UNIVERSITY OF Southampton
School of Social Sciences

Economics Division
University of Southampton
Southampton SO17 1BJ, UK

Discussion Papers in
Economics and Econometrics

Communication with an Intermediary

(Maksymilian Kwiek)

No. 1603

ISSN 0966-4246

Communication with an Intermediary

Maksymilian Kwiek*†
University of Southampton

June/2016

Abstract

A model of communication with two features is presented: inattentive final receivers who are heterogeneous in their inattention, and the presence of a monopolistic intermediary, whose preferences are not aligned with the preferences of the initial sender. The sender faces a dilemma how to shape her message to maneuver the intermediary into telling the final receivers what she considers important. The sender can engage in self-censoring – potentially beneficial because it prevents the intermediary from misusing the limited attention of the receivers. Less obvious is a policy of interdependence, whereby issues are correlated with each other; this policy is potentially beneficial because it forces the intermediary to talk about all issues tied together, including the ones important to the sender.

Keywords: inattention; communication, information theory.
JEL Code: D83

1 Introduction

Consider a think-tank expert, an activist, or some institution's press secretary, engaged constantly in a communication to an impatient or inattentive public. They can send their messages only via specialized intermediaries such

*m.kwiek@soton.ac.uk. Please do not hesitate to send me comments. If you want to do that anonymously, please use this form: goo.gl/JS9D5j

†I thank Jakub Steiner and the participants of 2016 CRETA Conference at the University of Warwick for helpful comments. I would like to thank Wioletta Dziuda, Ying Chen, Thomas Gall, and other participants of the Workshop at the University of Southampton and Ce^2 Workshop 2014 in Warsaw, for many insightful comments on the earlier version of this paper.

as news reporters. This paper investigates policies that the initial sender may apply in an effort to manipulate an intermediary into telling a story that is the closest to the sender's most preferred one.

The main premise of this study is that listeners are inattentive. Since not everything can be transmitted, the question is what should, and how it should be achieved, given the fact that the messenger's preferences are not the same as the preferences of the initial sender.

This study assumes that everyone prefers everyone to know more rather than less, so in a world without frictions all agents responsible for informing others would honestly reveal everything. This is meant to reflect the fact that experts and journalists might be intrinsically motivated to explain the details of the issue at hand, or, if not, reputation concerns might be enough to prevent outright lies. The communication problem arises because they simply do not have enough time to explain all the caveats. Consequently, they must behave strategically when deciding about which aspects of reality to emphasize and which to downplay, and whether to use other more subtle techniques.

Issues covered in this paper will be painfully familiar to all those engaged in public exposés and debates. Whenever elections are held, or a controversial reform is proposed, such as on Brexit, introduction of the Euro, right-to-die regulation, climate change etc., experts struggle to get what they consider balanced and unbiased message through to general public. Broadly understood popularization of science is another good example.[1] It is not a coincidence that experts and politicians experienced with the media almost never answer questions directly when they are being grilled by the journalists on air. In real life, there are problems other than the ones investigated here, for example players may have incentives to actually misrepresent or hide the facts, or even deceive (as the case might be for politicians or lobbyists). However, there is something to be learned about communication, *even if these complications do not play any role.* Next section summarizes these lessons.

Summary of the results

The intermediary (he) monopolizes the link between the sender (she) and the final receivers, some of whom are more attentive than others. Given the message that the intermediary was told by the initial sender, he has full control over what the final receivers learn – subject to communication

[1] A landmark report on the public understanding of science (Bodmer Report, 1985) very explicitly discusses the differences between the scientists (the sender) and the mass media (the intermediary), and recognizes that the essential fact of this relationship is that information must be packed into limited space.

constraints, of course. The sender and the intermediary derive utilities from informing the final receivers, but these preferences are not perfectly aligned.

The main tension is the trade-off faced by the sender trying to inform heterogeneous receivers. This dilemma stems from the fact that the sender cannot tailor her message to each segment of the final receivers, due to the lack of control of the transmission channel.

On the one hand, knowing that the intermediary will transmit her message word-for-word to the attentive receivers, the sender may provide the best message she can come up with; however, she must then be reconciled with the fact that the intermediary will adjust the composition of her message for the less attentive receivers – focusing only on those aspects that the intermediary finds interesting. On the other hand, the sender may prevent that latter effect by simply being silent on the issues that the intermediary would like to over-communicate; however, this is not the best way to inform attentive receivers.[2] Which of these two approaches is better depends on the parameters of the environment, and comparative statics is intuitive. We call the former policy myopic (honest, principled) and the latter one self-censorship.

A less obvious observation is that, apart from the above straightforward myopia-or-censorship choice, the sender has more subtle informational policies in her arsenal. The sender may tie the issues together in a way that is difficult to disentangle by the intermediary, and thus force the intermediary into transmitting a certain mix of issues to both attentive and inattentive receivers. I will call this policy interdependence.

As an illustration of these policies of censorship and interdependence, consider an expert in geopolitics, a presidential candidate, or certain government's spokesperson, who would like to convey her diagnosis on contemporary conflicts in various places, such as in Syria, Ukraine, Libya, Yemen, etc. A journalist may be interested mostly in one such place, say Syria, perhaps because it appears to be the most relevant. Self-censorship policy would lead the expert to discuss Syria only to a limited degree, and devote more time to other regions, forcing the journalist to do the same. An example of the interdependence policy could be perhaps discussing oil prices, desertification and migration, thus implicitly linking Syria, Libya and Yemen, or discussing president Putin's military expansion, thus implicitly linking Syria and Ukraine.

Finally, one contribution of this paper is in the method. I apply information-theoretic tools to games of communication. I also develop a graphical presentation that should look familiar and intuitive to economists, facilitating

[2]Good or bad informational policy is always meant to be taken from the perspective of a particular player. This paper does not develop normative analysis.

rich analysis.

More on modeling choices

This study takes advantage of information-theoretic tools describing the limits of communication when the receivers are inattentive. These limits are embodied in constraints involving mutual information between the state of nature and a recommended reaction to this state, communicated from one agent to another.

Why use this toolkit? In principle, one could endeavor to express this communication friction with an *ad hoc* and apparently simpler model. For example, information could be quantified as a number of "facts" or "stories" and the constraint could be framed as the maximum number of stories that the receivers are capable of accommodating. The role of the sender would be to create an optimal mix of stories.[3]

This simplified formulation captures some relevant aspects of the interaction between agents studied in this paper. There is a sense in which information resembles a physical quantity, like apples and oranges.[4] If the sender wants the intermediary to forward more stories on Libya to the final receiver, then one option is to limit the number of stories on Syria sent to the intermediary – akin to censorship policy. However, this simplified model misses other important considerations. For example, the role of correlation can be handled by tools of information theory with ease, while it probably cannot be discussed adequately within the framework of the simplified model. This study could not be written without those tools.[5]

Information theory is sometimes used in macroeconomic models of rational inattention to capture an aspect of bounded rationality (for example Sims (2003), (2006)). I assume fully rational agents who are impatient in some sense. The technique details are similar, but the distinction is quite important in interpretation, as the agents in the model below are assumed to be

[3] See for example Alaoui and Germano (2014).

[4] In an essay reviewing the role of competition in markets for news, Gentzkow and Shapiro (2008) caution against applying economic intuition from product markets to information markets. However, if there is a cost of processing information, then some of these differences are blurred.

[5] These tools are quite general, allowing considerable flexibility when assuming the initial distribution of the state of nature (particularity allowing non-independent issues), allowing many different forms of preferences, and allowing any type of hard communication policy, no matter how indirect or peculiar. General and robust properties can be derived. The fact that all examples in this paper use a popular Gaussian-quadratic framework does not matter, as the point of the presentation is to demonstrate certain possibilities and illustrate intuition.

capable of decoding very complex messages. For example, the expert might talk about desertification and use other euphemisms when talking about the conflicts South and East of Europe, and the readers can correctly decipher the intended message of the expert, given the limited time they devote to deciphering.

Next important assumption is about senders' commitment power. This paper presents a setup in which the initial sender commits to the communication policy first, then the intermediary, who gets to know this choice, commits to her own communication policy. The actual communication, as well as the receivers' interpretation of senders' messages occurs afterwards. This creates a simple perfect information game in informational policies, in which the intermediary is a follower, and the initial sender is a leader. Certainly, one can find many examples in which alternative formulations are realistic – for example, the intermediary might be a long-term player whose informational policy has to be accepted by a small-scale or short-term sender. More on the timing assumption in Section 4.[6]

Thirdly, I assume that the communication friction takes the form of a capacity constraint, rather than a smother, for example linear cost function. This is done for the sake of simplicity. A linear cost model, although perhaps more realistic, would contain all the results presented below, probably without much additional insight.

Finally, it is assumed that there is one sender and one intermediary. In many realistic cases, such one-to-one conversations represent only a fragment of a broader public debate, in which multiple experts of diverse interests compete with each other for the attention of the intermediaries, and so do the intermediaries for the attention of the final receivers and for access to experts. If there is a competition of this type, then the players in their roles as receivers obtain some bargaining power, and thus the power of senders to shape their message is expected to be weaker.

Literature

The literature on communication can be classified into two categories. The soft communication (cheap talk) approach adopts the assumption of the seminal paper by Crawford and Sobel (1982) that facts may be costlessly misreported. The verifiable (hard talk) approach assumes that information cannot

[6] Commitment can also refer to a relationship between two players in their roles as the sender and the receiver. In this paper, the sender does not have any incentives to misrepresent her knowledge, and so this type of commitment is not needed. However, one could assume some conflict of interest between the sender and the receiver. Commitment to informational policy would then play an important role.

be lied about, although it can be withheld (e.g. Admati and Pfleiderer (1986), Hörner and Skrzypacz (2011) or Kamenica and Gentzkow (2011), etc.). This study follows the second approach by assuming that the sender publicly commits to an editorial policy prior to the realization of the state of nature and prior to a purchase decision.

