number of variables to consider by noticing that equality constraints that define the four-fold product of the three dimensional simplex ensure that in $\lambda(\cdot|a)$ one of the numbers is determined by the other three for each $a \in A$, which would reduce the number of variables to 12, with 4 inequality constraints representing the 4-fold probability simplices, 12 non-negativity constraints, and 8 obedience constraints for a total of 24 constraints.

To simplify further the problem and make it tractable for analysis, we will focus on ***symmetric Bayes Correlated Equilibria***, hereafter SBCE. It is obvious that SBCE $\subseteq$ BCE, however, it will be shown that the optimal BCE for the designer actually occurs at a symmetric BCE. The set of SBCE can be parameterized as shown in figure 5.

Let $\boldsymbol{\lambda} = (\gamma_\alpha, \xi_\alpha, \gamma_\beta, \xi_\beta, \mu, \delta, \zeta)^\intercal \in \mathbb{R}^7$ represent the parameterized decision rule $(\lambda(\cdot|a))_{a \in A} \in \Delta_3^4$ as in figure 5. For this parameterization to be a valid system of conditional probability distributions

$\lambda(\cdot\|\alpha\alpha)$	W	S
W	γ_α	$\xi_\alpha - \gamma_\alpha$
S	$\xi_\alpha - \gamma_\alpha$	$\gamma_\alpha - 2\xi_\alpha + 1$

$\lambda(\cdot\|\alpha\beta)$	W	S
W	μ	δ
S	ζ	$1-\mu-\delta-\zeta$

$\lambda(\cdot\|\beta\alpha)$	W	S
W	μ	ζ
S	δ	$1-\mu-\delta-\zeta$

$\lambda(\cdot\|\beta\beta)$	W	S
W	γ_β	$\xi_\beta - \gamma_\beta$
S	$\xi_\beta - \gamma_\beta$	$\gamma_\beta - 2\xi_\beta + 1$

Figure 5: Symmetric BCE—Parameterized decision rule $\boldsymbol{\lambda}$

we need the following conditions to hold

$$2\xi_i - 1 \le \gamma_i \le \xi_i, \text{ for } i \in \{\alpha, \beta\}, \tag{A.3a}$$

$$\delta + \zeta + \mu \le 1, \tag{A.3b}$$

$$\xi_i \ge 0, \gamma_i \ge 0, \text{ for } i = 1, 4, \tag{A.3c}$$

$$\mu \ge 0, \delta \ge 0, \zeta \ge 0. \tag{A.3d}$$

In a SBCE, the two players are treated symmetrically, so the 8 constraints in (A.2), 4 for each player, are identical across players. Hence, we can consider only four of them. Rewriting the four obedience constraints using $\boldsymbol{\lambda}$ yields:

$$-\frac{1}{2}(1-2\phi)\psi\xi_\alpha + \mu\phi(1-\psi) - \delta(1-\phi)(1-\psi) \le 0, \tag{A.4a}$$

$$-\frac{1}{2}(1-2\phi)(1-\psi)\xi_\beta - \mu(1-\phi)\psi + \zeta\phi\psi \le 0, \tag{A.4b}$$

$$-\frac{1}{2}(1-2\phi)\psi\xi_\alpha - \mu(1-\phi)(1-\psi) - \delta(1-\phi)(1-\psi) - \zeta(1-\psi) \le \frac{1}{2}(2\phi+\psi-2), \tag{A.4c}$$

$$-\frac{1}{2}(1-2\phi)(1-\psi)\xi_\beta + \mu\phi\psi + \delta\psi + \zeta\phi\psi \le \frac{1}{2}(2\phi+\psi-1). \tag{A.4d}$$

If a player low ability (α) is told to work (W), then he will do it if constraint (A.4a) holds. Similarly, if a player of high ability (β) is told to work, he will do it if (A.4b) holds. On the other hand, a player of low ability who is told to shirk, will do so if constraint (A.4c) holds. Finally, a player of high ability who is told to shirk will do it if constraint (A.4d) holds. Partition $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)^\intercal$ where $\boldsymbol{\lambda}_1 = (\gamma_\alpha, \xi_\alpha, \gamma_\beta, \xi_\beta)^\intercal$ and $\boldsymbol{\lambda}_2 = (\mu, \delta, \zeta)^\intercal$. Define the matrix $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2]$ as the one that incorporates the left-hand side of the obedience constraints (A.4) and the simplex constraints (A.3a) and (A.3b), i.e. without taking into consideration the non-negativity constraints and $\boldsymbol{\eta}$ as the respective right-hand side, where $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are the component matrices that correspond to the previous partition of $\boldsymbol{\lambda}$

$$
\boldsymbol{\Omega}_1 = \begin{bmatrix} 0 & \left(\phi - \frac{1}{2}\right)\psi & 0 & 0 \\ 0 & 0 & 0 & \left(\phi - \frac{1}{2}\right)(1-\psi) \\ 0 & \left(\phi - \frac{1}{2}\right)\psi & 0 & 0 \\ 0 & 0 & 0 & \left(\phi - \frac{1}{2}\right)(1-\psi) \\ -1 & 2 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{A.5}
$$

$$
\boldsymbol{\Omega}_2 = \begin{bmatrix} \phi(1-\psi) & -(1-\phi)(1-\psi) & 0 \\ -(1-\phi)\psi & 0 & \phi\psi \\ -(1-\phi)(1-\psi) & -(1-\phi)(1-\psi) & -(1-\psi) \\ \phi\psi & \psi & \phi\psi \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{A.6}
$$

$$
\boldsymbol{\eta} = \begin{pmatrix} 0 \\ 0 \\ -(1-\psi)(1-\phi) + \psi\left(\phi - \frac{1}{2}\right) \\ (1-\psi)\left(\phi - \frac{1}{2}\right) + \psi\phi \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \tag{A.7}
$$

Notice that both $\boldsymbol{\Omega}$ and $\boldsymbol{\eta}$ depend on $\psi, \phi \in C$, but we supress the dependency for simplicity.

At this point, it will be convenient to introduce the following notation: let $\mathrm{d}_\alpha = k(\alpha)\alpha - \alpha$ and $\mathrm{d}_\beta = k(\beta)\beta - \beta$ denote the marginal productivity of putting effort by the low ability and high ability player respectively. We can use the previous notation and the parameterized decision rule $\boldsymbol{\lambda}$ to rewrite the expected payoff of the designer given by equation (3.3) as follows:

$$
2\psi^2 \mathrm{d}_\alpha \xi_1 + 2(1-\psi)^2 \mathrm{d}_\beta \xi_4 + 2\psi(1-\psi)\big((\mathrm{d}_\alpha + \mathrm{d}_\beta)\mu + \mathrm{d}_\alpha\delta + \mathrm{d}_\beta\zeta\big) + 2\big(\alpha\psi + \beta(1-\psi)\big). \tag{A.8}
$$

Define the vector $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)^\intercal$, where the subindexes correspond to the partition of $\boldsymbol{\lambda}$ and where

$\boldsymbol{\rho}_1 = (0, \psi^2 \mathrm{d}_\alpha, 0, (1-\psi)^2 \mathrm{d}_\beta)^\intercal$ and $\boldsymbol{\rho}_2 = (\mathrm{d}_\alpha + \mathrm{d}_\beta, \mathrm{d}_\alpha, \mathrm{d}_\beta)^\intercal$. Then (A.8) can be written as

$$2\boldsymbol{\rho}_1^\intercal \boldsymbol{\lambda}_1 + 2\psi(1-\psi)\boldsymbol{\rho}_2^\intercal \boldsymbol{\lambda}_2 + 2\big(\alpha\psi + \beta(1-\psi)\big). \tag{A.9}$$

We conclude that the objective (A.1) subject to the set of constraints (A.2) is equivalent to the following program.

$$\begin{gathered} \bar{V} = \max_{\boldsymbol{\lambda} \in \mathbb{R}^7} \left\{2\boldsymbol{\rho}_1^\intercal \boldsymbol{\lambda}_1 + 2\psi(1-\psi)\boldsymbol{\rho}_2^\intercal \boldsymbol{\lambda}_2 + 2\big(\alpha\psi + \beta(1-\psi)\big)\right\} \\ \text{subject to} \\ \boldsymbol{\Omega}_1 \boldsymbol{\lambda}_1 + \boldsymbol{\Omega}_2 \boldsymbol{\lambda}_2 \leq \boldsymbol{\eta} \text{ and } \boldsymbol{\lambda} \geq \mathbf{0}. \end{gathered} \tag{A.10}$$

Remark: $\bar{V}$ also depends on (ψ, ϕ, k) but we suppress the dependency for simplicity. [16]

A.2 Further simplifying reductions and auxiliary results

Inspection of the objective (A.9) and the obedience constraints (A.4) yields the following lemma, whose proof is trivial.

Lemma A.1. *In an optimal SBCE for the designer, when the players are similar, the probability that both work, γ_j and the probability that only one of them works, ξ_j are equal, i.e. $\gamma_j = \xi_j$ for $j \in \{\alpha, \eta\}$. Thus, in an optimal SBCE, when the players are similar either they are both working or both shirking or mixing between these two cases.*

Proof. Notice that γ_j, $j \in \{\alpha, \eta\}$ does not enter into the objective (A.9) nor appears in the constraints (A.4). The only place that these two variables appear is in the simplex constraints (A.3a) and the non-negativity constraints (A.3c). Thus, without loss of generality, we can put $\gamma_j = \xi_j$ for $j \in \{\alpha, \eta\}$ without altering the optimal value of the problem and without affecting the rest of the feasible set. ∎

Lemma A.1 yields the following corollaries.

Corollary A.2. *The simplex constraints* (A.3a) *pertaining to the states when the players are similar, i.e. $a \in \{\alpha\alpha, \beta\beta\}$, can be compressed into the single constraints $\xi_j \leq 1$ for $j \in \{\alpha, \beta\}$.*

Corollary A.3. *Define the reduced vector $\hat{\boldsymbol{\lambda}} = (\hat{\boldsymbol{\lambda}}_1, \boldsymbol{\lambda}_2)$ where $\hat{\boldsymbol{\lambda}}_1 = (\xi_\alpha, \xi_\beta)^\intercal$ and $\boldsymbol{\lambda}_2 = (\mu, \delta, \zeta)^\intercal$ remains the same as before. Furthermore, define $\hat{\boldsymbol{\rho}} = (\hat{\boldsymbol{\rho}}_1, \boldsymbol{\rho}_2)^\intercal$ where $\hat{\boldsymbol{\rho}}_1 = (\psi^2 \mathrm{d}_\alpha, (1-\psi)^2 \mathrm{d}_\beta)^\intercal$ and $\boldsymbol{\rho}_2$ remains as before. Finally, define the partitioned matrix $\hat{\boldsymbol{\Omega}} = [\hat{\boldsymbol{\Omega}}_1, \hat{\boldsymbol{\Omega}}_2]$ representing the reduced left-hand side obedience constraints and the reduced vector $\hat{\boldsymbol{\eta}}$ of the corresponding right-hand side as*

[16] Notation: if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

- $\mathbf{x} \geq \mathbf{y} \Leftrightarrow x_i \geq y_i, i = 1, \ldots, n$.
- $\mathbf{x} > \mathbf{y} \Leftrightarrow x_i \geq y_i, i = 1, \ldots, n$ and $\mathbf{x} \neq \mathbf{y}$.
- $\mathbf{x} \gg \mathbf{y} \Leftrightarrow x_i > y_i, i = 1, \ldots, n$.

follows:

$$\hat{\boldsymbol{\Omega}}_1 = \begin{bmatrix} \left(\phi - \frac{1}{2}\right)\psi & 0 \\ 0 & \left(\phi - \frac{1}{2}\right)(1-\psi) \\ \left(\phi - \frac{1}{2}\right)\psi & 0 \\ 0 & \left(\phi - \frac{1}{2}\right)(1-\psi) \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \tag{A.11}$$

$$\hat{\boldsymbol{\Omega}}_2 = \begin{bmatrix} \phi(1-\psi) & -(1-\phi)(1-\psi) & 0 \\ -(1-\phi)\psi & 0 & \phi\psi \\ -(1-\phi)(1-\psi) & -(1-\phi)(1-\psi) & -(1-\psi) \\ \phi\psi & \psi & \phi\psi \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \tag{A.12}$$

$$\hat{\boldsymbol{\eta}} = \begin{pmatrix} 0 \\ 0 \\ -(1-\psi)(1-\phi) + \psi\left(\phi - \frac{1}{2}\right) \\ (1-\psi)\left(\phi - \frac{1}{2}\right) + \psi\phi \\ 1 \\ 1 \\ 1 \end{pmatrix}. \tag{A.13}$$

Define the correspondence $\Upsilon : C \rightrightarrows \mathbb{R}^5$ as the solution set of the constraints, i.e.

$$\Upsilon(\psi,\phi) = \{\hat{\boldsymbol{\lambda}} = (\hat{\boldsymbol{\lambda}}_1, \boldsymbol{\lambda}_2) \in \mathbb{R}^5 \mid \hat{\boldsymbol{\Omega}}_1\hat{\boldsymbol{\lambda}}_1 + \hat{\boldsymbol{\Omega}}_2\boldsymbol{\lambda}_2 \leq \hat{\boldsymbol{\eta}},\ \hat{\boldsymbol{\lambda}} \geq \mathbf{0}.\} \tag{A.14}$$

Then, problem (A.10) *is equivalent to:*

$$\bar{V} = \max_{\hat{\boldsymbol{\lambda}} \in \Upsilon(\psi,\phi)} \left\{ 2\hat{\boldsymbol{\rho}}_1^\intercal \hat{\boldsymbol{\lambda}}_1 + 2\psi(1-\psi)\boldsymbol{\rho}_2^\intercal \boldsymbol{\lambda}_2 + 2\big(\alpha\psi + \beta(1-\psi)\big) \right\} \tag{A.15}$$

As mentioned before, the feasible set $\Upsilon(\psi,\phi)$ is a convex and compact set since it is a polytope.[17]

Since $\Upsilon(\psi,\phi)$ describes a convex polytope, then it can potentially be described by the convex hull of its vertices. The set of vertices of $\Upsilon(\psi,\phi)$ is the set of its *extremal points.*[18]

Thus, for any (ψ,ϕ) let $\mathsf{Ext}(\Upsilon(\psi,\phi))$ denote the finite set of extremal points of $\Upsilon(\psi,\phi)$, i.e. $\Upsilon(\psi,\phi) = \mathsf{conv}\big(\mathsf{Ext}(\Upsilon(\psi,\phi))\big)$.

[17] It is a bounded polyhedron, since it is described by a finite set of linear inequalities and it does not contain a ray.

[18] If K is a convex subset of $\mathbb{R}^n$, then a point $x \in K$ is an extreme point of K provided that $y, z \in K$, $0 < \theta < 1$, and $x = \theta y + (1-\theta)z$ imply $x = y = z$.

However, trying to search for all possible vertices of $\Upsilon(\psi, \phi)$ is a daunting task, since we have the following bounds (Salomone, Vaisman, and Kroese, 2016; Eckhoff, Gruber, and Wills, 1993):

1. By the McMullen upper bound theorem; the number of vertices $|\mathrm{Ext}(\Upsilon(\psi, \phi))|$ is at most

$$\binom{n - \lfloor (d+1)/2 \rfloor}{n-d} + \binom{n - \lfloor (d+2)/2 \rfloor}{n-d} = 2\binom{12}{7} = 72,$$

   where the number of half-spaces that define $\Upsilon(\psi, \phi)$ is $n = 12$, since we have 7 inequalities plus 4 non-negativity constraints, and $d = 5$ since $\hat{\boldsymbol{\lambda}} \in \mathbb{R}^5$.

2. By the Barnette lower bound theorem, the number of vertices $|\mathrm{Ext}(\Upsilon(\psi, \phi))|$ is at least

$$n(d-1) - (d+1)(d-2) = 30,$$

   where n and d are as before.

Thus, trying to find the set of extremal points of $\Upsilon(\psi, \phi)$ for all $(\psi, \phi) \in C$, and then evaluate the objective function at each of them in order to compare them is an intractable approach in general.

However, there is one situation in which we can completely describe $\Upsilon(\psi, \phi)$ and that is when $\phi = 1/2$. The following lemma records this observation.

Lemma A.4. *Let $\phi = 1/2$. Then, for any $\psi \in (0, 1)$, we have*

$$\Upsilon((\psi, 1/2)) = \{\hat{\boldsymbol{\lambda}} \in \mathbb{R}^5 | 0 \leq \xi_\alpha \leq 1, 0 \leq \xi_\beta \leq 1, \mu = \delta = \zeta = 1/4\}.$$

Proof. We start by substituting $\phi = 1/2$ in $\hat{\boldsymbol{\Omega}}_1$, $\hat{\boldsymbol{\Omega}}_2$ and $\hat{\boldsymbol{\eta}}$. After the substitution, notice that the upper portion of $\hat{\boldsymbol{\Omega}}_1$ vanishes, which implies that the only constraints affecting $\hat{\boldsymbol{\lambda}}_1 = (\xi_\alpha, \xi_\beta)$ are $0 \leq \xi_\alpha \leq 1$ and $0 \leq \xi_\beta \leq 1$.

On the other hand, we have after some simplifications the following constraints on $\boldsymbol{\lambda}_2 = (\mu, \delta, \zeta)$

$$\mu \leq \delta, \tag{A.16a}$$

$$\zeta \leq \mu, \tag{A.16b}$$

$$-\mu - \delta - 2\zeta \leq -1, \tag{A.16c}$$

$$\mu + 2\delta + \zeta \leq 1, \tag{A.16d}$$

$$\mu + \delta + \zeta \leq 1. \tag{A.16e}$$

Inequalities (A.16a) and (A.16b) imply that $\zeta \leq \mu \leq \delta$. Adding up inequalities (A.16c) and (A.16d) yields $\delta \leq \zeta$. The previous two inequalities implies that $\mu = \delta = \zeta$, which together with (A.16e) and the non-negativity constraints implies that these three numbers must equal $1/4$, which concludes the proof.

■

Careful observation of the objective in (A.15) and the matrix $\hat{\boldsymbol{\Omega}}_1$ yields further simplifications, which are recorded in the following lemmas.

Lemma A.5. *Let* $\hat{\boldsymbol{\lambda}}_1^* = (\xi_\alpha^*, \xi_\beta^*) = (1,1)$. *Define the correspondence* $\hat{\Upsilon}(\psi,\phi) : C \rightrightarrows \mathbb{R}^3$ *as*

$$\hat{\Upsilon}(\psi,\phi) = \{\boldsymbol{\lambda}_2 \in \mathbb{R}^3 \mid \hat{\boldsymbol{\Omega}}_2\boldsymbol{\lambda}_2 \leq \hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\Omega}}_1\hat{\boldsymbol{\lambda}}_1^*, \boldsymbol{\lambda}_2 \geq \mathbf{0}\}, \tag{A.17}$$

which represents the solution set of the reduced system of constraints. *Then* $\hat{\Upsilon}(\psi,\phi)$ *is non-empty for all* $(\psi,\phi) \in \{(\psi,\phi) \in C | \phi \leq 1/2\}$.

Proof. We need to show that the correspondence (A.17) is non-empty, i.e. it represents a valid system of inequalities. Towards this end, it is enough to show that it contains at least a point for each $(\psi,\phi) \in \{(\psi,\phi) \in C | \phi \leq 1/2\}$. First of all, notice that

$$\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\Omega}}_1\hat{\boldsymbol{\lambda}}_1^* = \begin{pmatrix} \psi\left(\frac{1}{2}-\phi\right) \\ (1-\psi)\left(\frac{1}{2}-\phi\right) \\ -(1-\psi)(1-\phi) \\ \psi\phi \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Since $(\psi,\phi) \in (0,1) \times (0,1/2]$ we have that $\psi\left(\frac{1}{2}-\phi\right) \geq 0$, $(1-\psi)\left(\frac{1}{2}-\phi\right) \geq 0$, $-(1-\psi)(1-\phi) < 0$ and $\psi\phi > 0$.

Now, consider the following point $\boldsymbol{\lambda}_2' = (\mu,\delta,\zeta)^\intercal = (\phi(1-\phi), \phi^2, (1-\phi)^2)^\intercal$, which we will show satisfies (A.17).[19] We have that after some simplifications

$$\hat{\boldsymbol{\Omega}}_2\boldsymbol{\lambda}_2' = \begin{bmatrix} 0 \\ 0 \\ -(1-\psi)(1-\phi) \\ \psi\phi \\ 0 \\ 0 \\ 1-\phi(1-\phi) \end{bmatrix}.$$

Noticing that $1-\phi(1-\phi) \leq 3/4$ if $\phi \leq 1/2$, it is easily seen that $\hat{\boldsymbol{\Omega}}_2\boldsymbol{\lambda}_2' \leq \hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\Omega}}_1\hat{\boldsymbol{\lambda}}_1^*$. Finally, it is easily seen that $\boldsymbol{\lambda}_2' \geq \mathbf{0}$ which completes the proof. ■

Lemma A.6. *Suppose* $\phi \leq 1/2$ *and let* $\hat{\boldsymbol{\lambda}}_1^* = (1,1)$ *as in lemma* (A.5). *Recall that* $\bar{V}$ *is the optimal value*

[19] The point $\boldsymbol{\lambda}_2'$ is part of the equilibrium distribution induced by the complete information disclosure rule $\mathcal{C}$, as described in 4.3.

of problem (A.15). *Then*

$$\bar{V} = \max_{\boldsymbol{\lambda}_2 \in \hat{\Upsilon}(\psi,\phi)} \left\{ 2\hat{\boldsymbol{\rho}}_1^{\mathsf{T}} \hat{\boldsymbol{\lambda}}_1^* + 2\psi(1-\psi)\boldsymbol{\rho}_2^{\mathsf{T}}\boldsymbol{\lambda}_2 + 2\big(\alpha\psi + \beta(1-\psi)\big) \right\} \tag{A.18}$$

That is, $\hat{\boldsymbol{\lambda}}_1^*$ *is part of an optimal solution.*

Proof. First of all, notice that in (A.15), the coefficient attached to $\hat{\boldsymbol{\lambda}}_1$, $\hat{\boldsymbol{\rho}}_1 = (\psi^2 \mathrm{d}_\alpha, (1-\psi)^2 \mathrm{d}_\beta)^{\mathsf{T}}$, has only positive components. Therefore, increasing the value of $\hat{\boldsymbol{\lambda}}_1$ as long as it is feasible increases the value of the objective.

Secondly, in the system of inequalities $\hat{\boldsymbol{\Omega}}_1\hat{\boldsymbol{\lambda}}_1 + \hat{\boldsymbol{\Omega}}_2\boldsymbol{\lambda}_2 \leq \hat{\boldsymbol{\eta}}$, in the matrix

$$\hat{\boldsymbol{\Omega}}_1 = \begin{bmatrix} \left(\phi - \frac{1}{2}\right)\psi & 0 \\ 0 & \left(\phi - \frac{1}{2}\right)(1-\psi) \\ \left(\phi - \frac{1}{2}\right)\psi & 0 \\ 0 & \left(\phi - \frac{1}{2}\right)(1-\psi) \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

the terms in the upper part, which correspond to the obedience constraints, are all non-positive if $\phi \leq 1/2$. Thus, by making $\hat{\boldsymbol{\lambda}}_1$ positive and large, we are actually making the obedience constraints less binding if $\phi < 1/2$ or leaving them the same if $\phi = 1/2$. Furthermore, the inequalities represented by the lower part of this matrix together with the non-negativity constraints imply that $\mathbf{0} \leq \hat{\boldsymbol{\lambda}}_1 \leq \mathbf{1}$.

Thus, the observations from the last two paragraphs imply that by putting $\hat{\boldsymbol{\lambda}}_1^* = (1,1)^{\mathsf{T}}$, we either strictly increase the value of the objective while making the obedience constraints less binding or leaving them unchanged while still satisfying the upper bound for this parameter.

Finally, to complete the proof notice that by lemma A.5 the system $\hat{\boldsymbol{\Omega}}_2\boldsymbol{\lambda}_2 \leq \hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\Omega}}_1\hat{\boldsymbol{\lambda}}_1^*$, $\boldsymbol{\lambda}_2 \geq \mathbf{0}$ is non-vacuous.

■

The previous two lemmas yield the following corollary.

Corollary A.7. *Let* $\phi \leq 1/2$. *Then we can find the optimum of problem* (A.15) *by solving a simpler problem:*

$$\boldsymbol{\lambda}_2^* \in \operatorname*{arg\,max}_{\boldsymbol{\lambda}_2 \in \hat{\Upsilon}(\psi,\phi)} \left\{ 2\hat{\boldsymbol{\rho}}_1^{\mathsf{T}} \hat{\boldsymbol{\lambda}}_1^* + 2\psi(1-\psi)\boldsymbol{\rho}_2^{\mathsf{T}}\boldsymbol{\lambda}_2 + 2\big(\alpha\psi + \beta(1-\psi)\big) \right\} \tag{A.19a}$$

$$\Longleftrightarrow$$

$$\boldsymbol{\lambda}_2^* \in \operatorname*{arg\,max}_{\boldsymbol{\lambda}_2 \in \hat{\Upsilon}(\psi,\phi)} \left\{ \boldsymbol{\rho}_2^{\mathsf{T}}\boldsymbol{\lambda}_2 \right\} \tag{A.19b}$$

Proof. Define $T : \mathbb{R} \to \mathbb{R}$ as $T(x) = 2\hat{\boldsymbol{\rho}}_1^{\mathsf{T}} \hat{\boldsymbol{\lambda}}_1^* + 2\big(\alpha\psi + \beta(1-\psi)\big) + 2\psi(1-\psi)x$. It is easily seen that T is a positive and increasing affine transformation, Since $\boldsymbol{\rho}_2^{\mathsf{T}}\boldsymbol{\lambda}_2 \in \mathbb{R}$, we have that $T(\boldsymbol{\rho}_2^{\mathsf{T}}\boldsymbol{\lambda}_2) =$

$2\hat{\boldsymbol{\rho}}_1^\intercal \hat{\boldsymbol{\lambda}}_1^* + 2\psi(1-\psi)\boldsymbol{\rho}_2^\intercal \boldsymbol{\lambda}_2 + 2\big(\alpha\psi + \beta(1-\psi)\big)$. Thus, the objective in (A.19a) is an affine transformation of the objective in (A.19b). Since the feasible sets from both problems are the same they necessarily have the same set of points at which the maximum is attained.

■

Corollary A.7 greatly simplifies the dimensionality and complexity of the problem, since for the parametric cases in which $\phi \leq 1/2$, we can reduce the number of choice variables to three and we can find their optimal values by maximizing a much simpler objective function.

The question remains if we can extend the result that $\hat{\boldsymbol{\lambda}}_1^* = (1, 1)$ is part of an optimum to the case in which $\phi > 1/2$. The answer is negative as is shown in lemma A.9. The proof of lemma A.9 uses the following version of Farkas's Lemma, which is a simple corollary of the version that appears in Gale (1989, Theorem 2.7, p.46).

Theorem A.8 (Farkas's Alternative)**.** *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix and let* $\mathbf{b} \in \mathbb{R}^m$*. Exactly one of the following alternatives hold. Either there exists an* $x \in \mathbb{R}^n$ *satisfying*

$$\begin{aligned} \mathbf{Ax} &\leq \mathbf{b}, \\ \mathbf{x} &\geq \mathbf{0}; \end{aligned} \tag{A.20}$$

or else there exists $\mathbf{y} \in \mathbb{R}^m$ *satisfying*

$$\begin{aligned} \mathbf{y}^\intercal \mathbf{A} &\geq \mathbf{0}, \\ \mathbf{y}^\intercal \mathbf{b} &< 0, \\ \mathbf{y} &> \mathbf{0}. \end{aligned} \tag{A.21}$$

Lemma A.9. *Let* $\hat{\boldsymbol{\lambda}}_1^* = (\xi_\alpha^*, \xi_\beta^*) = (1, 1)$*. Then the correspondence* $\hat{\Upsilon}(\psi, \phi)$ *is empty-valued for all* $\psi, \phi \in \{(\psi, \phi) \in C \,|\, \phi > 1/2\}$.