Some effort has been put into studying behavior under, and economic implications of, communication with capacity constraints, e.g. Chan and Suen (2008), Kwiek (2010), Glazer and Rubinstein, (2004), Sah and Stiglitz (1986) and a great deal of studies on recommendations or certification, such as Gill and Sgroi (2012). However, this literature does not use the information theoretic tools adopted by this paper.

Information theory has been used by economists under the headline of rational inattention, mainly in the somewhat different context of cognitive frictions in decision problems; see Sims (2003), (2006) and the macroeconomic literature that emerged since then, as well as Matějka and McKay (2015), or Matějka *et al.* (2015). Yang (2015) has used this type of behavioral constraint in studying games of coordination and global games. To my knowledge little has been published in economics on communication using information theory; an exception is a paper on organizational focus by Dessein at al. (2013).

2 Model basics

Agents. There are three types of agents: the Sender, the Intermediary and heterogeneous final receivers, but the latter are not treated as independent players. The Sender knows the realization of an n-dimensional state of nature, represented by a random variable X. For example, X_1 could be a state of a conflict in Syria, X_2 in Libya, etc. The sender's objective is to inform the final receivers about the realization of this state of nature. In all examples, the distribution of X will be Gaussian, $X \sim N(0, K)$, where K is $n \times n$ co-variance matrix.

Communication. The Sender cannot communicate to the final receivers directly; all communication has to go via the Intermediary. That is, the Intermediary is both the receiver of the Sender's message and an independent sender in his own right.

Although preferences of the Sender and the Intermediary are not perfectly aligned in the precise sense described below, I do assume that everyone prefers the final receivers to have more information. Thus, with unconstrained communication, the Sender could simply inform the Intermediary about the realization of X, who would then inform the final receivers, achieving the first best for all.

Instead, it is assumed that all players are inattentive in their role as information receivers; that is, in all instances of communication, there is an information-theoretic exogenous capacity constraint that the senders have to obey. Formally, these communication constraints can be captured by the concept of mutual information. Let $I(X;Z)$ be the mutual information[7] between two random variables X and Z. Roughly, it measures the minimal length of the message, i.e. how many digits on average a message needs to have, so that action Z can be taken at the receiving end of the communication in response to the realization of the source X.[8] Communication constraint will simply say that this length cannot exceed some exogenous number, called the rate. In contrast to engineering or computer science where the capacity constraints expressed in bits, bytes, etc., are taken literary, here they are a metaphor for what is meant to capture frictions in communication.[9]

When there is an intermediary, there are multiple stages of communication. Firstly, the Sender sends messages to the Intermediary at a rate at most s. Secondly, the Intermediary – in his role of a sender – faces two heterogeneous final receivers, one being more inattentive than the other. He sends messages to them at rates r and r^o, respectively, where $r < r^o$. For example, the receivers characterized by rate r^o could be the citizens interested in geopolitics, and those with r could represent the aloof part of the public. The top part of Figure 1 shows the communication structure schematically.

Based on the messages that those final receivers observe, they will take an action. These actions are interpreted as guesses of the realization of X. Let the estimate of X of the more attentive final receiver (whose capacity is r^o) be denoted Y and the one of the less attentive receiver (whose capacity is r) be denoted Z, both $n-$dimensional.

As the Sender designs her informational policy and the Intermediary his own, the probability distribution of (Y, Z) is endogenous, and the main object of study in this paper.

This paper focuses on the case $r < s \leq r^o$. The first inequality says that the Intermediary is more attentive than the less attentive final receiver. This

[7]I.e. $I(X;Z) = \sum_{xz} p_{XZ}(x,z) \log(p_{XZ}(x,z)/p_X(x)\,p_Z(z))$

[8]This paper will not explain information theory, but rather take it as given. For details, see Cover and Thomas (2006).

[9]Communication constraint should be understood as both receiver's capacity to listen as well as the sender's eloquence. The same sender may provide information at two different rates to two different receivers (as the receivers may be able to read at different rates), but also the same reader may receive information at two different rates from two different senders (as one of the senders may be more eloquent than the other one). This study treats them as exogenous capacity parameters characterizing the sender-receiver pairs.

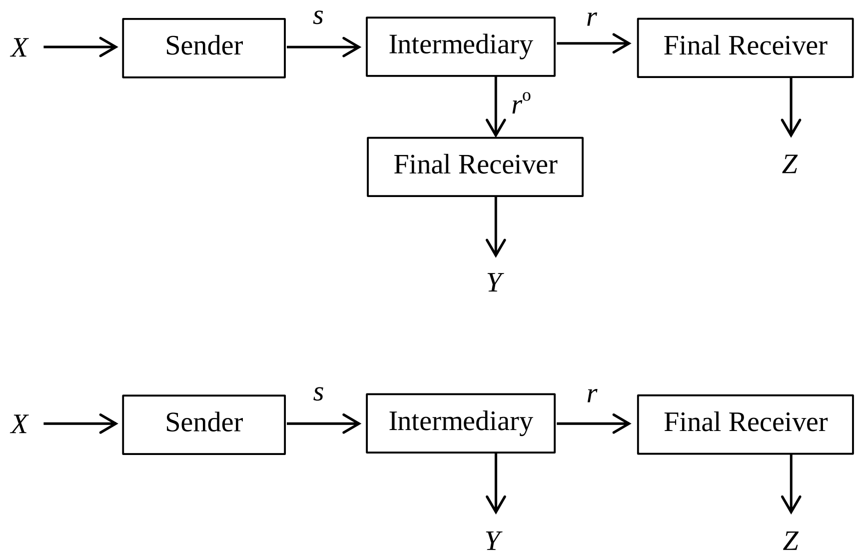

Figure 1: Communication structure

inequality must hold if the Intermediary is to have a non-trivial decision to make; since not everything that he learns from the Sender can be released to the final receiver, he must choose what to tell. Otherwise, the only optimal decision is just to repeat everything that the initial Sender said. The second inequality, $s \leq r^o$, is made mostly for simplicity – the Intermediary will tell the more attentive receiver everything that he learned from the Sender.[10] Under these assumptions, the communication structure can be simplified and illustrated as the bottom part of Figure 1, where Y is interpreted simply as the Intermediary's guess of X, while Z as the less attentive receiver's estimate of X. Communication between the sender and the intermediary will be called stage one, and the communication between the intermediary and the less attentive final receiver will be called stage two.

I postulate that the communication constraints in the two-stage communication problem described above are

$$I\left(X;Y,Z\right) \leq s \text{ and } I\left(X;Z\right) \leq r \tag{1}$$

see Yamamoto (1981). In words, it has to be feasible to transmit both Y and Z within capacity s, and, furthermore, it has to be feasible to transmit

[10]Alternatively, if $s > r^o$, then the Intermediary would have to design a policy for the more attentive receiver. It seems that this complication would not add any new insight to the analysis below.

Z within capacity r.[11]

Preferences. So far, it was loosely stated that all agents want the receivers to know more than less. Now, I can be more precise.

The Sender and the Intermediary want the public to get to know the reality as well as possible. In particular, neither the sender nor the intermediary have any pecuniary motives in selling the newspapers, or attracting advertisers;[12] nor they want to induce a particular action, like buy a product or vote for a candidate. Specifically, it is assumed that they suffer a loss if final receivers' guess of X is not equal to actual realization of X on a given dimension. This loss will be represented by the expected square difference between the source and its estimate, that is, by the vector of variances of $X - Y$ and $X - Z$. In what follows, let Σ denote the $n \times n$ co-variance matrix of a (not necessarily Gaussian) vector random variable $X - Y$, and similarly let Φ denote the co-variance matrix of $X - Z$. Therefore, the sender's loss associated with more attentive receiver's action Y will depend on variances of $X - Y$, that is, on the diagonal elements of Σ, denoted $\mathrm{d}(\Sigma) = (\Sigma_{11}, ..., \Sigma_{nn})$, and the loss associated with Z will depend on diagonal elements of Φ, denoted $\mathrm{d}(\Phi)$.

To be more concrete, assume that the sender has an increasing and quasi-convex loss functions $L_S^1(\mathrm{d}(\Sigma))$ and $L_S^2(\mathrm{d}(\Phi))$, associated with stage one and stage two, respectively. The ultimate loss is a combination of the two stages,

$$\lambda L_S^1(\mathrm{d}(\Sigma)) + (1 - \lambda) L_S^2(\mathrm{d}(\Phi)) \tag{2}$$

where $0 < \lambda < 1$ is a weight that the sender puts on stage one, or on more attentive receivers.[13] Often, but not always, it will be assumed that $L_S^t(...)$ is a linear combination of the relevant variances across dimensions, such as $L_S^t(\mathrm{d}(\Sigma)) = \sum_k \alpha_k^t \Sigma_{kk}$, where $\alpha_k^t \geq 0$ is a weight that the sender puts on dimension k in period t; let $\alpha^t = (\alpha_1^t, ..., \alpha_n^t)$.[14]

[11] It is tempting to say that these three random variables form a Markov chain, that is, that X and Z, conditional on Y, are independent. This will indeed be true under the preference assumption of this paper, but this is not assumed at the outset, and certainly it is not a part of the communication constraint. Random variables Y and Z, both conditional on X, may even be independent. The intermediary simply recovers Y and Z from the messages received from the sender, and chooses to take action Y and forwards coded Z to the less attentive final receiver. If these random variables happen to form a Markov chain, $X \to Y \to Z$, then we can take advantage of the relationship $I(X; Y, Z) = I(X; Y)$ to simplify the first constraint in (1).

[12] Since the size of the audience is fixed by assumption, the sender and intermediary do not have to worry about receivers' outside option.