Proof. The proof strategy will be to find an appropriate vector $\mathbf{y}$ as given by (A.21) in theorem A.8 which will provide a *certificate of infeasibility* for each $\psi, \phi = (\psi, \phi)$ such that $\phi > 1/2$.

The proof will proceed in three steps.

Step 1. Define the correspondence $Y : \{(\psi, \phi) \in C \,|\, \phi > 1/2\} \rightrightarrows \mathbb{R}_+^7$ as the set of vectors $\mathbf{y} = (y_i)_{i=1}^7$ that satisfy the following conditions: $y_1 = y_3 = \hat{y} > 0$, $y_2 = y_4 = \tilde{y} > 0$, $y_5 = y_6 = y_7 = 0$ and

$$\frac{1-\psi}{2\phi\psi} \leq \frac{\tilde{y}}{\hat{y}} < \frac{\psi(\phi - 1/2) + (1-\psi)(1-\phi)}{\psi\phi - (1-\psi)(\phi - 1/2)}, \qquad \text{if } \psi\phi - (1-\psi)(\phi - 1/2) > 0, \tag{A.22}$$

$$\frac{1-\psi}{2\phi\psi} \leq \frac{\tilde{y}}{\hat{y}} < \infty \qquad \text{if } \psi\phi - (1-\psi)(\phi - 1/2) \leq 0. \tag{A.23}$$

For illustration purposes, consider the case in which $\psi = 1/4$. Then we have that

$$Y(\phi) = \begin{cases} \frac{3}{2\phi} \leq \frac{\tilde{y}}{\hat{y}} < \frac{5-4\phi}{3-4\phi} & 1/2 < \phi < 3/4, \\ \frac{3}{2\phi} \leq \frac{\tilde{y}}{\hat{y}} < \infty & 3/4 \leq \phi < 1. \end{cases}$$

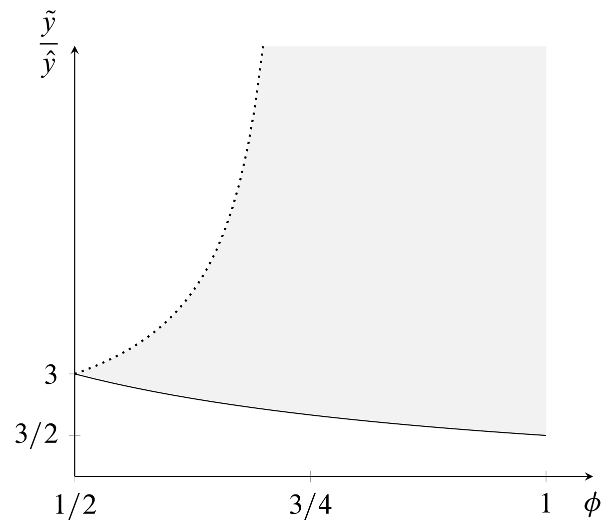


Figure 6: The correspondence $Y(\phi)$, when $\psi = 1/2$ in $(\phi, \tilde{y}/\hat{y})$-space.

Figure 6 plots the correspondence in $(\phi, \tilde{y}/\hat{y})$-space.

We need to show that $Y(\psi, \phi)$ is well-defined and non-empty valued for all $\psi, \phi \in \{(\psi, \phi) \in C \,|\, \phi > 1/2\}$. Notice that since $\phi > 1/2$

$$\frac{1-\psi}{2\phi\psi} > \frac{1-\psi}{\psi} > 0 \Rightarrow \frac{\tilde{y}}{\hat{y}} > 0.$$

The expression

$$\frac{\psi(\phi - 1/2) + (1-\psi)(1-\phi)}{\psi\phi - (1-\psi)(\phi - 1/2)},$$

has a positive numerator since $\phi > 1/2$. Furthermore, if the denominator is positive, i.e. $\psi\phi - (1-\psi)(\phi - 1/2) > 0$, then the whole expression is strictly positive. If on the other hand the denominator is strictly negative, i.e. $\psi\phi - (1-\psi)(\phi - 1/2) < 0$ then the whole expression is strictly negative.

We now show that if $\psi\phi - (1-\psi)(\phi - 1/2) > 0$ and $\phi > 1/2$, then

$$\frac{1-\psi}{2\phi\psi} < \frac{\psi(\phi - 1/2) + (1-\psi)(1-\phi)}{\psi\phi - (1-\psi)(\phi - 1/2)}, \tag{A.24}$$

which makes the condition regarding $\tilde{y}/\hat{y}$ in (A.22) well-defined. The inequality in (A.24) is equivalent to

$$0 < \frac{\big(\psi(\phi - 1/2) + (1-\psi)(1-\phi)\big)2\phi\psi - \big(\psi\phi - (1-\psi)(\phi - 1/2)\big)(1-\psi)}{\big(2\psi\phi\big)\big(\psi\phi - (1-\psi)(\phi - 1/2)\big)}.$$

After some algebraic manipulations, the previous inequality es equivalent to

$$\begin{aligned}0 &< \frac{(\phi - 1/2)\big(2\phi\psi^2 + (1-\psi)^2\big) - 2(1-\psi)\psi\phi(\phi - 1/2)}{(2\psi\phi)\big(\psi\phi - (1-\psi)(\phi - 1/2)\big)} \\ &= \frac{(\phi - 1/2)\big(2\phi\psi(2\psi - 1) + (1-\psi)^2\big)}{(2\psi\phi)\big(\psi\phi - (1-\psi)(\phi - 1/2)\big)}. \end{aligned} \tag{A.25}$$

In light of (A.25), we can see that inequality (A.24) is true if and only if the terms in parenthesis are either *all strictly positive or strictly negative at the same time.* Since by assumption $1/2 < \phi < 1$, $0 < \psi < 1$ so that $\phi\psi > 0$, and $\psi\phi - (1-\psi)(\phi - 1/2) > 0$ we only need to check the sign of $2\phi\psi(2\psi - 1) + (1-\psi)^2$. We have that

$$2\phi\psi(2\psi - 1) + (1-\psi)^2 > \psi(2\psi - 1) + (1-\psi)^2 = 3\psi^2 - 3\psi + 1 \geq 1/4 > 0, \tag{A.26}$$

where the first inequality in (A.26) is due to the fact that $\phi > 1/2$ and the final inequality is due to the fact that $3\psi^2 - 3\psi + 1$ is a convex function of ψ which achieves its minimum at $1/4$ when $\psi = 1/2$.

Therefore, the previous arguments show that $Y(\psi, \phi)$ is a well-defined non-empty correspondence.

Fix any $\mathbf{y} \in Y(\psi, \phi)$. We will show that $\mathbf{y}$ provides a *certificate of infeasibility.* From theorem A.8 we need the inequalities in (A.21) to hold:

$$\mathbf{y}^\mathsf{T}\hat{\boldsymbol{\Omega}}_2 \geq \mathbf{0}, \tag{A.27a}$$

$$\mathbf{y}^\mathsf{T}(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\Omega}}_1\hat{\boldsymbol{\lambda}}_1^*) < 0, \tag{A.27b}$$

$$\mathbf{y} > \mathbf{0}. \tag{A.27c}$$

Evidently, by step 2, each component of $\mathbf{y}$ is either zero or strictly positive so (A.27c) holds. From (A.27a) we get after rearranging and simplifying terms

$$\mathbf{y}^\mathsf{T}\hat{\boldsymbol{\Omega}}_2 = \begin{pmatrix} (2\phi - 1)\big((1-\psi)\hat{y} + \psi\tilde{y}\big) \\ -2\big((1-\phi)(1-\psi)\big)\hat{y} + \psi\tilde{y} \\ -(1-\psi)\hat{y} + 2\phi\psi\tilde{y} \end{pmatrix}^\mathsf{T} \geq \mathbf{0}. \tag{A.28}$$

Similarly, we get from (A.27b) after collecting terms and rearranging that

$$-\hat{y}\big(\psi(\phi - 1/2) + (1-\psi)(1-\phi)\big) + \tilde{y}\big(\psi\phi - (1-\psi)(\phi - 1/2)\big) < 0. \tag{A.29}$$

In (A.28), the first component is strictly positive because $0 < \psi < 1$ and $1/2 < \phi < 1$, so this inequality will not be binding. Thus, we need the following system to have a solution

$$\psi \tilde{y} \geq 2\big((1-\phi)(1-\psi)\big), \tag{A.30a}$$

$$2\phi\psi\tilde{y} \geq (1-\psi)\hat{y}, \tag{A.30b}$$

$$\tilde{y}\big(\psi\phi - (1-\psi)(\phi - 1/2)\big) < \hat{y}\big(\psi(\phi-1/2) + (1-\psi)(1-\phi)\big). \tag{A.30c}$$

The coefficients attached to $\tilde{y}$ and $\hat{y}$ in inequalities (A.30a) and (A.30b) are strictly positive by assumption, so these two are equivalent to

$$\max\left\{\frac{2(1-\phi)(1-\psi)}{\psi}, \frac{(1-\psi)}{2\phi\psi}\right\} \leq \frac{\tilde{y}}{\hat{y}} \tag{A.31}$$

Notice that in (A.31), the inequality

$$\frac{2(1-\phi)(1-\psi)}{\psi} < \frac{(1-\psi)}{2\phi\psi} \Leftrightarrow \frac{4\phi(1-\phi)-1}{2\phi\psi} < 0 \Leftrightarrow 4\phi(1-\phi) < 1$$

holds since $0 < \phi < 1$ and $4\phi(1-\phi) < 1$ because $\phi > 1/2$.

On the other hand, in inequality (A.30c), the coefficient attached to $\hat{y}$ is strictly positive when $\phi > 1/2$ but the coefficient of $\tilde{y}$ can change sign. If it is strictly positive, i.e. $\phi\psi - (1-\psi)(\phi - 1/2) > 0$, then the inequality does provide an upper bound on $\tilde{y}/\hat{y}$:

$$\frac{\tilde{y}}{\hat{y}} < \frac{\psi(\phi-1/2)+(1-\psi)(1-\phi)}{\psi\phi-(1-\psi)(\phi-1/2)},$$

and provides no restriction in any other case, i.e. $\phi\psi - (1-\psi)(\phi-1/2) \leq 0$.

Combining all these observations, we have that all of them correspond to the definition of $Y(\psi, \phi)$, and thus we have shown that the vector $\mathbf{y}$ that we arbitrarily chose indeed is a certificate of optimality. This concludes the proof.

■

A.3 Full characterization of the optimal SBCE

After all the work done to simplify the problem in the previous subsections, we are now ready to move forward with the characterization of an optimal Symmetric Bayes Correlated Equilibrium.

Before proceeding there are a few remarks that are in order. We are looking to describe the set of optima of problem (A.10) as a function of all contests $(\psi, \phi) \in C$ and all productivities $k \in K$. Problem (A.10) is a parametric linear programming problem and although is structurally simple it still poses a challenge to solve fully. The simplifications from the previous subsections in this appendix will help in the full characterization. The search for an optimal solution involves making sure that we are finding positive solutions. On the other hand, the objective function depends on the marginal productivities $(\mathrm{d}_\alpha, \mathrm{d}_\beta)$. Therefore, the structure and shape of the set of optimizers will be impacted by these two considerations.

We need some preparations before describing the full characterization.

Boundaries: The next set of equations describe boundaries in the regions from C that we will describe next.

$$\ell_1(\psi) = \psi \qquad \ell_4(\psi) = 2 - \psi$$

$$\ell_2(\psi) = 1 - \psi \qquad \ell_5(\psi) = \frac{1 - 2\psi - \psi^2 + \sqrt{1 - 4\psi + 10\psi^2 - 4\psi^3 + \psi^4}}{2(1-\psi)}$$

$$\ell_3(\psi) = 1 + \psi \qquad \ell_6(\psi) = \frac{2 + \psi^2 - \sqrt{4 - 8\psi + 4\psi^2 + \psi^4}}{2\psi}$$

$$\ell_7(\psi) = \frac{3 - 2\psi + \psi^2 - \sqrt{1 - 4\psi + 10\psi^2 - 4\psi^3 + \psi^4}}{2(1-\psi)}$$

$$\ell_8(\psi) = \frac{5 - \psi - \sqrt{9 - 10\psi + \psi^2}}{2}$$

$$\ell_9(\psi) = \frac{-2 + 6\psi - \psi^2 + \sqrt{4 - 16\psi + 24\psi^2 - 12\psi^3 + \psi^4}}{2\psi}$$

Regions: After having described the boundaries in the previous paragraph, we will be able to describe concisely the regions in C that will define the characterization of the optimal SBCE. We do this next.

$$C_1 = \{(\psi,\phi) \in C : 2\phi \le \ell_1(\psi)\} \qquad C_{11} = \{(\psi,\phi) \in C : \max(\ell_7(\psi), \ell_8(\psi)) \le 2\phi \le \ell_3(\psi)\}$$

$$C_2 = \{(\psi,\phi) \in C : \ell_1(\psi) \le 2\phi \le \ell_5(\psi)\} \qquad C_{12} = \{(\psi,\phi) \in C : 1/2 \le \psi, 1 < 2\phi \le \ell_4(\psi)\}$$

$$C_3 = \{(\psi,\phi) \in C : \ell_5(\psi) \le 2\phi \le 1\} \qquad C_{13} = \{(\psi,\phi) \in C : \psi \le 1/2, 1 < 2\phi \le \ell_9(\psi)\}$$

$$C_4 = \{(\psi,\phi) \in C : 1/2 \le \psi, 1 < 2\phi \le \ell_6(\psi)\} \qquad C_{14} = \{(\psi,\phi) \in C : \psi \le 1/2, \ell_9(\psi) \le 2\phi \le \ell_3(\psi)\}$$

$$C_5 = \{(\psi,\phi) \in C : \psi \le 1/2, 1 < 2\phi \le \ell_7(\psi)\} \qquad C_{15} = \{(\psi,\phi) \in C : 2\phi = 1\}$$

$$C_6 = \{(\psi,\phi) \in C : \max(\ell_6(\psi), \ell_7(\psi)) \le 2\phi \le \min(\ell_2(\psi), \ell_3(\psi))\}$$

$$C_7 = \{(\psi,\phi) \in C : \ell_3(\psi) \le 2\phi \le \ell_2(\psi)\}$$

$$C_8 = \{(\psi,\phi) \in C : \ell_3(\psi) \le 2\phi \le \ell_4(\psi)\}$$

$$C_9 = \{(\psi,\phi) \in C : \max(1+\psi, 2-\psi) \le 2\phi\}$$

$$C_{10} = \{(\psi,\phi) \in C : 1/2 \le \psi, \ell_6(\psi) \le 2\phi \le \ell_8(\psi)\}$$

Productivity Regions: As mentioned earlier, the shape of the set of optimizers also depends on the marginal productivities of the high versus low ability type player. Specifically, the optimum depends on the ratio $\mathrm{d}_\beta/\mathrm{d}_\alpha$. Figure 7 shows the relevant regions.

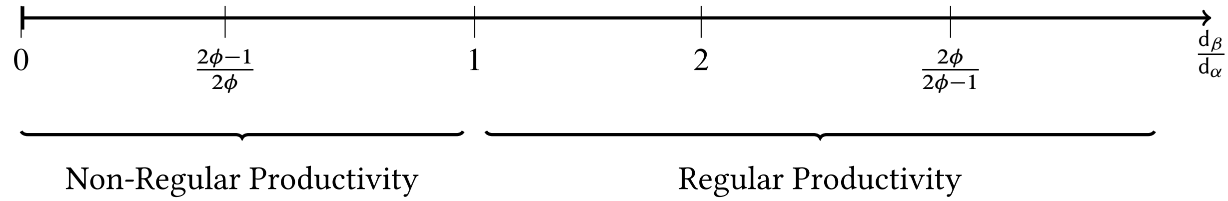


Figure 7: Ratio of marginal productivities—relevant regions

We reproduce here again the diagram about the productivity regions that appeared in the main body of the paper. In figure 7 each mark represents a cutoff of the ratio $\mathrm{d}_\beta/\mathrm{d}_\alpha$ at which the shape of the optimizers changes. All cutoffs except or the first and last one are independent of the particular contest (ψ, ϕ) considered. The first and last cutoffs depend on the normalized value of the contest ϕ when $\phi > 1/2$. The optimizer will be unique except when $\mathrm{d}_\beta/\mathrm{d}_\alpha$ is equal to one of the cutoffs, at which multiple optimizers are possible. As shown in the figure, when the productivity differential ratio is less than one corresponds to the case of non-regular productivity as described in the main body of the paper. Similarly, when the ratio is greater than one corresponds to the case of regular productivity.

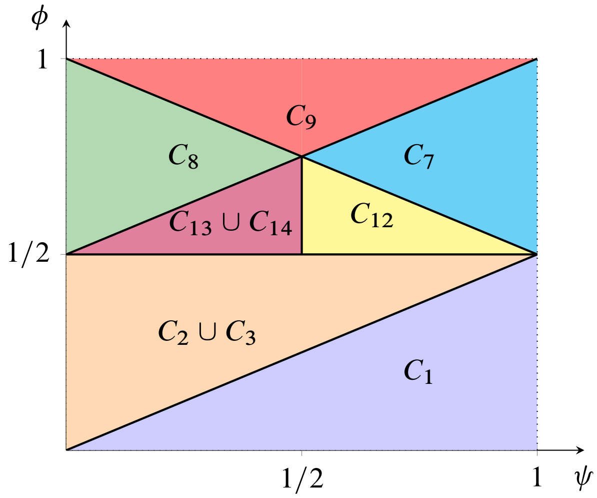
We will present the characterization as a set of lemmas. In what follows, let $\hat{\boldsymbol{\lambda}}^* : C \times K \rightrightarrows \mathbb{R}^5$ denote the set of points at which the maximum of problem (A.15) is achieved, i.e.

$$\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k) = \operatorname*{arg\,max}_{\hat{\boldsymbol{\lambda}} \in \Upsilon(\psi,\phi)} \left\{ 2\hat{\boldsymbol{\rho}}_1^\intercal \hat{\boldsymbol{\lambda}}_1 + 2\psi(1-\psi)\boldsymbol{\rho}_2^\intercal \boldsymbol{\lambda}_2 + 2\big(\alpha\psi + \beta(1-\psi)\big) \right\},$$

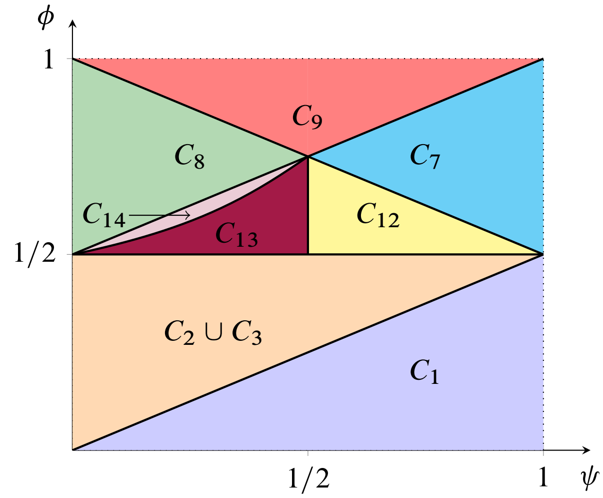
where $\hat{\boldsymbol{\lambda}}^* = (\hat{\boldsymbol{\lambda}}_1^*, \boldsymbol{\lambda}_2^*)$, $\hat{\boldsymbol{\lambda}}_1^* = (\xi_\alpha^*, \xi_\beta^*)$ and $\boldsymbol{\lambda}_2^* = (\mu^*, \delta^*, \zeta^*)$.

As mentioned before, the structure of $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$ depends, via k on the region in which the ratio $\mathrm{d}_\beta/\mathrm{d}_\alpha$ falls. Again, we reproduce one of the diagrams that has already appeared on the main body. Figure 8 represents the large scale structure of $\hat{\boldsymbol{\lambda}}^*$. Each color represents a region in $C \times K$ at which two things happen: $\hat{\boldsymbol{\lambda}}^*$ is constant with respect to k as long as the respective condition on the ratio of the productivity differential is met, and the algebraic structure of $\hat{\boldsymbol{\lambda}}^*$ with respect to (ψ, ϕ) is similar. Intuitively, in each of the regions in figure 8, different obedience and simplex constraints are binding at the optimal SBCE distribution $\hat{\boldsymbol{\lambda}}^*$. Therefore, at the optimal BNE induced by such an optimal decision rule, a player who is recommended to follow a particular action for which the obedience constraint is binding will be indifferent between following it or not. However, following the recommendation will be part of the optimal BNE.

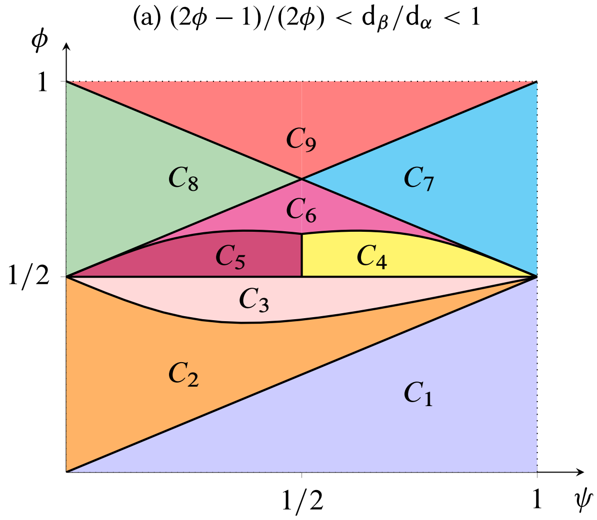
There is a pair of regions in which $\hat{\boldsymbol{\lambda}}^*$ is a constant function of (ψ, ϕ, k). The next two lemmas (A.10 and A.11) deal with these two cases, which occur when (ψ, ϕ) belongs to the regions C_1 and C_{15}, i.e. when $2\phi \le \psi$ and $\phi = 1/2$ respectively.



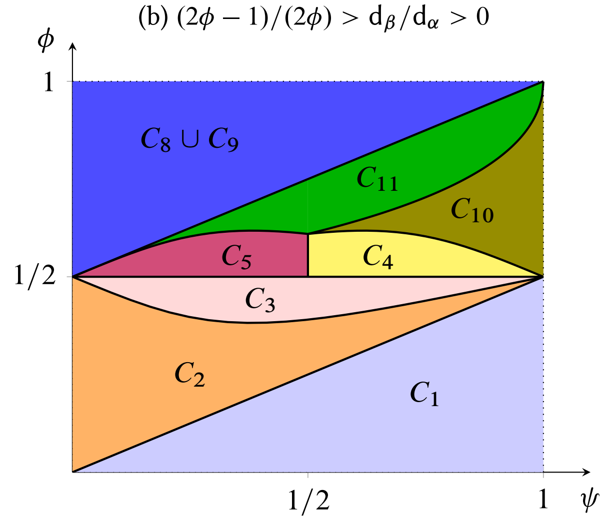
(a) $(2\phi - 1)/(2\phi) < \mathrm{d}_\beta/\mathrm{d}_\alpha < 1$



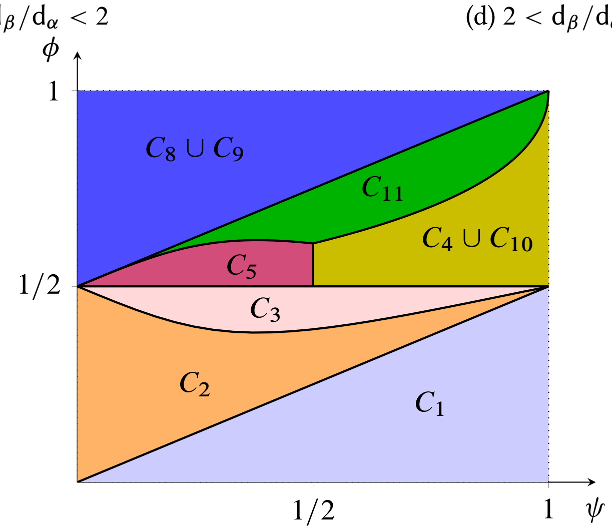
(b) $(2\phi - 1)/(2\phi) > \mathrm{d}_\beta/\mathrm{d}_\alpha > 0$



(c) $1 < \mathrm{d}_\beta/\mathrm{d}_\alpha < 2$



(d) $2 < \mathrm{d}_\beta/\mathrm{d}_\alpha < (2\phi)/(2\phi - 1)$



(e) $(2\phi)/(2\phi - 1) < \mathrm{d}_\beta/\mathrm{d}_\alpha$

Figure 8: Structure of $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$

Lemma A.10. *For all* $(\psi,\phi,k)\in C_1\times K$*,* $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ *is constant and is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = (1,1), \quad \boldsymbol{\lambda}_2^* = (1,0,0).$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *gives an optimal value of*

$$\bar{V}(\psi,\phi,k) = 2\big(\psi(\mathrm{d}_\alpha+\alpha)+(1-\psi)(\mathrm{d}_\beta+\beta)\big) = 2\big(\psi \mathscr{k}(\alpha)\alpha + (1-\psi)\mathscr{k}(\beta)\beta\big)$$

Proof. The optimality of $\hat{\boldsymbol{\lambda}}_1^* = (1,1)$ follows from lemma A.6 since $\phi \leq 1/2$ in the region C_1. For the same reason, we can apply corollary A.7 to find the optimal value of $\boldsymbol{\lambda}_2^*$ by solving the problem $\max\{\boldsymbol{\rho}_2^\intercal\boldsymbol{\lambda}_2\}$ subject to $\boldsymbol{\lambda}_2 \in \hat{\Upsilon}(\psi,\phi)$. Since $\hat{\Upsilon}(\psi,\phi)$ includes the non-negativity constraints and simplex constraint for $\boldsymbol{\lambda}_2 = (\mu,\delta,\zeta)$ and $\boldsymbol{\rho}_2^\intercal\boldsymbol{\lambda}_2 = \mu(\mathrm{d}_\alpha+\mathrm{d}_\beta)+\delta\mathrm{d}_\alpha+\zeta\mathrm{d}_\beta$, it follows that the objective is a weighted sum with weights given by $\boldsymbol{\lambda}_2$ and in which the largest term is $\mathrm{d}_\alpha+\mathrm{d}_\beta$. Thus, it is sufficient to show that $\boldsymbol{\lambda}_2^* \in \hat{\Upsilon}(\psi,\phi)$. The point $\boldsymbol{\lambda}_2^*$ satisfies the non-negativity conditions so we need to show that $\hat{\boldsymbol{\Omega}}_2\boldsymbol{\lambda}_2^* \leq \hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\Omega}}_1\hat{\boldsymbol{\lambda}}_1^*$. The last expression is equivalent to

$$\begin{pmatrix} (1-\psi)\phi \\ -\psi(1-\phi) \\ -(1-\psi)(1-\phi) \\ \psi\phi \\ 0 \\ 0 \\ 1 \end{pmatrix} \leq \begin{pmatrix} \psi\left(\frac{1}{2}-\phi\right) \\ (1-\psi)\left(\frac{1}{2}-\phi\right) \\ -(1-\psi)(1-\phi) \\ \psi\phi \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

From the last expression we can see that the inequalities from rows three to seven are trivially satisfied. From the first row after some algebraic manipulations we get that $2\phi \leq \psi$. From the second row we get that $2\phi \leq 1+\psi$. Since the first one implies the second one for all $(\psi,\phi)\in C$ we get that the expression in the previous display is equivalent to $2\phi\leq\psi$. Noticing that this conditions is precisely the one that defines the boundary of region C_1, we have that $\boldsymbol{\lambda}_2^*$ is feasible if and only if $(\psi,\phi)\in C_1$.