[13] Alternatively, λ is a fraction or influence of more attentive receivers in the general population.

[14] Only relative weights matter, α_k^t / α_l^t for $k \neq l$. We assume that the vector of weights is normalized, for example by setting $\alpha_2^t = 1$.

The intermediary's preferences, very much like the sender's, are defined on $\mathrm{d}(\Sigma)$ and $\mathrm{d}(\Phi)$. However, the optimal behavior of the intermediary vis-a-vis the more attentive receivers is to repeat Y that has been communicated to him. So, from the point of view of the intermediary, the first stage variance vector $\mathrm{d}(\Sigma)$ is already a given constant. Thus, it is enough to assume that the preferences of the intermediary are defined solely on $\mathrm{d}(\Phi)$, and the particular form selected will be

$$L_I(\mathrm{d}(\Phi)) = \sum_k \beta_k \Phi_{kk}$$

where β_k is the Intermediary's weight on dimension k in stage two.

To restate what has already been described: all agents agree that lower elements of $\mathrm{d}(\Sigma), \mathrm{d}(\Phi)$ are better than higher and that nothing else matters; however, they potentially disagree about which elements of these vectors are relatively more important.

3 Dictator problem

This section considers a hypothetical situation in which the sender has full control over the policy of the intermediary, in addition to her own policy. Since the intermediary is an automaton, $L_I(...)$ is irrelevant. I will refer to this as the *dictator* problem. This scenario is not the main case of interest, but it is presented here as the benchmark against which the game between independent players will be compared; its didactic role is also important.

The sender sends messages to the intermediary who is programmed to forward them to the final receivers. The objective of the sender is to find a distribution of Y, Z conditional on X, denoted $\tilde{g}(y, z|x)$, in order to minimize the loss function (2), subject to the two capacity constraints (1).

Before stating the main result, define the minimal mutual information in the single-stage problem, for a given co-variance matrix Σ. Namely,

$$R(\Sigma) = \min_{g(Y|X):\mathrm{COV}(X-Y)=\Sigma} I(X;Y)$$

Notation $\Phi \succeq \Sigma$, or $\Phi - \Sigma \succeq 0$, means that $\Phi - \Sigma$ is a symmetric positive semidefinite matrix, and thus it is an acceptable co-variance matrix.

It turns out that quite a lot can be said about the optimal behavior of the dictator.

Proposition 1. *If $X \sim N(0, K)$, then a solution $\tilde{g}(y, z|x)$ to the two-stage dictator problem has the following properties.*

1. *The triple X, Y, Z has a joint Gaussian distribution*

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \sim N\left(0, \begin{bmatrix} K & K-\Sigma & K-\Phi \\ K-\Sigma & K-\Sigma & K-\Phi \\ K-\Phi & K-\Phi & K-\Phi \end{bmatrix}\right) \tag{3}$$

2. *The solution is*

$$\begin{aligned} (\Sigma^*, \Phi^*) &= \arg\min_{\Sigma,\Phi} \lambda L_S^1 \left(d\left(\Sigma\right)\right) + \left(1-\lambda\right) L_S^2 \left(d\left(\Phi\right)\right) \\ & K \succeq \Phi \succeq \Sigma \succeq 0 \\ & R\left(\Sigma\right) \leq s, \\ & R\left(\Phi\right) \leq r, \end{aligned} \tag{4}$$

where function $R(...)$ takes a form

$$R\left(\Sigma\right) = (1/2)\log\left(\det K / \det \Sigma\right)$$

3. *Moreover, if K is diagonal, then so are Σ^*, Φ^*.*

The proof is in the Appendix. This proposition is a version of a known characterization of optimal coding with a Gaussian source and quadratic loss function. Instead of finding optimal distribution – a daunting task – Proposition 1 asserts that we have to worry only about an optimal co-variance matrices – a considerably simpler problem. One can solve the problem in point 2 of the Proposition for any number of dimensions n, and any co-variance matrix K, even when dimensions are not independent, as long as the loss function is linear and its weights $\alpha = \alpha^t$ are the same in both stages. The process involves a version of a well-known water-filling procedure on eigenvalues of matrix K, modified to take weights α into account. The application of this procedure to the current version of the model is stated next for completeness and can be skipped.

Let A be a diagonal matrix which has the vector of dictator's weights, $\alpha = \alpha^1 = \alpha^2$, on the diagonal. Define $\hat{K} = A^{1/2} K A^{1/2}$. Furthermore, let $\hat{K} = Q \Lambda Q'$ be the eigendecomposition of $\hat{K}$, where Λ is a diagonal matrix of eigenvalues and Q is a matrix of eigenvectors, and Q' is its inverse (and transpose). The candidate solution will be constructed from diagonal matrices $\tilde{\Sigma}$ and $\tilde{\Phi}$, which we now define. Let diagonal elements of $\tilde{\Sigma}$ be $\tilde{\Sigma}_{kk} = \min\{\Lambda_{kk}, \sigma\}$, where scalar $\sigma \geq 0$ is selected so that the resulting product of these diagonal elements is $\tilde{\Sigma}_{11} \times ... \times \tilde{\Sigma}_{nn} = e^{-2s} \det \Lambda$. Likewise, $\tilde{\Phi}$ has diagonal elements $\tilde{\Phi}_{kk} = \min\{\Lambda_{kk}, \phi\}$, where the scalar $\phi \geq 0$ is selected so that the resulting product is $\tilde{\Phi}_{11} \times ... \times \tilde{\Phi}_{nn} = e^{-2r} \det \Lambda$. These are all well-defined.

Proposition 2. *If $\alpha = \alpha^1 = \alpha^2$ then the solution to point 2 of Proposition 1 is $\Sigma^* = A^{-1/2}Q\tilde{\Sigma}Q'A^{-1/2}$ and $\Phi^* = A^{-1/2}Q\tilde{\Phi}Q'A^{-1/2}$.*

Note, for example, that this solution does not depend on weights λ, which is, of course, specific to this special case.

Much of the discussion later will be conducted in terms of examples with two dimensions. Here are all semidefinitness constraints (4) written explicitly for the case of two dimensions:

1. Matrix $\Phi - \Sigma \succeq 0$ is positive semi-definite, if and only if all leading principal minors are positive, that is, $\Phi_{11} - \Sigma_{11} \geq 0$ and $(\Phi_{11} - \Sigma_{11})(\Phi_{22} - \Sigma_{22}) - (\Phi_{21} - \Sigma_{21})^2 \geq 0$. In other words

$$\Sigma_{kk} \leq \Phi_{kk} \text{ for } k = 1, 2 \tag{5}$$

$$\Sigma_{21} - \Phi_{21} \in \left[\pm\sqrt{\Phi_{11} - \Sigma_{11}}\sqrt{\Phi_{22} - \Sigma_{22}}\right] \tag{6}$$

2. Likewise, matrix $K - \Phi \succeq 0$, if and only if

$$\Phi_{kk} \leq K_{kk} \text{ for } k = 1, 2 \tag{7}$$

$$K_{21} - \Phi_{21} \in \left[\pm\sqrt{K_{11} - \Phi_{11}}\sqrt{K_{22} - \Phi_{22}}\right] \tag{8}$$

3. Matrix $\Sigma \succeq 0$, if and only

$$0 \leq \Sigma_{kk} \text{ for } k = 1, 2 \tag{9}$$

$$\Sigma_{21} \in \left[\pm\sqrt{\Sigma_{11}\Sigma_{22}}\right] \tag{10}$$

Single-stage capacity frontier

One can consider a degenerate version of the two-stage dictator policy, in which stage two does not count, $\lambda = 1$, and random variable Z is simply ignored. Solving this *single-stage problem* will help developing tools of analysis to be employed later.

If X is $N(0, K)$, then by standard arguments used in Proposition 1, the single-stage problem becomes

$$\min_{\Sigma: K \succeq \Sigma \succeq 0 \text{ and } R(\Sigma) \leq s} L_S^1(\mathrm{d}(\Sigma)) \tag{11}$$

One can solve this problem in two steps, the first step dealing with the covariances only, for given variances, and the second one dealing with variances.

Namely, if for a vector of desired variances, $D \in R^n_+$, I define first $\bar{R}(D) = \min_{\Sigma:\mathrm{d}(\Sigma)\leq D, K\succeq\Sigma\succeq 0} R(\Sigma)$, then (11) becomes simply $\min_{D:\bar{R}(D)\leq s} L^1_S(D)$.

Function $\bar{R}(...)$ gives rise to a useful concept of a *single-stage capacity frontier*. It is the set of lowest distortion profiles that are achievable in stage one when available capacity is s. Formally, the single-stage capacity frontier is defined as

$$CF(s) = \left\{D : \bar{R}(D) \leq s \text{ and } D' \leq D \Rightarrow \bar{R}(D') > s\right\}$$

If only capacity s is available, then distortions lying above $CF(s)$ are achievable in a single stage, and no distortions below $CF(s)$ are.

In some cases, the dictator does not have to leave the single-stage capacity frontier when trying to take the stage-two loss into account.

Proposition 3. *Suppose either (i) issues are independent, that is, K is a diagonal matrix, or (ii) weights are equal across stages, $\alpha^1 = \alpha^2$. Then the optimal distortions generated in the two-stage dictator problem belong to their respective single-stage capacity frontiers, $\mathrm{d}(\Sigma^*) \in CF(s)$ and $\mathrm{d}(\Phi^*) \in CF(r)$.*

The conclusion of Proposition 3 does not necessarily hold if issues are not independent.

Claim 1. If not all dimensions are independent and $\alpha^1 \neq \alpha^2$. Then it is possible that the optimal distortions generated in the two-stage dictator problem do not belong to their respective single-stage capacity frontiers.

Sketch of proof in the Appendix.

Example of the dictator policy

Example 1. Suppose that the source X has two statistically independent dimensions with unit variance, that is, matrix K is a two-dimensional identity matrix. The sender's loss function is linear, $L^t_S(D) = \alpha D_1 + D_2$ for $t = 1, 2$, where scalar α is the slope of the iso-loss line and is the same in both stages. To avoid some corner solutions, assume that $e^{-2r} \leq \alpha \leq e^{2r}$.

Claim 2. A dictator's optimal policy in Example 1 is to select a Gaussian distribution described in expression (3), with Σ^* and Φ^* having zero co-variance, $\Sigma^*_{12} = \Phi^*_{12} = 0$, and the diagonal elements being $\mathrm{d}(\Sigma^*) = \left(\sqrt{1/\alpha}e^{-s}, \sqrt{\alpha}e^{-s}\right)$ and $\mathrm{d}(\Phi^*) = \left(\sqrt{1/\alpha}e^{-r}, \sqrt{\alpha}e^{-r}\right)$.

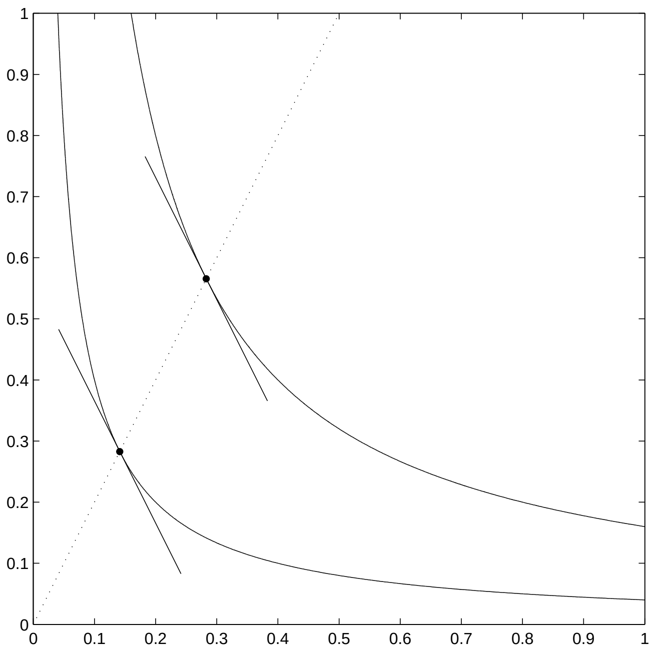

Figure 2: Dictator policy.

To see that this claim is true, notice that Proposition 1 directly implies that the distribution is Gaussian, and the co-variances are zero. The variances can be obtained from point 2 of Proposition 1. Specifically, the semidefinitness constraints (5)-(10) translate into $1 \geq \Phi^*_{kk} \geq \Sigma^*_{kk} \geq 0$, and the capacity constraints into $\Sigma^*_{11}\Sigma^*_{22} \geq e^{-2s}$ and $\Phi^*_{11}\Phi^*_{22} \geq e^{-2r}$. The proposed point $\mathrm{d}\left(\Sigma^*\right)$ minimizes the loss in stage one, subject to the capacity constraint in that stage, and, similarly, $\mathrm{d}\left(\Phi^*\right)$ does this in stage two. The semidefinitness constraints are not binding. Therefore, this is an overall solution, regardless of λ.

Figure 2 depicts the distortions resulting from the optimal choice of the dictator. Two axes measure distortions on two dimensions, so point $(0, 0)$ represents perfect communication, while $(1, 1)$ represents no gains from communication whatsoever. Two curves represent the single-stage capacity frontiers $CF\left(s\right)$ and $CF\left(r\right)$. The dotted line indicates a locus where the capacity frontiers for different capacities are tangent to the iso-loss lines, whose slope is negative α. Black dots represent the solution obtained in Claim 2.

As noted above, the sender does not have to have the same slope of the iso-loss line across stages. If in stage two the sender cared more about the second dimension than in stage one, then iso-loss line would be flatter than α, and thus the stage two distortion point would move along the capacity frontier $CF(r)$ to the right. In an extreme case, the constraint $\Sigma_{22} \leq \Phi_{22}$ may become binding.

4 Sender's policy with an independent Intermediary

Now, I turn to the main section of this study: the case of an intermediary who is an independent decision maker.

The aim is to analyze how the intermediary reacts to the choice of the informational policy of the sender, and how the sender can handle this reaction. It will be assumed that the sender is the first mover who commits to her informational policy before the intermediary selects his informational policy. Thus, in game-theoretic parlance, this section considers a perfect-information game in which the sender is the Stackelberg leader, and the intermediary is the follower, and the solution concept is the subgame perfect Nash equilibrium. The rest of the environment is the same as before.

This timing assumption is meant to capture a situation in which the sender is a long-run and slow-changing player relative to the intermediary. For example, the sender could be a Ministry of Foreign Affairs, an important world leader, or a reputable think tank, while the intermediary could be one of the news outlets, who has to adjust to whatever these senders say.

This timing assumption would not be valid in many other cases. For example, the intermediary could be a long-term player capable of committing to a policy to which the sender has to react (say, a well-known news anchor interviewing various experts as guests). In fact, in some cases it could be plausible that none of the players has a commitment power, and thus assuming a Nash equilibrium of some simultaneous-move game in informational policies would be more appropriate.

To formalize this timing assumption, I postulate that the sender's strategy is a probability distribution of Y conditional on X, denoted $g(y|x)$; the strategy of the intermediary is a probability distribution of Z conditional on Y, denoted $h(z|y)$, as a function of the announced choice of $g(...)$. Recall from the previous section that the dictator chooses both $g(y|x)$ and $h(z|y)$, such that $\tilde{g}(y,z|x) = g(y|x)\,h(z|y)$.[15]

[15] A choice of $g(y|x)$ is restricted only by $I(X;Y) \leq s$, and so g can be selected by the sender independently of what the intermediary is choosing. However, the choice of the intermediary, $h(z|y)$, is constrained by the second capacity constraint $I(X;Z) \leq r$ which

The dilemma of the Sender can be summarized as follows. The Sender can select a policy that is optimal in stage one (the best for the more attentive receiver), achieving the same loss as the dictator. Obviously, the Intermediary has different objectives than the Sender, so his informational policy in the second stage would be different than the one of the dictatorial Sender. I will call this policy of the Sender *myopic*. Alternatively, the Sender can select a different policy. One the one hand, this will generate a higher loss from stage one, but, on the other, this will affect the set of feasible policies available to the Intermediary. If the Sender can induce the Intermediary to behave in the way that is closer to what she thinks is best in stage two, then this alternative policy may be better overall, despite a higher loss in stage one.

The analysis of this section is conducted by presenting a series of examples. In the first class of examples, I suppose that the sender is committed to a policy that puts her stage-one distortions on the single-stage capacity frontier. I will call this class of policies *capacity* policies, for obvious reason. We will see that there are only two capacity policies that may be optimal, one of which is myopic. The other policy, which I call *censorship*, limits information delivered to the intermediary on one dimension, in an attempt to achieve a better communication in some other dimension.

After that, I will consider sender's strategies that lead to stage-one distortions strictly inferior than the single-stage capacity frontier. It will be shown that such a policy of inducing correlation between dimensions, while myopically sub-optimal, may be strictly better overall. I will call this class of informational policies *interdependence*. This result stays in contrast with the dictator's case, who, as we saw in the previous section, would not adjust the co-variances to create interdependence if the source is uncorrelated.

All results will be possibility results. They will be presented as examples with Gaussian source and two independent dimensions.[16]

---

restricts the distribution $\hat{g}(z|x) = \sum_y g(y|x)\, h(z|y)$. Thus, the feasibility of intended h cannot be determined without the knowledge of realized g. The assumption on timing made in this paper solves this problem by letting the intermediary know exactly what the actual g is, and therefore allowing him to select h such that $I(X;Z) \leq r$.

[16] In all examples, the sender will propose a Gaussian $g(...)$. Facing such a sender, the intermediary will find it optimal to react with an $h(...)$ that is also Gaussian. However, I do not know if proposing a Gaussian $g(...)$ is optimal. This approach will show that certain outcomes for the sender are possible (in particular certain rankings can be established), but global optimality is not established.

Myopia and censorship

Example 2. Consider the same environment as in Example 1, but add preferences of the intermediary; assume that he has a linear loss function $L_I(\mathrm{d}(\Phi)) = \beta\Phi_{11} + \Phi_{22}$, where β is interpreted graphically as the negative slope of the iso-loss line. To avoid certain boundary solutions later, assume that $e^{-2r} \leq \beta \leq e^{2r}$. Without loss of generality, let $\alpha > \beta$, so that the the intermediary cares about dimension one to a lesser degree than the sender does.

Capacity policies considered in this subsection lead to distortions $\mathrm{d}(\Sigma) \in CF(s)$, where $CF(s) = \{\mathrm{d}(\Sigma) \leq (1,1) : \Sigma_{11}\Sigma_{22} = e^{-2s}\}$. Thus, the sender must be using a Gaussian distribution of the following type

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(0, \begin{bmatrix} K & K-\Sigma \\ K-\Sigma & K-\Sigma \end{bmatrix}\right) \tag{12}$$

where $\Sigma_{21} = 0$.

The Sender might, but does not have to behave in the same way as the dictator; she may use other capacity policies. There exists a very powerful, if unsophisticated, method of affecting the Intermediary's behavior. The fact is that the Intermediary can never generate a lower stage-two distortion than the one achieved by the initial sender in stage one. Formally, $\mathrm{d}(\Phi)$ is bounded below by $\mathrm{d}(\Sigma)$, as stated by constraint (5).