Finally, substituting the optimal values $\hat{\boldsymbol{\lambda}}_1^*$ and $\boldsymbol{\lambda}_2^*$ into the objective A.15 and simplifying terms yields the following expression for the value of the information design problem:

$$\begin{aligned}\bar{V} &= 2\hat{\boldsymbol{\rho}}_1^\intercal\hat{\boldsymbol{\lambda}}_1^* + 2\psi(1-\psi)\boldsymbol{\rho}_2^\intercal\boldsymbol{\lambda}_2^* + 2\big(\alpha\psi+\beta(1-\psi)\big) \\ &= 2\big(\psi(\mathrm{d}_\alpha+\alpha)+(1-\psi)(\mathrm{d}_\beta+\beta)\big) \\ &= 2\big(\psi\mathscr{k}(\alpha)\alpha+(1-\psi)\mathscr{k}(\beta)\beta\big).\end{aligned}$$

■

Lemma A.11. *If* $(\psi,\phi)\in C_{15}$*, i.e.* $(\psi,\phi) = (0,1)\times\{1/2\}$ *and for all* $k\in K$*, we have that* $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ *is constant and is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = (1,1), \quad \boldsymbol{\lambda}_2^* = (1/4,1/4,1/4).$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *gives, for all* $\psi \in (0,1)$ *and* $k \in K$*, an optimal value of*

$$\bar{V}(\psi, 1/2, k) = 2(\alpha\psi + \beta(1-\psi)) + \mathrm{d}_\alpha \psi(1+\psi) + \mathrm{d}_\beta(1-\psi)(2-\psi)$$

Proof. We have again that the optimality of $\hat{\boldsymbol{\lambda}}_1^* = (1,1)$ follows from lemma A.6 since $\phi = 1/2$ in the region C_{16}. Lemma A.4 implies that $\boldsymbol{\lambda}_2^* = (1/4, 1/4, 1/4)$ is the only feasible point of $\hat{\Upsilon}(\psi,\phi)$ for all $\psi \in (0,1)$. The previous fact together with corollary A.7 implies that $\boldsymbol{\lambda}_2^*$ is optimal.

Finally, substituting the values $\hat{\boldsymbol{\lambda}}_1^*$ and $\boldsymbol{\lambda}_2^*$ into the objective A.15 and simplifying terms yields the value of the information design problem as shown in the statement of the lemma. ■

The following lemmas complete the characterization of $\hat{\boldsymbol{\lambda}}^*$ over $\big(C \setminus (C_1 \cup C_{15})\big) \times K$.

Lemma A.12. *If* $k \in \left\{k' \in K : 1 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha}\right\}$ *and* $(\psi,\phi) \in C_2$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ *is given by*

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_1^* &= (1,1), \\ \boldsymbol{\lambda}_2^* &= \left(\tfrac{\psi(1-2\phi)}{2\phi(1-\psi)}, 0, \tfrac{2\phi-\psi}{2\phi(1-\psi)}\right). \end{aligned}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi,\phi,k) = 2\mathrm{d}_\beta(1-\psi) + \mathrm{d}_\alpha \frac{\psi^2}{\phi} + 2(\alpha\psi + \beta(1-\psi)).$$

Lemma A.13. *If* $k \in \left\{k' \in K : 1 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha}\right\}$ *and* $(\psi,\phi) \in C_3$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ *is given by*

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_1^* &= (1,1), \\ \boldsymbol{\lambda}_2^* &= \begin{pmatrix} \frac{2\psi+\phi(3-4\psi-\psi^2)-2\phi^2(1-\psi)-1}{2\psi(1-\psi)} \\ \frac{2\phi^2(1-\psi)-\psi^2+\phi(1-2\psi-\psi^2)}{2(1-\psi)\psi} \\ \frac{2\phi^3(1-\psi)+\psi^2+\phi(2-2\psi-3\psi^2)+\phi^2(6\psi+\psi^2-5)}{2\phi\psi(1-\psi)} \end{pmatrix}. \end{aligned}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\begin{aligned} \bar{V}(\psi,\phi,k) = {} & \mathrm{d}_\alpha\left(2\phi(1-\psi) + 2\psi + \psi^2 - 1\right) \\ & + \mathrm{d}_\beta\left(3 - 2\phi(1-\psi) - 4\psi - \psi^2 + \frac{\psi^2}{\phi}\right) + 2(\alpha\psi + \beta(1-\psi)). \end{aligned}$$

Lemma A.14. *If* $k \in \left\{k' \in K : 0 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} < 1\right\}$ *and* $(\psi,\phi) \in C_2 \cup C_3$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ *is given by*

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_1^* &= (1,1), \\ \boldsymbol{\lambda}_2^* &= \left(\tfrac{2\phi^2-2\phi^3+\psi-3\phi\psi+\phi^2\psi}{2\phi(1-\psi)}, \tfrac{\phi(2\phi-\psi)}{2(1-\psi)}, \tfrac{(1-\phi)^2(2\phi-\psi)}{2\phi(1-\psi)}.\right) \end{aligned}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ implies that the optimal value for the designer is

$$\bar{V}(\psi,\phi,k) = \mathrm{d}_\beta\left(2-2(1+\phi)\psi+\psi^2\right)+\mathrm{d}_\alpha\left(2\phi\psi-\psi^2+\frac{\psi^2}{\phi}\right)+2(\alpha\psi+\beta(1-\psi)).$$

Lemma A.15. *If $k \in \left\{k' \in K : 1 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} < \frac{2\phi}{2\phi-1}\right\}$ and $(\psi,\phi) \in C_4$ then $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \left(\frac{\phi(2\psi-1)}{\psi^2}, 1\right)$$
$$\boldsymbol{\lambda}_2^* = \begin{pmatrix} \frac{2\phi+2\psi-2\phi\psi-2\phi^2\psi-\psi^2+\phi\psi^2-1}{2(1-\psi)\psi} \\ \frac{\phi(2\phi-\psi)}{2(1-\psi)} \\ \frac{1-2\phi+2\phi^2\psi-\phi\psi^2}{2(1-\psi)\psi} \end{pmatrix}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ implies that the optimal value for the designer is

$$\bar{V}(\psi,\phi,k) = \mathrm{d}_\beta\left(2-2(1+\phi)\psi+\psi^2\right)+\mathrm{d}_\alpha\left(2(1+\phi)\psi-\psi^2-1\right)+2(\alpha\psi+\beta(1-\psi)).$$

Lemma A.16. *If $k \in \left\{k' \in K : 1 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha}\right\}$ and $(\psi,\phi) \in C_5$ then $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \left(0, \frac{1-\phi-2\psi+2\phi\psi+\psi^2}{(1-\psi)^2}\right)$$
$$\boldsymbol{\lambda}_2^* = \begin{pmatrix} \frac{(1-\phi)(2\phi+\psi-1)}{2\psi} \\ \frac{\phi(2\phi+\psi-1)}{2\psi} \\ \frac{3\phi-2\phi^2-2\phi\psi+2\phi^2\psi+\phi\psi^2-1}{2(1-\psi)\psi} \end{pmatrix}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ implies that the optimal value for the designer is

$$\bar{V}(\psi,\phi,k) = \mathrm{d}_\beta\left(2+2\phi(-1+\psi)-2\psi+\psi^2\right)+\mathrm{d}_\alpha(1-\psi)(2\phi+\psi-1)+2(\alpha\psi+\beta(1-\psi)).$$

Lemma A.17. *If $k \in \left\{k' \in K : 1 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} < 2\right\}$ and $(\psi,\phi) \in C_6$ then $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \begin{pmatrix} \frac{3\phi-2\phi^2-2\phi\psi+2\phi^2\psi+\phi\psi^2-1}{(2\phi-1)\psi^2} \\ \frac{(1-\phi)\psi(2-2\phi-\psi)}{(2\phi-1)(1-\psi)^2} \end{pmatrix}$$
$$\boldsymbol{\lambda}_2^* = \left(\frac{2-2\phi-\psi}{2(1-\psi)}, \frac{2\psi-1}{2(1-\psi)\psi}, 0\right)$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ implies that the optimal value for the designer is

$$\bar{V}(\psi,\phi,k) = \frac{\mathrm{d}_\beta\psi(2-2\phi-\psi)+\mathrm{d}_\alpha\left(\psi^2+2\phi(1+\psi)-1-2\psi\right)}{2\phi-1}+2(\alpha\psi+\beta(1-\psi)).$$

Lemma A.18. *If* $k \in \left\{k' \in K : 0 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} < 2\right\}$ *and* $(\psi, \phi) \in C_7$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$ *is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \left(\tfrac{(1-\phi)(1-\psi)(2\phi+\psi-1)}{(2\phi-1)\psi^2}, 0\right),$$
$$\boldsymbol{\lambda}_2^* = \left(0, \phi + \tfrac{\psi}{2} - 1, 0.\right)$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi, \phi, k) = \frac{\mathrm{d}_\alpha(1-\psi)(2\phi+\psi-1)}{2\phi-1} + 2(\alpha\psi + \beta(1-\psi)).$$

Lemma A.19. *If* $k \in \left\{k' \in K : 0 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} < 2\right\}$ *and* $(\psi, \phi) \in C_8$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$ *is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \left(1, \tfrac{(1-\phi)\psi(2-2\phi-\psi)}{(2\phi-1)(1-\psi)^2}\right),$$
$$\boldsymbol{\lambda}_2^* = \left(\tfrac{2-2\phi-\psi}{2(1-\psi)}, \tfrac{2\phi-\psi}{2(1-\psi)}, 0\right)$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi, \phi, k) = \frac{\psi\big(\mathrm{d}_\alpha(4\phi-2) - \mathrm{d}_\beta(2\phi+\psi-2)\big)}{2\phi-1} + 2(\alpha\psi + \beta(1-\psi)).$$

Lemma A.20. *If* $k \in \left\{k' \in K : 0 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} < 2\right\}$ *and* $(\psi, \phi) \in C_9$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$ *is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \left(\tfrac{2(1-\phi)(1-\psi)}{\psi(2\phi-1)}, 0\right),$$
$$\boldsymbol{\lambda}_2^* = \left(0, 1, 0\right)$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi, \phi, k) = \frac{\mathrm{d}_\alpha(1-\psi)\psi}{2\phi-1} + 2(\alpha\psi + \beta(1-\psi)).$$

Lemma A.21. *If* $k \in \left\{k' \in K : 2 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha}\right\}$ *and* $(\psi, \phi) \in C_8 \cup C_9$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$ *is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \left(0, \tfrac{2(1-\phi)^2\psi}{(2\phi-1)(1-\psi)}\right),$$
$$\boldsymbol{\lambda}_2^* = \left(1-\phi, \phi, 0\right)$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi, \phi, k) = \frac{2\mathrm{d}_\beta\psi(1-\phi-\psi+\phi\psi)}{2\phi-1} + 2\mathrm{d}_\alpha(1-\psi)\psi + 2(\alpha\psi + \beta(1-\psi)).$$

Lemma A.22. *If* $k \in \left\{k' \in K : 2 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} < \frac{2\phi}{2\phi-1}\right\}$ *and* $(\psi, \phi) \in C_{10}$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$ *is given by*

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_1^* &= \left(\tfrac{\phi(1-\psi)(2-5\phi+2\phi^2+\phi\psi)}{(1-\phi)(2\phi-1)\psi^2}, 1\right), \\ \boldsymbol{\lambda}_2^* &= \left(\tfrac{(2\phi-1)(1-\psi)}{2(1-\phi)\psi}, \tfrac{\phi-2\phi^2+\phi\psi}{2(1-\phi)\psi}, 0\right)\end{aligned}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi, \phi, k) = \mathrm{d}_\beta\left(\frac{(1-\psi)^2}{1-\phi}\right) + \mathrm{d}_\alpha\left(\frac{(1-\psi)(1-\phi-2\phi^2-\psi+3\phi\psi)}{(2\phi-1)(1-\phi)}\right) + 2(\alpha\psi + \beta(1-\psi)).$$

Lemma A.23. *If* $k \in \left\{k' \in K : 2 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha}\right\}$ *and* $(\psi, \phi) \in C_{11}$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$ *is given by*

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_1^* &= \left(0, \tfrac{(1-\phi)^2(2\phi+\psi-1)}{(2\phi-1)(1-\psi)}\right), \\ \boldsymbol{\lambda}_2^* &= \left(\tfrac{(1-\phi)(2\phi+\psi-1)}{2\psi}, \tfrac{\phi(2\phi+\psi-1)}{2\psi}, 0\right)\end{aligned}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi, \phi, k) = \frac{\mathrm{d}_\beta(1-\phi)(1-\psi)(2\phi+\psi-1)}{2\phi-1} + \mathrm{d}_\alpha(1-\psi)(2\phi+\psi-1) + 2(\alpha\psi + \beta(1-\psi)).$$

Lemma A.24. *If* $k \in \left\{k' \in K : \frac{2\phi}{2\phi-1} < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha}\right\}$ *and* $(\psi, \phi) \in C_4 \cup C_{10}$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$ *is given by*

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_1^* &= \left(0, 1\right), \\ \boldsymbol{\lambda}_2^* &= \left(\tfrac{(1-\phi)(2\phi+\psi-1)}{2\psi}, \tfrac{\phi(2\phi+\psi-1)}{2\psi}, \tfrac{2-5\phi+2\phi^2+\phi\psi}{2\psi}\right)\end{aligned}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi, \phi, k) = \mathrm{d}_\beta(1-\psi)(3-2\phi_\psi) + \mathrm{d}_\alpha(1-\psi)(2\phi+\psi-1) + 2(\alpha\psi + \beta(1-\psi)).$$

Lemma A.25. *If* $k \in \left\{k' \in K : 0 < \frac{\mathrm{d}_\beta}{\mathrm{d}_\alpha} < 1\right\}$ *and* $(\psi, \phi) \in C_{12}$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi, \phi, k)$ *is given by*