For example, if the Sender is desperate to make Φ_{11} lower than the one selected by the Intermediary, then she can censor the second dimension, that is, she can increase Σ_{22} in order to force Φ_{22} up as well. With the capacity constraint present, this leaves the intermediary with no other option but to inform about the first dimension and thus reduce Φ_{11}. It is worth emphasizing that the censorship of the second dimension occurs not because the sender intrinsically wants to hide this information from the receivers – she rather thinks that informing about the second dimension is a poor use of the receivers' attention, and one way to inform about the first dimension is to free up some of that resource.

In the context of Example 2, let us define two specific capacity policies:

- The policy which is optimal in a single-stage problem, $(\Sigma_{11}^m, \Sigma_{22}^m) = \left(\sqrt{1/\alpha}e^{-s}, \sqrt{\alpha}e^{-s}\right)$ will be called *myopic*.
- The policy in which the sender censors information that the intermediary would like to over-provide will be called *censorship*, $(\Sigma_{11}^c, \Sigma_{22}^c) = \left(\sqrt{1/\alpha}\left(e^{-2s}/\tau\right), \tau\sqrt{\alpha}\right)$, where $\tau = \sqrt{\lambda e^{-2s} + (1-\lambda)e^{-2r}}$.

Optimal behavior in Example 2 is described in the Claim below, which is proved in the Appendix. However, the intuition can be shown easier in a graphical way, which is presented right after the Claim.

Claim 3. Suppose that the sender must use a capacity policy. Then

1. The best response to any capacity policy $\mathrm{d}(\Sigma)$ is a Gaussian distribution characterized by co-variance matrix Φ, where $\Phi_{21} = 0$ and where the diagonal elements are as follows

   (a) $\mathrm{d}(\Phi) = (\Sigma_{11} e^{2s-2r}, e^{-2s}/\Sigma_{11})$ if $\Sigma_{11} \leq \sqrt{1/\beta} e^{r-2s}$

   (b) $\mathrm{d}(\Phi) = \left(\sqrt{1/\beta} e^{-r}, \sqrt{\beta} e^{-r}\right)$ if $\sqrt{1/\beta} e^{r-2s} \leq \Sigma_{11} \leq \sqrt{1/\beta} e^{-r}$

   (c) $\mathrm{d}(\Phi) = (\Sigma_{11}, e^{-2r}/\Sigma_{11})$ if $\sqrt{1/\beta} e^{-r} \leq \Sigma_{11}$,

2. The sender's optimal policy is:

   (a) If $\Sigma_{11}^m < \sqrt{1/\beta} e^{r-2s}$ then the solution is censorship $(\Sigma_{11}^c, \Sigma_{22}^c)$.

   (b) If $\Sigma_{11}^c > \sqrt{1/\beta} e^{r-2s}$ then the solution is myopic $(\Sigma_{11}^m, \Sigma_{22}^m)$.

   (c) If $\Sigma_{11}^c \leq \sqrt{1/\beta} e^{r-2s} \leq \Sigma_{11}^m$ then the optimal policy is either the myopic policy $(\Sigma_{11}^m, \Sigma_{22}^m)$, or the censorship $(\Sigma_{11}^c, \Sigma_{22}^c)$, depending on the values of the parameters.

Figure 3 represents the distortions achieved by the myopic strategy as two black dots. The distortion profile obtained in stage one is the same, by definition, as in the dictator policy illustrated on Figure 2. However, in stage two the intermediary's choice does not replicate the dictator's favorite outcome. Since $\beta < \alpha$, the intermediary's choice is to the right on the stage-two capacity frontier. The sender's iso-loss line going through that point represents a higher loss associated with not being able to replicate the dictator's distortion profile shown on Figure 2.

The effect of censorship policy – the alternative to the myopic policy – is presented on Figure 3 as white dots. As the sender provides more information in the first dimension, she must provide less in the second, moving to the left along the capacity frontier. Soon enough, constraint (5) bites, the intermediary is forced to provide less information on the second dimension too, moving the stage-two distortions along the stage-two capacity frontier to the left. The effect of this censorship policy from the point of view of the sender is higher loss from stage one, but a lower loss in stage two, which, depending on parameters, may be a lower overall loss than sticking to the myopic policy.

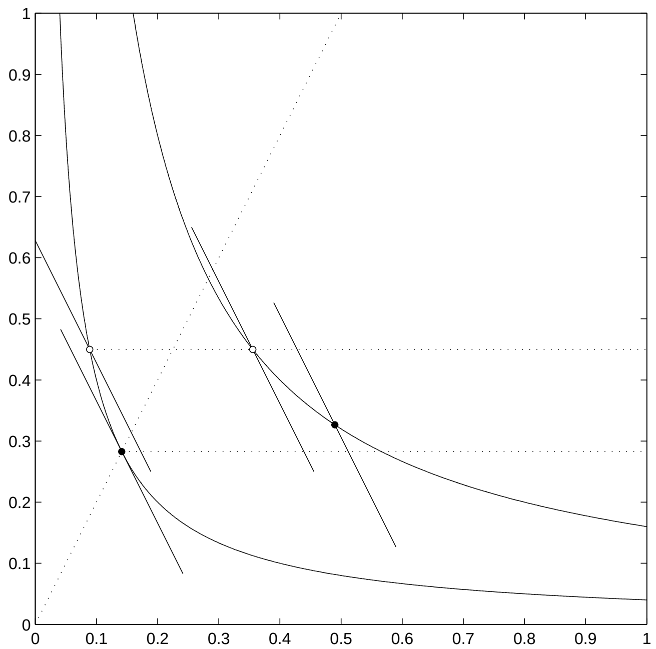

Figure 3: Myopia and censorship

Interdependence

Censorship might be better than myopia because there is a trade-off involving some sacrifice in stage one, in order to improve the payoff in stage-two. The policy of interdependence will play the same card but in a different manner.

Recall that the dictator's optimal co-variance is zero. The point of interdependence policy is to generate some non-zero co-variance between the dimensions. If the sender selects $\Sigma_{21} < 0$ then the estimate y presented by the sender to the intermediary will have some positive co-variance across dimensions, $K_{21} - \Sigma_{21} = -\Sigma_{21} > 0$. This ties intermediary's hands somewhat, as it may give rise to some co-variance in the second stage as well. Since the sender wants to inform about both dimensions (especially in the example below), exposing the intermediary to correlation between both dimensions may be somewhat advantageous to the sender. We will now see that, indeed, this is possible.

Example 3. Assume the environment as in example 2 specified above, but replace the sender's linear loss functions with the following extremely nonlinear one, in which the dimensions are perfect complements; for $t = 1, 2$

$$\lambda \max\{\Sigma_{11}, \Sigma_{22}\} + (1 - \lambda) \max\{\Phi_{11}, \Phi_{22}\} \tag{13}$$

Let also $\lambda = 64/89$, $\beta = 1/4$, $e^{-2s} = 0.04$ and $e^{-2r} = 0.16$.

The rest of this section calculates and compares the losses of the sender from different policies.

For reference, we state the dictator policy first; quite obviously it is the best capacity policy: $\Sigma_{21} = \Phi_{21} = 0$, $\Sigma_{11} = \Sigma_{22} = 0.2$ and $\Phi_{11} = \Phi_{22} = 0.4$. The resulting distortions would lie on the 45^o line on Figure 4.

As far as capacity policies with an independent intermediary are concerned, let us start with the myopic one. The intermediary's behavior is described in the following claim.

Claim 4. The intermediary's best response to the myopic policy $\Sigma_{21} = 0$, and $\Sigma_{11} = \Sigma_{22} = 0.2$ is $\Phi_{21} = 0$, $\Phi_{11} = 0.8$ and $\Phi_{22} = 0.2$. Sender's loss is approximately 0.3685.

Proof. Given this Σ, the best response must involve $\Phi_{21} = 0$. The intermediary chooses $\mathrm{d}(\Phi)$ to solve $\min_{\mathrm{d}(\Phi)} \beta\Phi_{11} + \Phi_{22}$, such that $\Phi_{11}\Phi_{22} = 0.16$ and $0.2 \leq \Phi_{kk} \leq 1$. Solving this without the latter constraints yields the candidate solution $\Phi_{11} = 0.8$ and $\Phi_{22} = 0.2$. Since it satisfies the constraints, it is the overall solution. Sender's overall loss is $(64/89) \times 0.2 + (1 - 64/89) \times 0.8 \approx 0.3685$. □

Moving on to censorship policies we obtain the following result.

Claim 5. The best capacity policy is censorship $\Sigma_{21} = 0$, and $\Sigma_{22} = 0.25$ and $\Sigma_{11} = 0.16$. Sender's loss is approximately 0.3596

Proof. The only chance to benefit from the censorship policy is to decrease Σ_{11} and increase Σ_{22} along the single-stage capacity frontier described by $\Sigma_{11}\Sigma_{22} = 0.04$, relative to the myopic point. This increases the loss in stage one, but forces the intermediary to increase Φ_{22} from 0.2 (recall that $\Sigma_{22} \leq \Phi_{22}$). Since the intermediary will select a point along the single-stage capacity frontier $\Phi_{11}\Phi_{22} = 0.16$, this maneuver will make Φ_{11} less than 0.8, something that makes the sender strictly happier. This trade-off will work until $\Sigma_{22} = 0.4$, past which the loss will unambiguously start rising. The overall sender's loss in (13) can be written as a function only of $0.2 \leq \Sigma_{22} \leq 0.4$.

$$L_S = \lambda\Sigma_{22} + (1 - \lambda)\, 0.16/\Sigma_{22}$$

To minimize, let us use the f.o.c. with respect Σ_{22}, ignoring the bounds:

$$\Sigma_{22} = 0.4\sqrt{(1-\lambda)/\lambda}$$

which is $\Sigma_{22} = 0.25$, since $\lambda = 64/89$, which satisfies the bounds. The associated overall loss is $(64/89)\,0.25 + (1 - 64/89)\,0.64 \approx 0.3596$ □

Now, for the first time, I introduce an example of an *interdependence policy.*

Claim 6. If the sender's policy is $\Sigma_{11} = \Sigma_{22} = 0.25$ and $\Sigma_{21} = -0.15$, then the intermediary responds by selecting $\Phi_{11} \approx 0.6367$, $\Phi_{22} \approx 0.2619$ and $\Phi_{21} \approx -0.0822$. The sender's loss is ≈ 0.3586.

To explain this interdependence policy, one has to start with clarifying the effect of changing the stage-one co-variance to a non-zero value. There are two effects.

Firstly, since this co-variance is sub-optimal in a single-stage problem, the distortion point $\mathrm{d}(\Sigma)$ can no longer be on the single-stage capacity frontier, the fact that can be seen from the capacity constraint $\Sigma_{11}\Sigma_{22} - (\Sigma_{21})^2 = 0.04$. The point $\Sigma_{11} = \Sigma_{22} = 0.25$ is just feasible when $\Sigma_{21} = -0.15$. Obviously, the resulting stage-one loss is greater relative to otherwise achievable $\Sigma_{11} = \Sigma_{22} = 0.2$.

The other effect is the change of the set of intermediary's feasible co-variance matrices. As always, the diagonal elements $\mathrm{d}(\Phi)$ have to be at least as high as now higher $\mathrm{d}(\Sigma)$, according to constraint (5). In addition to this, constraint (6) is also binding, forcing Φ_{21} to be negative, like Σ_{21}. The effect of this change can be depicted graphically in Figure 4. The bold solid curve represents the set of feasible distortions for the intermediary in stage two. Of course, this curve has to lie to the North-East of the familiar single-stage capacity frontier; it also has to lie to the North-East of the dotted lines $\Sigma_{kk} = 0.25$. This is not all, however, as there are still some pieces missing in the corners, where $\Sigma_{kk} \approx \Phi_{kk}$. For example, a policy previously feasible as a response to censorship, $\mathrm{d}(\Phi) = (0.64, 0.25)$, is not feasible now.

With $\beta = 1/4$, the intermediary has a relatively flat iso-loss line and his optimal point turns out to be on this relatively flat part of the stage-two capacity frontier, represented by the empty dot.

From the point of view of the sender, the loss calculation in every scenario is very easy. The best capacity policy is censorship, and has exactly the same loss in stage one as the interdependence policy – we compare point $(0.16, 0.25)$ generated by censorship, with point $(0.25, 0.25)$ generated by interdependence. However, the interdependence policy is slightly better in stage two – we compare point $(0.64, 0.25)$ from censorship, with point $(0.6367, 0.2619)$ from interdependence.

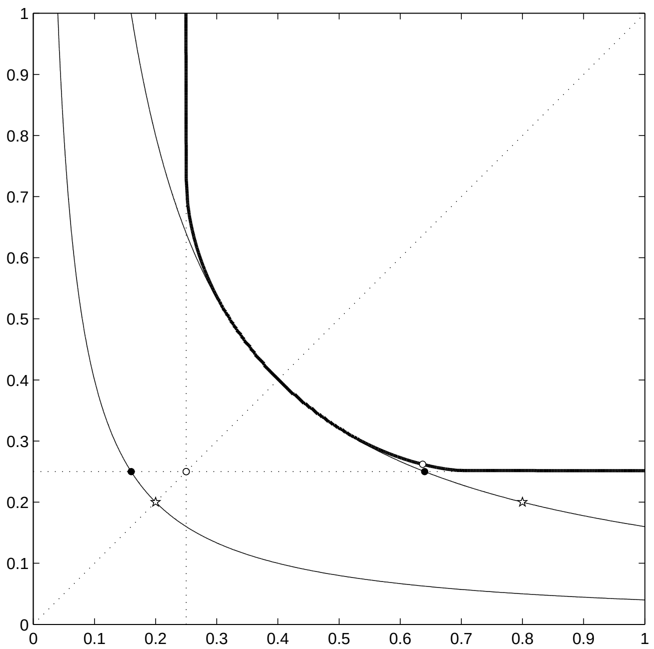

Figure 4: Myopic policy - stars. Censorship - black dots. Interdependence - empty dots

We can see that the shape of the loss function in (13) is quite important. If instead the sender had a linear loss function like in previous sections, it would be difficult to come up with an interdependence policy that is better than the best capacity policy. Inducing censorship locally has only a second-order negative effect on the losses in stage one, while increasing co-variance has an immediate first-order negative effect on those losses – this is one of the reasons why censorship often easily beats interdependence.

To summarize this section, Table 1 compares all three policies discussed in the context of Example 3. The punchline is that the best capacity policy is not as good as a proposed interdependence policy.

5 Conclusions

The main objective of this study is to investigate the act of communication and its observable implications. The main innovation is the introduction of a plausible communication constraint; it stems from the fact that the

Policy	Σ	Φ	L_S
Myopic	$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0 \\ 0 & 0.2 \end{bmatrix}$	0.3685
Censorship (best capacity policy)	$\begin{bmatrix} 0.16 & 0 \\ 0 & 0.25 \end{bmatrix}$	$\begin{bmatrix} 0.64 & 0 \\ 0 & 0.25 \end{bmatrix}$	0.3596
Interdependence	$\begin{bmatrix} 0.25 & -0.15 \\ -0.15 & 0.25 \end{bmatrix}$	$\begin{bmatrix} 0.6367 & -0.0822 \\ -0.0822 & 0.2619 \end{bmatrix}$	0.3586
Dictator (for comparison)	$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix}$	0.2562

Table 1: Comparisons of three policies of this section

senders are not eloquent enough, and the readers are not patient enough for communication to be completely friction-less. Tools of information theory developed for engineering and computer science are successfully employed. The approach is indirect; I am not modeling the details of the communication protocol, but rather focus on general statistical properties of the equilibrium behavior, such as which aspects of reality are communicated, and what is the correlation profile.

The main trade-off faced by the initial sender is whether to focus on patient readers willing to understand the details of the sender's position, or on those who merely want to learn the basics. Since achieving both is impossible, the sender may decide to give up on the former in order to gain on the latter. We could call this phenomenon *rational populism*, to be understood as a policy of promoting superficial instead of deep understanding.

More specifically, both strategies of self-censorship and interdependence are populist in this sense. A policy such as self-censorship is very intuitive; it is also undoubtedly powerful in altering the intermediary's behavior, should the sender decide to use it. A policy of interdependence is more subtle (and probably more difficult to test). It is here where information theory is really useful, as it would be virtually impossible to formalize interdependence with a different, *ad hoc* model.

6 Proofs

6.1 Proof of Proposition 1

Mutual information between source X and its estimates Y and Z can be bounded below. Function η is entropy.

$$\begin{aligned}
I(X;Y,Z) &= \eta(X) - \eta(X|Y,Z) \\
&= \eta(X) - \eta(X-Y|Y,Z) \\
&\geq \eta(X) - \eta(X-Y) \\
&\geq \eta(X) - \eta(N(0,\Sigma)) \\
&= (1/2)\log\left((2\pi e)^n \det K\right) - (1/2)\log\left((2\pi e)^n \det \Sigma\right) \\
&= (1/2)\log\left(\det K / \det \Sigma\right) \qquad (14)
\end{aligned}$$

where the first inequality comes from the fact that conditioning reduces entropy. The second follows from the fact that Gaussian distribution maximizes entropy among all distributions of a given co-variance. Similarly,

$$\begin{aligned}
I(X;Z) &\geq \eta(X) - \eta(N(0,\Phi)) \\
&= (1/2)\log\left(\det K / \det \Phi\right) \qquad (15)
\end{aligned}$$

Note that the joint distribution in expression (3) has the following properties. Firstly, $X \sim N(0,K)$, as required. Random variables $X-Y$ and Z are independent, as well as $X-Y$ and Y are independent, which means that the first inequality in (14) holds with equality. Moreover, $X-Y \sim N(0,\Sigma)$, which means that the second inequality holds with equality as well. Similarly, $X-Z$ and Z are independent and $X-Z \sim N(0,\Phi)$ meaning that (15) holds with equality. Thus, for given co-variance matrices Σ, Φ, replacing any policy $\tilde{g}(y,z|x)$ with (3) will not change the value of the objective function but will relax the capacity constraints to their lowest bounds; these bounds are now $R(\Sigma) \leq s$ and $R(\Phi) \leq r$.

The last remaining issue is to set endogenous matrices Σ, Φ to minimize the loss function, subject to these capacity constrains, and subject to the constraint that the co-variance matrix in (3) has to be positive semi-definite. To simplify this last condition, transform the variables into $\begin{bmatrix} X-Y, Y-Z, Z \end{bmatrix}^T$. This vector of random variables has a co-variance matrix

$$\begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Phi-\Sigma & 0 \\ 0 & 0 & K-\Phi \end{bmatrix} \qquad (16)$$

Clearly, co-variance in (3) is positive semi-definite, if and only if (16) is. The latter is positive semi-definite if and only if the diagonal blocks are, denoted as $\Sigma \succeq 0$, $\Phi - \Sigma \succeq 0$ and $K - \Phi \succeq 0$, or simply as $K \succeq \Phi \succeq \Sigma \succeq 0$. This establishes the program in point 2.

Moving to point 3, if K is diagonal and Φ, Σ are not, then one can replace all their off-diagonal elements with zero and achieve lower $\det(\Sigma)$ and $\det(\Phi)$, by Hadamard's inequality, thus relaxing the constraints $R(\Sigma) \le s$ and $R(\Phi) \le r$ further. Note that all constraints $K - \Phi \succeq 0$, $\Phi - \Sigma \succeq 0$, $\Sigma \succeq 0$ still hold after this modification, and the objective function is unaffected.