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_1^* &= \left(\tfrac{1-\phi-2\psi+2\phi\psi+\psi^2}{\psi^2}, 0\right), \\ \boldsymbol{\lambda}_2^* &= \begin{pmatrix}\frac{\phi(2-2\phi-\psi)}{2(1-\psi)} \\ \frac{-1+2\phi+2\psi-4\phi\psi+2\phi^2\psi-\psi^2+\phi\psi^2}{2(1-\psi)\psi} \\ \frac{(1-\phi)(2-2\phi-\psi)}{2(1-\psi)}\end{pmatrix}\end{aligned}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi, \phi, k) = \mathrm{d}_\beta\psi(2-2\phi-\psi) + \mathrm{d}_\alpha(1-2(1-\phi)\psi+\psi^2) + 2(\alpha\psi + \beta(1-\psi)).$$

Lemma A.26. *If* $k \in \left\{k' \in K : \frac{2\phi-1}{2\phi} < \frac{d_\beta}{d_\alpha} < 1\right\}$ *and* $(\psi,\phi) \in C_{13} \cup C_{14}$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ *is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \left(1, \tfrac{(1-\phi)(1-2\psi)}{(1-\psi)^2}\right),$$
$$\boldsymbol{\lambda}_2^* = \begin{pmatrix} \frac{3\phi-2\phi^2+2\psi-4\phi\psi+2\phi^2\psi-\phi\psi^2-1}{2(1-\psi)\psi} \\ \frac{2\phi^2+2\phi\psi-2\phi^2\psi-\psi^2+\phi\psi^2-\phi}{2(1-\psi)\psi} \\ \frac{(1-\phi)(1-2\phi+\psi)}{2\psi} \end{pmatrix}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi,\phi,k) = d_\beta\big(2-2\phi(1-\psi)-2\psi-\psi^2\big) + d_\alpha\big(2\phi(1-\psi)+2\psi+\psi^2-1\big) + 2(\alpha\psi+\beta(1-\psi)).$$

Lemma A.27. *If* $k \in \left\{k' \in K : \frac{2\phi-1}{2\phi} > \frac{d_\beta}{d_\alpha} > 0\right\}$ *and* $(\psi,\phi) \in C_{13}$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ *is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \left(1, 0\right),$$
$$\boldsymbol{\lambda}_2^* = \begin{pmatrix} \frac{3\phi-2\phi^2+2\psi-6\phi\psi+6\phi^2\psi-2\phi^3\psi-\phi^2\psi^2-1}{2\phi(1-\psi)\psi} \\ \frac{2\phi+2\psi-4\phi\psi+2\phi^2\psi-\psi^2+\phi\psi^2-1}{2(1-\psi)\psi} \\ \frac{1-3\phi+2\phi^2-2\psi+8\phi\psi-8\phi^2\psi+2\phi^3\psi-\phi\psi^2+\phi^2\psi^2}{2\phi(1-\psi)\psi} \end{pmatrix}$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi,\phi,k) = d_\beta\big(2-2\phi-\psi\big) + d_\alpha\left(2-4\psi+2\phi\psi+\psi^2-\frac{1-2\psi}{\phi}\right) + 2(\alpha\psi+\beta(1-\psi)).$$

Lemma A.28. *If* $k \in \left\{k' \in K : \frac{2\phi-1}{2\phi} > \frac{d_\beta}{d_\alpha} > 0\right\}$ *and* $(\psi,\phi) \in C_{14}$ *then* $\hat{\boldsymbol{\lambda}}^*(\psi,\phi,k)$ *is given by*

$$\hat{\boldsymbol{\lambda}}_1^* = \left(1, \tfrac{2\phi+2\psi-6\phi\psi+2\phi^2\psi+\phi\psi^2-1}{(2\phi-1)(1-\psi)^2}\right),$$
$$\boldsymbol{\lambda}_2^* = \left(\tfrac{2-2\phi-\psi}{2(1-\psi)}, \tfrac{2\phi-\psi}{2(1-\psi)}, 0\right)$$

Furthermore, $\hat{\boldsymbol{\lambda}}^*$ *implies that the optimal value for the designer is*

$$\bar{V}(\psi,\phi,k) = d_\beta\left(\frac{\psi^2+2\psi+\phi(4-6\psi)-2}{2\phi-1}\right) + 2((\alpha+d_\alpha)\psi+\beta(1-\psi)).$$

B Example: detailed calculations

In this appendix we solve an example of program A.15 for the case in which the parameters are $(\psi,\phi) = (1/2, 1/3)$, i.e. each player believes with probability one half that the rival has low ability and the normalized cost of putting effort is one third.

As discussed in lemma A.6 and corollary A.7, the program A.15 can be reduced to a simpler one since $\phi = 1/3$.

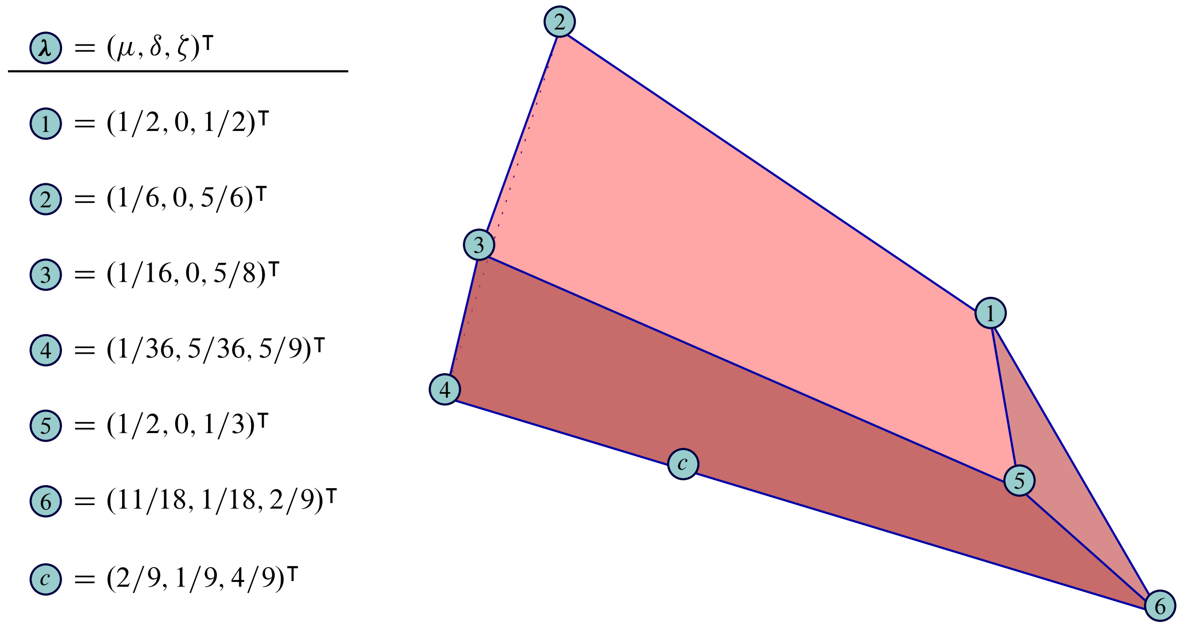


Figure 9: Constraint set for $\phi = 1/3$ and $\psi = 1/2$.

The simpler program is given by the expression (A.19b):

$$\max_{\boldsymbol{\lambda}_2 \in \hat{\Upsilon}(1/2,1/3)} \{\boldsymbol{\rho}_2^{\mathsf{T}} \boldsymbol{\lambda}_2\} \tag{B.1}$$

where $\boldsymbol{\lambda}_2 = (\mu, \delta, \zeta)$ and $\hat{\Upsilon}(1/2, 1/3) = \{\boldsymbol{\lambda}_2 \in \mathbb{R}^3 \mid \hat{\boldsymbol{\Omega}}_2 \boldsymbol{\lambda}_2 \leq \hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\Omega}}_1 \hat{\boldsymbol{\lambda}}_1^*, \boldsymbol{\lambda}_2 \geq \mathbf{0}\}$

The feasible set $\Upsilon(1/\hat{2}, 1/3)$ is a convex polytope in $\mathbb{R}^3$. Thus, it can be represented as the convex hull of its extreme points. The task of finding the set of extreme points of $\hat{\Upsilon}(\psi, \phi)$ is quite demanding and cumbersome as discussed in appendix A. However, for the case in the example $(\psi, \phi) = (1/2, 1/3)$, we can readily find the set of extreme points using standard computational implementations of the simplex algorithm. In the example we have that the feasible set is equivalent to the set of vectors (μ, δ, ζ) that belong to

$$Conv\left\{\begin{pmatrix}1/2\\0\\1/2\end{pmatrix}, \begin{pmatrix}1/6\\0\\5/6\end{pmatrix}, \begin{pmatrix}1/16\\0\\5/8\end{pmatrix}, \begin{pmatrix}1/36\\5/36\\5/9\end{pmatrix}, \begin{pmatrix}1/2\\0\\1/3\end{pmatrix}, \begin{pmatrix}11/18\\1/18\\2/9\end{pmatrix}\right\}. \tag{B.2}$$

The polytope described by (B.2) is illustrated in figure 9.

Although evaluating the objective in (B.1) at each of the extreme points in (B.2) and comparing the results would yield the optimal BCE distribution, we will not pursue this approach since it does not generalize well to the general case. Instead we will show how to solve the optimization program using the simplex algorithm since in this way we will obtain the intuition of how to solve the general case.

We start by substituting the values $(\psi, \phi) = (1/2, 1/3)$ into (B.1) which yields the following:

$$\max_{(\mu,\delta,\zeta)^\top \in \mathbb{R}^3} \mu(\mathrm{d}_\alpha + \mathrm{d}_\beta) + \delta \mathrm{d}_\alpha + \zeta \mathrm{d}_\beta$$
$$\text{subject to}$$
$$\begin{aligned}
OB'_{W,\alpha}: &\quad \frac{\mu}{3} - \frac{2\delta}{3} \leq \frac{1}{6},\\
OB'_{W,\beta}: &\quad -\frac{2\mu}{3} + \frac{\zeta}{3} \leq \frac{1}{6},\\
OB'_{S,\alpha}: &\quad -\frac{2\mu}{3} - \frac{2\delta}{3} - \zeta \leq -\frac{2}{3},\\
OB'_{S,\beta}: &\quad \frac{\mu}{3} + \delta + \frac{\zeta}{3} \leq \frac{1}{3},\\
S_D: &\quad \mu + \delta + \zeta \leq 1,\\
NN_D: &\quad \mu \geq 0,\ \delta \geq 0,\ \zeta \geq 0.
\end{aligned} \tag{B.3}$$

We will introduce the slack variables $x_{w\alpha}$, $x_{w\beta}$, $x_{s\alpha}$, $x_{s\beta}$ and x_d for each of the constraints and the objective variable z as follows:

$$\begin{array}{rcrllll}
& & \text{Basic} & : & \multicolumn{3}{l}{x_{w\alpha}, x_{w\beta}, x_{s\alpha}, x_{s\beta}, x_d} \\
& & \text{Non-Basic} & : & \multicolumn{3}{l}{\mu, \delta, \zeta} \\
x_{w\alpha} & = & \frac{1}{6} & -\frac{1}{3}\mu & +\frac{2}{3}\delta & \\
x_{w\beta} & = & \frac{1}{6} & +\frac{2}{3}\mu & & -\frac{1}{3}\zeta \\
x_{s\alpha} & = & -\frac{2}{3} & +\frac{2}{3}\mu & +\frac{2}{3}\delta & +\zeta \\
x_{s\beta} & = & \frac{1}{3} & -\frac{1}{3}\mu & -\delta & -\frac{1}{3}\zeta \\
x_d & = & 1 & -\mu & -\delta & -\zeta \\
\hline
z & = & & (\mathrm{d}_\alpha + \mathrm{d}_\beta)\mu & +\mathrm{d}_\alpha\delta & +\mathrm{d}_\beta\zeta
\end{array} \tag{B.4}$$

We call the system (B.4) a *dictionary*[20], since they translate any choice of right-hand side values of the *non-basic variables* into corresponding values of the left-hand side variables or *basic variables.* Dictionary (B.4) allows us to write the maximization problem as

$$\max z \quad \text{subject to} \quad \mu \geq 0,\ \delta \geq 0,\ \zeta \geq 0,\ x_{w\alpha} \geq 0,\ x_{w\beta} \geq 0,\ x_{s\alpha} \geq 0,\ x_{s\beta} \geq 0,\ x_d \geq 0.$$

However, notice that dictionary (B.4) violates the restriction of giving a feasible solution to the problem, since by setting all the non-basic variables equal to zero, we get a negative value for the basic variable $x_{s\alpha}$. In order to get hold of an initial feasible dictionary, we will need to solve first an auxiliary

[20]Chvatal (1983, chap. 2, p.17)

problem:

$$\min_{(\mu,\delta,\zeta,x_0)^\intercal\in\mathbb{R}^4} x_0$$

subject to

$$\frac{\mu}{3}-\frac{2\delta}{3}-x_0\le\frac{1}{6},$$
$$-\frac{2\mu}{3}+\frac{\zeta}{3}-x_0\le\frac{1}{6},$$
$$-\frac{2\mu}{3}-\frac{2\delta}{3}-\zeta-x_0\le-\frac{2}{3},$$
$$\frac{\mu}{3}+\delta+\frac{\zeta}{3}-x_0\le\frac{1}{3},$$
$$\mu+\delta+\zeta-x_0\le 1,$$
$$\mu\ge 0,\ \delta\ge 0,\ \zeta\ge 0,$$

which is actually equivalent to maximizing $-x_0$. Again, writing down the slack variables $x_{w\alpha}$, $x_{w\beta}$, $x_{s\alpha}$, $x_{s\beta}$ and x_d and the objective function $w=-x_0$ we obtain the dictionary:

$$\begin{array}{ll} \text{Basic} & : \ x_{w\alpha}, x_{w\beta}, x_{s\alpha}, x_{s\beta}, x_d \\ \text{Non-Basic} & : \ \mu,\delta,\zeta,x_0 \end{array}$$

$$\begin{array}{rcrllll} x_{w\alpha} & = & \frac{1}{6} & -\frac{1}{3}\mu & +\frac{2}{3}\delta & & +x_0 \\ x_{w\beta} & = & \frac{1}{6} & +\frac{2}{3}\mu & & -\frac{1}{3}\zeta & +x_0 \\ x_{s\alpha} & = & -\frac{2}{3} & +\frac{2}{3}\mu & +\frac{2}{3}\delta & +\zeta & +x_0 \\ x_{s\beta} & = & \frac{1}{3} & -\frac{1}{3}\mu & -\delta & -\frac{1}{3}\zeta & +x_0 \\ x_d & = & 1 & -\mu & -\delta & -\zeta & +x_0 \\ \hline w & = & & & & & -x_0 \end{array} \tag{B.5}$$