6.2 Proof of Proposition 2

This problem can be written as

$$\begin{array}{ll} \arg\min_{\Sigma,\Phi} & \lambda \mathrm{tr}\left(A^{1/2}\Sigma A^{1/2}\right) + (1-\lambda)\,\mathrm{tr}\left(A^{1/2}\Phi A^{1/2}\right) \\ & K \succeq \Phi \succeq \Sigma \succeq 0 \\ & (1/2)\log\left(\det K / \det \Sigma\right) \le s \\ & (1/2)\log\left(\det K / \det \Phi\right) \le r \end{array}$$

Let $\hat{\Sigma} = A^{1/2}\Sigma A^{1/2}$, and $\hat{\Phi} = A^{1/2}\Phi A^{1/2}$. Since $K \succeq \Phi$ if and only if $\hat{K} \succeq \hat{\Phi}$, etc., and $\det K / \det \Sigma = \det \hat{K} / \det \hat{\Sigma}$, etc., the problem can be stated alternatively as

$$\begin{array}{ll} \arg\min_{\hat{\Sigma},\hat{\Phi}} & \lambda \mathrm{tr}\hat{\Sigma} + (1-\lambda)\,\mathrm{tr}\hat{\Phi} \\ & \hat{K} \succeq \hat{\Phi} \succeq \hat{\Sigma} \succeq 0 \\ & (1/2)\log\left(\det \hat{K} / \det \hat{\Sigma}\right) \le s \\ & (1/2)\log\left(\det \hat{K} / \det \hat{\Phi}\right) \le r \end{array}$$

Define $\check{\Sigma} = Q'\hat{\Sigma}Q$ and $\check{\Phi} = Q'\hat{\Phi}Q$. Since $\mathrm{tr}\hat{\Sigma} = \mathrm{tr}\left(Q'\hat{\Sigma}Q\right)$, the problem can be further changed to

$$\begin{array}{ll} \arg\min_{\check{\Sigma},\check{\Phi}} & \lambda \mathrm{tr}\check{\Sigma}, + (1-\lambda)\,\mathrm{tr}\check{\Phi} \\ & \Lambda \succeq \check{\Phi} \succeq \check{\Sigma} \succeq 0 \\ & (1/2)\log\left(\det \Lambda / \det \check{\Sigma}\right) \le s \\ & (1/2)\log\left(\det \Lambda / \det \check{\Phi}\right) \le r \end{array}$$

Now, some properties of the optimal $\check{\Sigma}, \check{\Phi}$ are apparent. In particular, they must be diagonal (for if not, then one could replace it with another matrix, which has the same diagonal elements but zeros elsewhere; the objective function and the semipositiveness constraints would not be affected, but capacity constraints would be relaxed, by Hadamard's inequality). Hence, the problem can be written in a more tractable way as

$$
\begin{aligned}
\arg\min_{\check{\Sigma},\check{\Phi}} \quad & \lambda \sum_k \check{\Sigma}_{kk}, + (1-\lambda) \sum_k \check{\Phi}_{kk} \\
& \Lambda_k \geq \check{\Phi}_{kk} \geq \check{\Sigma}_{kk} \geq 0 \text{ for all } k \\
& \check{\Sigma}_{11} \times ... \times \check{\Sigma}_{nn} \leq e^{-2s} \det \Lambda \\
& \check{\Phi}_{11} \times ... \times \check{\Phi}_{nn} \leq e^{-2r} \det \Lambda
\end{aligned}
$$

It is easy to see that the capacity constraints must hold with equality. To see that proposed $\tilde{\Sigma}$ and $\tilde{\Phi}$ are a solution to this problem, note that they satisfy the the associated Kuhn-Tucker optimality conditions. Changing variables back to the original problem leads to the Proposition.

6.3 Proof of Proposition 3

(i) If K is diagonal, then so are Σ^*, Φ^*, by Proposition 1. Suppose $\mathrm{d}(\Sigma^*)$ does not belong to a single stage capacity frontier. In other words, there is another diagonal matrix Σ^0, which is feasible, $\bar{R}(\mathrm{d}(\Sigma^0)) \leq s$, but better in at least one dimension $\mathrm{d}(\Sigma^*) \geq \mathrm{d}(\Sigma^0)$. Notice that $\mathrm{d}(\Phi^*) \geq \mathrm{d}(\Sigma^0) \geq 0$ and thus all positive semidefinitness constraints are satisfied. Since the loss is lower when a feasible Σ^0 is applied, Σ^* could not have been optimal, a contradiction. A similar argument applies to Φ^*.

(ii) Suppose that weights are equal across stages, $\alpha^1 = \alpha^2$. The solution of the problem in Proposition 2 does not depend on λ, even if $\lambda = 0$ or $\lambda = 1$. These two extreme cases generate distortion profiles on the single stage capacity frontiers.

6.4 Proof of Claim 1 (sketch)

Suppose that there are two dimensions, and let the co-variance between them be $K_{21} > 0$. The off-diagonal elements Σ^*_{21} and Φ^*_{21} must be as close to zero as possible, or, specifically, for given diagonal elements $(\mathrm{d}(\Sigma^*), \mathrm{d}(\Phi^*))$, they must hit the bounds (6) and (8). Suppose now that s and r are very close to zero, indicating a very severe capacity constraint. Then the optimal $(\mathrm{d}(\Sigma^*), \mathrm{d}(\Phi^*))$ must be close to $\mathrm{d}(K)$, which implies that these off-diagonal elements are actually strictly positive.

$$
\begin{aligned}
\Phi^*_{21} &= K_{21} - \sqrt{K_{11} - \Phi^*_{11}}\sqrt{K_{22} - \Phi^*_{22}} \\
\Sigma^*_{21} &= \Phi^*_{21} - \sqrt{\Phi^*_{11} - \Sigma^*_{11}}\sqrt{\Phi^*_{22} - \Sigma^*_{22}}
\end{aligned}
$$

Combining these two equations to eliminate Φ_{21}^* we obtain

$$K_{21} - \Sigma_{21}^* = \sqrt{K_{11} - \Phi_{11}^*}\sqrt{K_{22} - \Phi_{22}^*} + \sqrt{\Phi_{11}^* - \Sigma_{11}^*}\sqrt{\Phi_{22}^* - \Sigma_{22}^*} \tag{17}$$

At the same time, solving a single-stage problem, to find a single-stage capacity frontier, in a similar way leads to

$$K_{21} - \Sigma_{21}^* = \sqrt{K_{11} - \Sigma_{11}^*}\sqrt{K_{22} - \Sigma_{22}^*} \tag{18}$$

Equations (17) and (18), describing optimal Σ_{21}^* in these two different problems are not compatible in general, and hence the two stage dictator problem may lead to a Σ^* that does not achieve the single stage capacity frontier.

6.5 Proof of Claim 3

For any capacity policy of the sender, defined by $\mathrm{d}(\Sigma)$, the best response of the intermediary is a Gaussian distribution characterized by Φ, with zero co-variance $\Phi_{21} = 0$, and where the diagonal elements are a solution to $\min_{\mathrm{d}(\Phi)} \beta\Phi_{11} + \Phi_{22}$, such that $\Phi_{11}\Phi_{22} = e^{-2r}$ and $\Sigma_{kk} \leq \Phi_{kk} \leq 1$ for $k = 1, 2$. Point 1 of the claim provides the solution to this program (note that $\Sigma_{11}\Sigma_{22} = e^{-2s} < e^{-2r}$, so either $\Sigma_{11} \leq \Phi_{11}$ is binding, or $\Sigma_{22} \leq \Phi_{22}$, or none of the two; moreover, by assumption on β in Example 1, the constraint $\mathrm{d}(\Phi) \leq 1$ is never binding).

Now turn to the behavior of the sender. The sender's payoff is

$$\lambda(\alpha\Sigma_{11} + \Sigma_{22}) + (1 - \lambda)(\alpha\Phi_{11} + \Phi_{22}) \tag{19}$$

where $\mathrm{d}(\Phi)$ depends on $\mathrm{d}(\Sigma)$ through the intermediary's best response described above, and $\mathrm{d}(\Sigma)$ is restricted to satisfy $\Sigma_{11}\Sigma_{22} = e^{-2s}$. Hence, there are three regions

1. If $\sqrt{1/\beta}e^{r-2s} \leq \Sigma_{11} \leq \sqrt{1/\beta}e^{-r}$, then the intermediary's choice is in the interior and thus independent on the selection of Σ_{11}. In this interval, the sender should obviously select the best myopic choice in the first stage $\Sigma_{11}^m = \sqrt{1/\alpha}e^{-s}$, as long as it is not lower than the lower bound of this interval. If on the other hand $\Sigma_{11}^m < \sqrt{1/\beta}e^{r-2s}$, the solution is equal to that lower bound. Note that $\Sigma_{11}^m \leq \sqrt{1/\beta}e^{-r}$ always.

2. If $\Sigma_{11} \leq \sqrt{1/\beta}e^{r-2s}$, then the sender may face a trade-off. Lowering Σ_{11} and increasing Σ_{22} may increase the stage-one loss, but this choice forces the intermediary to increase Φ_{22} (equal to Σ_{22}), and thus will

lower the loss in stage two. Overall, the sender's payoff (19), written as a function of Σ_{22} only, is

$$\lambda\left(\alpha\frac{1}{\Sigma_{22}}e^{-2s}+\Sigma_{22}\right)+(1-\lambda)\left(\alpha\frac{1}{\Sigma_{22}}e^{-2r}+\Sigma_{22}\right)$$

Optimality is achieved at $\Sigma_{22}^c=\tau\sqrt{\alpha}$ and $\Sigma_{11}^c=\sqrt{1/\alpha}\,(e^{-2s}/\tau)$, as long as $\Sigma_{11}^c\leq\sqrt{1/\beta}e^{r-2s}$. Otherwise, the upper bound of this interval should be selected.

3. Finally, if $\sqrt{1/\beta}e^{-r}\leq\Sigma_{11}$ then the sender unambiguously obtains a grater loss in stage one and in stage two. The best choice in this interval is its lower bound.

Now we can compare the best choices in each interval to obtain the best Σ_{11} overall. Notice, that the best choice in point 1 is always better than in point 3. Secondly, it can be easily verified that the best choice in point 1 and point 2 cannot be both at the end of the interval. Since the loss function is continuous at the endpoint of these intervals, the overall solution must be either Σ_{11}^m, or Σ_{11}^c. It is the latter if point 1 constraint is binding, and the former if point 2 constraint is binding. If none of them is binding, then either Σ_{11}^m or Σ_{11}^c may be optimal depending on direct payoff comparison.

6.6 Proof of Claim 6

This Σ is feasible, because $0\leq\Sigma_{kk}\leq 1$ and $\Sigma_{11}\Sigma_{22}-(\Sigma_{21})^2=0.04$.

The rest of the proof focuses on the best response of the intermediary, that is, on the following problem

$$\min_{\Phi_{11},\Phi_{22},\Phi_{21}}\beta\Phi_{11}+\Phi_{22}$$

subject to

$$\begin{aligned}0.25 &\leq \Phi_{kk}\leq 1 \text{ for } k=1,2 && (20)\\ \Phi_{21} &\in \left[\Sigma_{21}\pm\sqrt{\Phi_{11}-\Sigma_{11}}\sqrt{\Phi_{22}-\Sigma_{22}}\right] && (21)\\ 0 &\leq \Phi_{11}\Phi_{22}-(\Phi_{21})^2-e^{-2r} && (22)\end{aligned}$$

The plan of the proof is to ignore the constraints (20), and use only (21) and (22). Constraints (20) will be verified ex post.

Step 1. Φ_{21} should be as close to zero as possible such that (21) holds. The reason is that making it close to zero only relaxes the constraint (22) and does not affect the objective function. Since $\Sigma_{21} < 0$, this co-variance must be

$$\Phi_{21} = \min\left\{0, \Sigma_{21} + \sqrt{\Phi_{11} - \Sigma_{11}}\sqrt{\Phi_{22} - \Sigma_{22}}\right\} \tag{23}$$

Consider the case in which Φ_{21} is negative, that is,

$$\Phi_{21} = \Sigma_{21} + \sqrt{\Phi_{11} - \Sigma_{11}}\sqrt{\Phi_{22} - \Sigma_{22}} < 0$$

It will be later verified that it indeed is. Denote the right-hand side of (22) as

$$RHS\left(\mathrm{d}\left(\Phi\right)\right) = \Phi_{11}\Phi_{22} - \left(\Sigma_{21} + \sqrt{\Phi_{11} - \Sigma_{11}}\sqrt{\Phi_{22} - \Sigma_{22}}\right)^2 - e^{-2r} \tag{24}$$

The problem becomes $\min_{\mathrm{d}(\Phi)} \beta\Phi_{11} + \Phi_{22}$, subject to

$$0 \leq RHS\left(\mathrm{d}\left(\Phi\right)\right) \tag{25}$$

Step 2. Constraint (25) must be binding. If it is not then a small decrease of Φ_{11} will not break this constraint, but will improve the value of the objective.

Step 3. Construct the Lagrangian $\beta\Phi_{11} + \Phi_{22} - \mu RHS\left(\mathrm{d}\left(\Phi\right)\right)$, where μ is the multiplier. The first order conditions, after eliminating μ, give the tangency condition:

$$\beta = \frac{RHS_1\left(\mathrm{d}\left(\Phi\right)\right)}{RHS_2\left(\mathrm{d}\left(\Phi\right)\right)} \tag{26}$$

Equation (24) can be simplified somewhat to

$$RHS\left(\mathrm{d}\left(\Phi\right)\right) = \Sigma_{11}\Phi_{22} + \Sigma_{22}\Phi_{11} - 2\Sigma_{21}\sqrt{\Phi_{11} - \Sigma_{11}}\sqrt{\Phi_{22} - \Sigma_{22}} - Const \tag{27}$$

where $Const = \left(\Sigma_{21}\right)^2 + \Sigma_{11}\Sigma_{22} + e^{-2r}$. Note that after defining

$$\Lambda = \frac{\sqrt{\Phi_{22} - \Sigma_{22}}}{\sqrt{\Phi_{11} - \Sigma_{11}}} > 0 \tag{28}$$

the derivatives of the $RHS\left(\mathrm{d}\left(\Phi\right)\right)$ can be expressed as

$$RHS_1\left(\mathrm{d}\left(\Phi\right)\right) = \Sigma_{22} - \Sigma_{21}\Lambda > 0 \tag{29}$$

$$RHS_2\left(\mathrm{d}\left(\Phi\right)\right) = \Sigma_{11} - \Sigma_{21}\frac{1}{\Lambda} > 0 \tag{30}$$

The tangency condition (26) determines a quadratic equation in Λ

$$\Sigma_{21}\Lambda^2 + \left(\beta\Sigma_{11} - \Sigma_{22}\right)\Lambda - \beta\Sigma_{21} = 0$$

leading to an explicit value for Λ

$$\Lambda = \left(\frac{1}{2\Sigma_{21}}\right)\left(-\left(\beta\Sigma_{11}-\Sigma_{22}\right)-\sqrt{\left(\beta\Sigma_{11}-\Sigma_{22}\right)^2+4\beta\left(\Sigma_{21}\right)^2}\right) \tag{31}$$

Using (28) to recover the family of tangency points, we realize that it is therefore linear

$$\Phi_{22} = \Lambda^2\left(\Phi_{11}-\Sigma_{11}\right)+\Sigma_{22} \tag{32}$$

Combining (32) with the constraint (27) and Step 2, we obtain an explicit solution for Φ_{11}

$$\Phi_{11} = \frac{\left(\Sigma_{21}-\Sigma_{11}\Lambda\right)^2+e^{-2r}}{\Sigma_{11}\Lambda^2+\Sigma_{22}-2\Sigma_{21}\Lambda} \tag{33}$$

Then equation (32) can be used to recover Φ_{22}, and equation (23) to recover Φ_{21}.

Step 4. Plugging in the parameters of the problem into these expressions we obtain $\Lambda \approx 0.1754$, $\Phi_{11} \approx 0.6367$, $\Phi_{22} \approx 0.2619$, and $\Phi_{21} \approx -0.0822$. Constraints (20) are not binding, and indeed $\Phi_{21} < 0$, as asserted in Step 1.

Step 5. The payoff of the sender from using this policy is

$$\lambda \max\{0.25, 0.25\} + (1-\lambda)\max\{0.2619, 0.6367\} = 0.3586$$

Literature

Admati, Anat R. and Paul Pfleiderer, (1986), A Monopolistic Market for Information, *Journal of Economic Theory*, 39, pp. 400-438.

Alaoui Larbi, and Fabrizio Germano, (2014), Time Scarcity and the Market for News, B*arcelona GSE Working Paper Series*, Working Paper nº 675.

Chan, Jimmy and Wing Suen, (2008), A Spatial Theory of News Consumption and Electoral Competition, *Review of Economics Studies*, 75(3): 699-728.

Cover, Thomas M. and Joy A. Thomas, (2006), *Elements of Information Theory*, John Willey & Sons.

Crawford, Vincent P. and Joel Sobel, (1982), Strategic Information Transmission, *Econometrica*, Vol. 50, No. 6, pp. 1431-1451.

Dessein, Wouter, Andrea Galeotti and Tano Santos, (2013), Rational Inattention and Organizational Focus, mimeo.

Gentzkow, Matthew and Jesse M. Shapiro, (2008), Competition and Truth in the Market for News, Journal of Economic Perspectives, Vol 22, No 2, pp 133-154.

Gill, David, Daniel Sgroi, (2012), The Optimal Choice of Pre-Launch Reviewer, *Journal of Economic Theory*, Volume 147, Issue 3, pp. 1247–1260.

Glazer, Jacob and Ariel Rubinstein, (2004), On Optimal Rules of Persuasion, *Econometrica*, Vol. 72, No. 6, pp. 1715-1736.

Hörner, Johannes and Andrzej Skrzypacz, (2011), Selling Information, Cowles Foundation Discussion Paper No. 1743R, Yale University

Kamenica, Emir, and Matthew Gentzkow, (2011), Bayesian Persuasion, *American Economic Review*, 101(6): 2590–2615.

Kwiek, Maksymilian, (2010), Competition among mass media, Southampton Discussion Paper in Economics and Econometrics 1013.

Matějka, F. and A. McKay, (2015), Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model, *American Economic Review*, 105(1): 272-98.

Matějka, F., J. Steiner and C. Stewart, (2015), Rational Inattention Dynamics: Inertia and Delay in Decision-Making, manuscript.

Sah, R. K. and J. E. Stiglitz, (1986), The architecture of economic systems: Hierarchies and polyarchies, *American Economic Review*, 76, 716–727.

Sims, Christopher A., (2003), Implications of Rational Inattention, *Journal of Monetary Economics*, 50(3), 665-690.

Sims, Christopher A., (2006), Rational Inattention: A Research Agenda, mimeo.

Yang, Ming (2015), Coordination with flexible information acquisition, *Journal of Economic Theory*, 158, Part B (0) Symposium on Information, Coordination, and Market Frictions, pp. 721-738.

Yamamoto, Hirosuke, (1981), Source Coding Theory for Cascade and Branching Communication Systems, *IEEE Transactions on Information Theory*, Vol. IT-27, No. 3, May 1981 299