Let $s \rightleftharpoons t$ denote the replacement-pivoting operation in which variable s enters the basis and variable t leaves it. Dictionary (B.5) is also infeasible but it can be transformed into a feasible one by the operation $x_0 \rightleftharpoons x_{s\alpha}$, which yields the dictionary:

$$\begin{array}{ll} \text{Basic} & : \ x_0, x_{w\alpha}, x_{w\beta}, x_{s\beta}, x_d \\ \text{Non-Basic} & : \ \mu,\delta,\zeta,x_{s\alpha} \end{array}$$

$$\begin{array}{rcrllll} x_0 & = & \frac{2}{3} & -\frac{2}{3}\mu & -\frac{2}{3}\delta & -\zeta & +x_{s\alpha} \\ x_{w\alpha} & = & \frac{5}{6} & -\mu & & -\zeta & +x_{s\alpha} \\ x_{w\beta} & = & \frac{5}{6} & & -\frac{2}{3}\delta & -\frac{4}{3}\zeta & +x_{s\alpha} \\ x_{s\beta} & = & 1 & -\mu & -\frac{5}{3}\delta & -\frac{4}{3}\zeta & +x_{s\alpha} \\ x_d & = & \frac{5}{3} & -\frac{5}{3}\mu & -\frac{5}{3}\delta & -2\zeta & +x_{s\alpha} \\ \hline w & = & -\frac{2}{3} & +\frac{2}{3}\mu & +\frac{2}{3}\delta & +\zeta & -x_{s\alpha} \end{array} \tag{B.6}$$

The next dictionary is obtained by the operation $\zeta \rightleftharpoons x_{w\beta}$, which yields:

$$
\begin{array}{lcl}
\text{Basic} & : & x_0, \zeta, x_{w\alpha}, x_{s\beta}, x_d \\
\text{Non-Basic} & : & \mu, \delta, x_{w\beta}, x_{s\alpha}
\end{array}
$$

$$
\begin{array}{rcrrrrr}
\zeta & = & \frac{5}{8} & & -\frac{1}{2}\delta & -\frac{3}{4}x_{w\beta} & +\frac{3}{4}x_{s\alpha} \\
x_0 & = & \frac{1}{24} & -\frac{2}{3}\mu & -\frac{1}{6}\delta & +\frac{3}{4}x_{w\beta} & +\frac{1}{4}x_{s\alpha} \\
x_{w\alpha} & = & \frac{5}{24} & -\mu & +\frac{1}{2}\delta & +\frac{3}{4}x_{w\beta} & +\frac{1}{4}x_{s\alpha} \\
x_{s\beta} & = & \frac{1}{6} & -\mu & -\delta & +x_{w\beta} & \\
x_d & = & \frac{5}{12} & -\frac{5}{3}\mu & -\frac{2}{3}\delta & +\frac{3}{2}x_{w\beta} & -\frac{1}{2}x_{s\alpha} \\
\hline
w & = & -\frac{1}{24} & +\frac{2}{3}\mu & +\frac{1}{6}\delta & -\frac{3}{4}x_{w\beta} & -\frac{1}{4}x_{s\alpha}
\end{array}
\tag{B.7}
$$

The next and final dictionary is obtained by the operation $\mu \rightleftharpoons x_0$, which yields:

$$
\begin{array}{lcl}
\text{Basic} & : & \mu, \zeta, x_{w\alpha}, x_{s\beta}, x_d \\
\text{Non-Basic} & : & \delta, x_0, x_{w\beta}, x_{s\alpha}
\end{array}
$$

$$
\begin{array}{rcrrrrr}
\mu & = & \frac{1}{16} & -\frac{1}{4}\delta & -\frac{3}{2}x_0 & +\frac{9}{8}x_{w\beta} & \\
\zeta & = & \frac{5}{8} & -\frac{1}{2}\delta & & -\frac{3}{4}x_{w\beta} & +\frac{3}{4}x_{s\alpha} \\
x_{w\alpha} & = & \frac{7}{48} & +\frac{3}{4}\delta & +\frac{3}{2}x_0 & -\frac{3}{8}x_{w\beta} & -\frac{1}{8}x_{s\alpha} \\
x_{s\beta} & = & \frac{5}{48} & -\frac{3}{4}\delta & +\frac{3}{2}x_0 & -\frac{1}{8}x_{w\beta} & -\frac{3}{8}x_{s\alpha} \\
x_d & = & \frac{5}{16} & -\frac{1}{4}\delta & +\frac{5}{2}x_0 & +\frac{3}{8}x_{w\beta} & -\frac{9}{8}x_{s\alpha} \\
\hline
w & = & & & -x_0 & &
\end{array}
\tag{B.8}
$$

Dictionary (B.8) is the final optimal dictionary since $x_0 = 0$ implies that the optimal value of $w = -x_0$ is equal to zero. Thus, this dictionary points to a feasible solution of the original problem (B.3).

We can obtain an initial feasible dictionary for the original problem (B.3) by omitting all terms involving x_0 and writing z in (B.4) by substituting the values of the basic variables μ and δ. Doing this yields the initial feasible dictionary:

$$
\begin{array}{lcl}
\text{Basic} & : & \mu, \zeta, x_{w\alpha}, x_{s\beta}, x_d \\
\text{Non-Basic} & : & \delta, x_{w\beta}, x_{s\alpha}
\end{array}
$$

$$
\begin{array}{rcrrrr}
\mu & = & \frac{1}{16} & -\frac{1}{4}\delta & +\frac{9}{8}x_{w\beta} & \\
\zeta & = & \frac{5}{8} & -\frac{1}{2}\delta & -\frac{3}{4}x_{w\beta} & +\frac{3}{4}x_{s\alpha} \\
x_{w\alpha} & = & \frac{7}{48} & +\frac{3}{4}\delta & -\frac{3}{8}x_{w\beta} & -\frac{1}{8}x_{s\alpha} \\
x_{s\beta} & = & \frac{5}{48} & -\frac{3}{4}\delta & -\frac{1}{8}x_{w\beta} & -\frac{3}{8}x_{s\alpha} \\
x_d & = & \frac{5}{16} & -\frac{1}{4}\delta & +\frac{3}{8}x_{w\beta} & -\frac{9}{8}x_{s\alpha} \\
\hline
z & = & \frac{1}{16}(\mathrm{d}_\alpha + 11\mathrm{d}_\beta) & -\frac{3}{4}(\mathrm{d}_\beta - \mathrm{d}_\alpha)\delta & +\frac{3}{8}(3\mathrm{d}_\alpha + \mathrm{d}_\beta)x_{w\beta} & +\frac{3}{8}(\mathrm{d}_\alpha + 3\mathrm{d}_\beta)x_{s\alpha}
\end{array}
\tag{B.9}
$$

Notice that the initial feasible solution found in (B.9) is precisely vertex number 3 in figure (9).

From the last row of (B.9), we see that in the expression for z, in order to determine which variable will enter the basis we need to consider what is the relationship between d_α and d_β. We will thus break the analysis into three cases.

Case 1: $\mathrm{d}_\beta > \mathrm{d}_\alpha$**.** In this case the coefficient on δ in the last row of (B.9) is negative and the coefficient on $x_{w\beta}$ is positive and strictly smaller than the coefficient on $x_{s\alpha}$. The next dictionary is obtained by the operation $x_{s\alpha} \rightleftharpoons x_{s\beta}$ which yields the dictionary:

$$
\begin{array}{ll}
\text{Basic} & : \mu, \zeta, x_{w\alpha}, x_{s\alpha}, x_d \\
\text{Non-Basic} & : \delta, x_{w\beta}, x_{s\beta}
\end{array}
$$

$$
\begin{array}{rcrllll}
\mu & = & \frac{1}{6} & -\delta & +x_{w\beta} & -x_{s\beta} \\
\zeta & = & \frac{5}{6} & -2\delta & -x_{w\beta} & -2x_{s\beta} \\
x_{w\alpha} & = & \frac{1}{9} & +\delta & -\frac{1}{3}x_{w\beta} & +\frac{1}{3}x_{s\beta} \\
x_{s\alpha} & = & \frac{5}{18} & -2\delta & -\frac{1}{3}x_{w\beta} & -\frac{8}{3}x_{s\beta} \\
x_d & = & & +2\delta & & +3x_{s\beta} \\
\hline
z & = & \frac{1}{6}\mathrm{d}_\alpha + \mathrm{d}_\beta & -3\mathrm{d}_\beta\delta & +\mathrm{d}_\alpha x_{w\beta} & -(3\mathrm{d}_\beta + \mathrm{d}_\alpha)x_{s\beta}
\end{array} \tag{B.10}
$$

The next and final dictionary is obtained by the operation $x_{w\beta} \rightleftharpoons x_{w\alpha}$ which yields:

$$
\begin{array}{ll}
\text{Basic} & : \mu, \zeta, x_{w\beta}, x_{s\alpha}, x_d \\
\text{Non-Basic} & : \delta, x_{w\alpha}, x_{s\beta}
\end{array}
$$

$$
\begin{array}{rcrllll}
\mu & = & \frac{1}{2} & +2\delta & -3x_{w\alpha} & \\
\zeta & = & \frac{1}{2} & -5\delta & +3x_{w\alpha} & -3x_{s\beta} \\
x_{w\beta} & = & \frac{1}{3} & +3\delta & -3x_{w\alpha} & +x_{s\beta} \\
x_{s\alpha} & = & \frac{1}{6} & -3\delta & +x_{w\alpha} & -3x_{s\beta} \\
x_d & = & & +2\delta & & +3x_{s\beta} \\
\hline
z & = & \frac{1}{2}\mathrm{d}_\alpha + \mathrm{d}_\beta & -3(\mathrm{d}_\beta - \mathrm{d}_\alpha)\delta & -3\mathrm{d}_\alpha x_{w\alpha} & -3\mathrm{d}_\beta x_{s\beta}
\end{array} \tag{B.11}
$$

The optimal solution implies that $\delta = x_{w\alpha} = x_{w\beta} = 0$. Thus we have that

$$
(\mu, \delta, \zeta, x_{w\alpha}, x_{w\beta}, x_{s\alpha}, x_{s\beta}, x_d)^\top = (1/2, 0, 1/2, 0, 1/3, 1/6, 0, 0)^\top. \tag{B.12}
$$

Case 2: $\mathrm{d}_\beta < \mathrm{d}_\alpha$**.** In this case all the coefficients in the last row of (B.9) are positive. However, the coefficient on $x_{w\beta}$ is the largest of three. The next dictionary is obtained by performing the operation $x_{w\beta} \rightleftharpoons x_{w\alpha}$ which yields the dictionary

$$
\begin{array}{ll}
\text{Basic} & : \mu, \zeta, x_{w\beta}, x_{s\beta}, x_d \\
\text{Non-Basic} & : \delta, x_{w\alpha}, x_{s\alpha}
\end{array}
$$

$$
\begin{array}{rcrllll}
\mu & = & \frac{1}{2} & +2\delta & -3x_{w\alpha} & \\
\zeta & = & \frac{1}{3} & -2\delta & +2x_{w\alpha} & +x_{s\alpha} \\
x_{w\beta} & = & \frac{7}{18} & +2\delta & -\frac{8}{3}x_{w\alpha} & -\frac{1}{3}x_{s\alpha} \\
x_{s\beta} & = & \frac{1}{18} & -\delta & +\frac{1}{3}x_{w\alpha} & -\frac{1}{3}x_{s\alpha} \\
x_d & = & \frac{1}{6} & -\delta & +x_{w\alpha} & -x_{s\alpha} \\
\hline
z & = & \frac{1}{2}\mathrm{d}_\alpha + \frac{5}{6}\mathrm{d}_\beta & +3\mathrm{d}_\alpha\delta & -(3\mathrm{d}_\alpha + \mathrm{d}_\beta)x_{w\alpha} & +\mathrm{d}_\beta x_{s\alpha}
\end{array} \tag{B.13}
$$

The next and final dictionary is obtained by the operation $\delta \rightleftharpoons x_{s\beta}$ which yields

$$\begin{array}{ll} \text{Basic} & : \ \mu, \delta, \zeta, x_{w\beta}, x_d \\ \text{Non-Basic} & : \ x_{w\alpha}, x_{s\alpha}, x_{s\beta} \end{array}$$

$$\begin{array}{rclllll} \mu & = & \frac{11}{18} & -\frac{7}{3}x_{w\alpha} & -\frac{2}{3}x_{s\alpha} & -2x_{s\beta} \\ \delta & = & \frac{1}{18} & +\frac{1}{3}x_{w\alpha} & -\frac{1}{3}x_{s\alpha} & -x_{s\beta} \\ \zeta & = & \frac{2}{9} & +\frac{4}{3}x_{w\alpha} & +\frac{5}{3}x_{s\alpha} & +2x_{s\beta} \\ x_{w\beta} & = & \frac{1}{2} & -2x_{w\alpha} & -x_{s\alpha} & -2x_{s\beta} \\ x_d & = & \frac{1}{9} & +\frac{2}{3}x_{w\alpha} & -\frac{2}{3}x_{s\alpha} & +x_{s\beta} \\ \hline z & = & \frac{2}{3}\mathrm{d}_\alpha + \frac{5}{6}\mathrm{d}_\beta & -(2\mathrm{d}_\alpha + \mathrm{d}_\beta)x_{w\alpha} & -(\mathrm{d}_\alpha - \mathrm{d}_\beta)x_{s\alpha} & -3\mathrm{d}_\alpha x_{s\beta} \end{array} \tag{B.14}$$

The optimal solution implies that $x_{w\alpha} = x_{s\alpha} = x_{s\beta} = 0$. Thus we have that

$$(\mu, \delta, \zeta, x_{w\alpha}, x_{w\beta}, x_{s\alpha}, x_{s\beta}, x_d)^\intercal = (11/18, 1/18, 2/9, 0, 1/2, 0, 0, 1/9)^\intercal. \tag{B.15}$$

Case 3: $\mathrm{d}_\beta = \mathrm{d}_\alpha = \mathrm{d}$. In this case, we have that in the last row of (B.9) the coefficient on δ is equal to zero. Furthermore, that row simplifies to $z = \frac{3}{4}\mathrm{d} + \frac{3}{2}\mathrm{d}x_{w\beta} + \frac{3}{2}\mathrm{d}x_{s\alpha}$. Thus, the next dictionary is obtained by performing the operation $x_{w\beta} \rightleftharpoons x_{w\alpha}$ which yields

$$\begin{array}{ll} \text{Basic} & : \ \mu, \zeta, x_{w\beta}, x_{s\beta}, x_d \\ \text{Non-Basic} & : \ \delta, x_{w\alpha}, x_{s\alpha} \end{array}$$

$$\begin{array}{rcllll} \mu & = & \frac{1}{2} & +2\delta & -3x_{w\alpha} & \\ \zeta & = & \frac{1}{3} & -2\delta & +2x_{w\alpha} & +x_{s\alpha} \\ x_{w\beta} & = & \frac{7}{18} & +2\delta & -\frac{8}{3}x_{w\alpha} & -\frac{1}{3}x_{s\alpha} \\ x_{s\beta} & = & \frac{1}{18} & -\delta & +\frac{1}{3}x_{w\alpha} & -\frac{1}{3}x_{s\alpha} \\ x_d & = & \frac{1}{6} & -\delta & +x_{w\alpha} & -x_{s\alpha} \\ \hline z & = & \frac{4}{3}\mathrm{d} & +3\mathrm{d}\delta & -4\mathrm{d}x_{w\alpha} & +\mathrm{d}x_{s\alpha} \end{array} \tag{B.16}$$

The next and final dictionary is obtained by the operation $\delta \rightleftharpoons x_{s\beta}$ which yields

$$\begin{array}{ll} \text{Basic} & : \ \mu, \delta, \zeta, x_{w\beta}, x_d \\ \text{Non-Basic} & : \ x_{w\alpha}, x_{s\alpha}, x_{s\beta} \end{array}$$

$$\begin{array}{rcllll} \mu & = & \frac{11}{18} & -\frac{7}{3}x_{w\alpha} & -\frac{2}{3}x_{s\alpha} & -2x_{s\beta} \\ \delta & = & \frac{1}{18} & +\frac{1}{3}x_{w\alpha} & -\frac{1}{3}x_{s\alpha} & -x_{s\beta} \\ \zeta & = & \frac{2}{9} & +\frac{4}{3}x_{w\alpha} & +\frac{5}{3}x_{s\alpha} & +2x_{s\beta} \\ x_{w\beta} & = & \frac{1}{2} & -2x_{w\alpha} & -x_{s\alpha} & -2x_{s\beta} \\ x_d & = & \frac{1}{9} & +\frac{2}{3}x_{w\alpha} & -\frac{2}{3}x_{s\alpha} & +x_{s\beta} \\ \hline z & = & \frac{3}{2}\mathrm{d} & -3\mathrm{d}x_{w\alpha} & & -3\mathrm{d}x_{s\beta} \end{array} \tag{B.17}$$

Notice that the optimal solution has $x_{w\alpha} = x_{s\beta} = 0$ but it says nothing about $x_{s\alpha}$. Thus, if we let

$x_{s\alpha} = t$ we can parameterize the solution as

$$\begin{pmatrix} \mu \\ \delta \\ \zeta \\ x_{w\beta} \\ x_d \end{pmatrix} = \begin{pmatrix} 11/18 \\ 1/18 \\ 2/9 \\ 1/2 \\ 1/9 \end{pmatrix} + t \begin{pmatrix} -2/3 \\ -1/3 \\ 5/3 \\ -1 \\ -2/3 \end{pmatrix} \quad \text{for } t \in [0, 1/6]. \tag{B.18}$$

Notice that if we put $\theta = 1 - 6t$ we get that

$$\begin{pmatrix} \mu \\ \delta \\ \zeta \\ x_{w\beta} \\ x_d \end{pmatrix} = \theta \begin{pmatrix} 11/18 \\ 1/18 \\ 2/9 \\ 1/2 \\ 1/9 \end{pmatrix} + (1-\theta) \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \\ 1/3 \\ 0 \end{pmatrix} \quad \text{for } \theta \in [0, 1].$$

We summarize the previous results in the following proposition.

Proposition B.1. *Suppose that $\psi = 1/2$ and $\phi = 1/3$. Consider the following family decision rules $\{\lambda(\theta) : \theta \in [0, 1]\}$ given by:*

$\lambda(\cdot\|\alpha\alpha)$	W	S
W	1	0
S	0	0

$\lambda(\cdot\|\alpha\beta)$	W	S
W	$\frac{1}{2} + \frac{\theta}{9}$	$\frac{\theta}{18}$
S	$\frac{1}{2} - \frac{5\theta}{18}$	$\frac{\theta}{9}$

$\lambda(\cdot\|\beta\alpha)$	W	S
W	$\frac{1}{2} + \frac{\theta}{9}$	$\frac{1}{2} - \frac{5\theta}{18}$
S	$\frac{\theta}{18}$	$\frac{\theta}{9}$

$\lambda(\cdot\|\beta\beta)$	W	S
W	1	0
S	0	0

Then the following statements are true.

1. *If $\mathrm{d}_\beta > \mathrm{d}_\alpha$ then $\lambda(0)$ solves problem* (B.1) *and gives an expected payoff of*

$$\bar{V} = \alpha + \beta + \frac{3}{4}\mathrm{d}_\alpha + \mathrm{d}_\beta. \tag{B.19}$$

2. *If $\mathrm{d}_\beta < \mathrm{d}_\alpha$ then $\lambda(1)$ solves problem* (B.1) *and gives an expected payoff of*

$$\bar{V} = \alpha + \beta + \frac{5}{6}\mathrm{d}_\alpha + \frac{11}{12}\mathrm{d}_\beta. \tag{B.20}$$

3. *If $\mathrm{d}_\beta = \mathrm{d}_\alpha = \mathrm{d}$, then for any $\theta \in [0, 1]$, $\lambda(\theta)$ solves problem* (B.1) *and gives an expected payoff of*

$$\bar{V} = \alpha + \beta + \frac{7}{4}\mathrm{d}. \tag{B.21}$$

There are some features about the optimum characterized in B.1 that are worth remarking. In the search for an optimum, it is important to note that we always take that the basic variables at each dictionary are positive. In the general case this will provide new restrictions that the parameters (ψ, ϕ) need to satisfy. Also, the termination condition in the simplex algorithm requires that all coefficients of the non-basic variables that appear in the last row of a dictionary to be negative. Notice from our calculations that in the last row z of the dictionary all the coefficients have to do with functions of d_β and d_α. In particular, the determining condition is whether the ratio of these two, $\mathrm{d}_\beta/\mathrm{d}_\alpha$ is strictly greater, strictly less or equal to one. Thus the shape of the optimum depends on how the marginal productivities behave. Furthermore, notice that multiplicity in this case occurs when $\mathrm{d}_\beta/\mathrm{d}_\alpha = 1$ i.e. when the marginal productivites are the same. These type of conditions arise in a similar way in the general case.

References

Angeletos, G.-M. and A. Pavan (2007): "Efficient use of information and social value of information," *Econometrica*, 75, 1103–1142.

Aumann, R. J. (1974): "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics*, 1, 67–96.

——— (1987): "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica*, 1–18.

Aumann, R. J., M. Maschler, and R. E. Stearns (1995): *Repeated games with incomplete information*, MIT press.

Battigalli, P. (2018): "Game Theory, Analysis of Strategic Thinking," Bocconi University, URL: http://didattica.unibocconi.eu/mypage/doc.php?idDoc=29501&IdUte=48808&idr=703&Tipo=m&lingua=eng.

Benoît, J.-P. and J. Dubra (2011): "Apparent overconfidence," *Econometrica*, 79, 1591–1625.

Bergemann, D. and S. Morris (2013): "Bayes correlated equilibrium and the comparison of information structures in games," *Technical Report 1822R, Cowles Foundation for Research in Economics, Yale University.*

——— (2016): "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 11, 487–522.

——— (2018): "Information Design: A Unified Perspective," Cowles Foundation Discussion Paper No. 2075R3.

Brandenburger, A. and E. Dekel (1993): "Hierarchies of beliefs and common knowledge," *Journal of Economic Theory*, 59, 189–198.

Brocas, I. and J. D. Carrillo (2007): "Influence through ignorance," *The RAND Journal of Economics*, 38, 931–947.

Carroll, G. (2016): "Informationally robust trade and limits to contagion," *Journal of Economic Theory*, 166, 334–361.

Chvatal, V. (1983): *Linear Programming*, New York: W.H. Freeman.

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

Denter, P., J. Morgan, and D. Sisak (2014): "'Where Ignorance is Bliss,'Tis Folly to Be Wise': Transparency in Contests," Working paper.

Dubey, P. (2013): "The role of information in contests," *Economics Letters*, 120, 160–163.

Eckhoff, J., P. Gruber, and J. Wills (1993): *Handbook of Convex Geometry*, North-Holland Amsterdam, Netherlands.

Einy, E., D. Moreno, and B. Shitovitz (2017): "The value of public information in common-value Tullock contests," *Economic Theory*, 63, 925–942.

Ely, J., A. Frankel, and E. Kamenica (2015): "Suspense and surprise," *Journal of Political Economy*, 123, 215–260.

Epstein, G. S. and Y. Mealem (2013): "Who gains from information asymmetry?" *Theory and decision*, 75, 305–337.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Vardi (2004): *Reasoning about knowledge*, MIT press.

Fey, M. (2008): "Rent-seeking contests with incomplete information," *Public Choice*, 135, 225–236.

Fu, Q., O. Gürtler, and J. Münster (2013): "Communication and commitment in contests," *Journal of Economic Behavior & Organization*, 95, 1–19.

Fu, Q., J. Lu, and J. Zhang (2016): "Disclosure policy in Tullock contests with asymmetric stochastic entry," *Canadian Journal of Economics/Revue canadienne d'économique*, 49, 52–75.

Gale, D. (1989): *Theory of linear economic models*, Chicago: University of Chicago Press, reprint of the 1960 edition published by McGraw-Hill.

Gentzkow, M. and E. Kamenica (2014): "Costly persuasion," *American Economic Review*, 104, 457–62.

Gürtler, O., J. Münster, and P. Nieken (2013): "Information policy in tournaments with sabotage," *The Scandinavian Journal of Economics*, 115, 932–966.

Harsanyi, J. C. (1967): "Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model," *Management science*, 14, 159–182.

——— (1968a): "Games with incomplete information played by "Bayesian" players part II. Bayesian equilibrium points," *Management Science*, 14, 320–334.

——— (1968b): "Games with incomplete information played by "Bayesian" players, part III. The basic probability distribution of the game," *Management Science*, 14, 486–502.

Kamenica, E. (2018): "Bayesian persuasion and information design," Working paper, University of Chicago.

Kamenica, E. and M. Gentzkow (2011): "Bayesian persuasion," *The American Economic Review*, 101, 2590–2615.

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017): "Persuasion of a privately informed receiver," *Econometrica*, 85, 1949–1964.

Kramm, M. (2018): "Information Design in Multi-Task Contests—Whom to Inform When the Importance of Tasks is Uncertain," Working paper.

Lim, W. and A. Matros (2009): "Contests with a stochastic number of players," *Games and Economic Behavior*, 67, 584–597.

Lipnowski, E. and L. Mathevet (2018): "Disclosure to a Psychological Audience," *American Economic Journal: Microeconomics*, 10, 67–93.

Maschler, M., E. Solan, and S. Zamir (2013): *Game Theory*, Cambridge University Press.

Mathevet, L., J. Perego, and I. Taneva (2016): "Information Design: The Belief Hierarchy Approach," Working paper.

Mertens, J.-F. and S. Zamir (1985): "Formulation of Bayesian analysis for games with incomplete information," *International Journal of Game Theory*, 14, 1–29.

Moldovanu, B. and A. Sela (2001): "The optimal allocation of prizes in contests," *American Economic Review*, 91, 542–558.

Morath, F. and J. Münster (2013): "Information acquisition in conflicts," *Economic Theory*, 54, 99–129.

Morris, S. and H. S. Shin (2002): "Social value of public information," *American Economic review*, 92, 1521–1534.

Münster, J. (2009): "Repeated contests with asymmetric information," *Journal of Public Economic Theory*, 11, 89–118.

Myerson, R. B. (1979): "Incentive compatibility and the bargaining problem," *Econometrica*, 61–73.

——— (1991): *Game theory: analysis of conflict*, Harvard University Press, 1st Paperback Edition, 1997.

Salomone, R., R. Vaisman, and D. P. Kroese (2016): "Estimating the Number of Vertices in Convex Polytopes," in *Proceedings of the International Conference on Operations Research and Statistics*, Citeseer, 96–105.

Taneva, I. (2015): "Information design," Working Paper, Department of Economics, University of Edinburgh.

Vives, X. (1988): "Aggregation of information in large Cournot markets," *Econometrica: Journal of the Econometric Society*, 851–876.

Zamir, S. (2009): "Bayesian Games: Games with Incomplete Information," in *Encyclopedia of Complexity and Systems Science*.

Zhang, J. and J. Zhou (2016): "Information disclosure in contests: A Bayesian persuasion approach," *The Economic Journal*, 126, 2197–2217.