UNIVERSITY OF Southampton
School of Social Sciences

Economics Division
University of Southampton
Southampton SO17 1BJ, UK

Discussion Papers in
Economics and Econometrics

Title Mass Media and Constrained Communication

By Maksymilian Kwiek

No. 1204

This paper is available on our website
http://www.southampton.ac.uk/socsci/economics/research/papers

ISSN 0966-4246

Mass Media and Constrained Communication*

Maksymilian Kwiek†

May 04, 2012

Abstract

This paper studies how a sender optimally communicates information to receivers when communication is coarse, but the sender's interests are aligned with those of the receivers. Results from information theory, based on the notion of mutual information, are used to derive equilibria characterizing monopolistic behavior in this constrained communication model. The main finding is that when communication constraints are present, the correlation between receivers' actions is typically greater than in a frictionless environment. If constraints are severe enough, this correlation is one. The result appears to be robust to different specification of industrial structure and suggests a theoretical background to studies of mass media as a source of herding or contagion.

Keywords: Communication, mass media,

JEL: C72, D83

*Preliminary

†University of Southampton, M.Kwiek@soton.ac.uk

1 Introduction

Newspapers provide information to many diverse readers. At the same time the readers are busy and have just enough time to flip through the headlines. How should the newspaper select the topics, whom should they inform and how should they do it? What is an effect of mass media on readers' actions?[1]

This note formulates a simple *economic model of constrained communication*, in which the receivers have to obey a capacity constraint on how much information they can read. This approach is novel in economics, because the focus is not – as is in the current rich and growing literature on communication games – on the conflict of interests between the sender and the receiver (e.g. Crawford and Sobel (1982), Hörner and Skrzypacz (2011) or Kamenica and Gentzkow (2011), etc.). It is assumed here that there is no such conflict, so – without communication constraint – perfect communication should ensue.[2] The main finding is that when communication constraints are present, the correlation between receivers' actions is typically greater than in a frictionless environment. If constraints are severe enough, this correlation is one.

There are three innovations in this paper.

Constrained communication. Even if the presentation below interprets the sender as a mass medium such as a newspaper or a TV channel,[3] there are other examples of social institutions in which constrained communication of the type studied here seems to play some role. Rating agencies must express a complicated situation by means of a simple rating. Educators must

[1] The following quote emphasises both the source of friction and its economic repercussion: "Seen from London and other places that might still be called the commanding heights of global finance, the countries of eastern Europe look much of a muchness. With the big exception of Russia, the rest tend to merge when viewed by crisis-weary traders glued to their screens. Sell one, sell all has been the motto. (...) Journalists too, including this one, often put everything together under one headline. (...) With little space and time what else can be done?" – "How to annoy someone from central or eastern Europe", by Stefan Wagstyl, *The Financial Times*, February 27, 2009.

[2] Since this is not a cheap talk game in the Crawford and Sobel (1982) tradition, but rather the one in which the sender commits to a decision on an informational policy prior to receivers choosing their actions, cheap talk bubbling equilibria do not emerge.

[3] A recent survey of political economy literature on the role of mass media in politics and policy is Prat and Strömberg (2011).

decide what to put in a curriculum. Pundits have just minutes to explain complex social issues on evening news, and politicians have only political slogans to present their election platforms to voters. Some effort has been put into studying behavior under, and economic implications of, communication with capacity constraints,[4] but this literature does not use the tools adopted by this paper. Techniques essential to handle the model studied below are provided by information theory, started by the seminal Shannon's 1948 paper. To my knowledge, information theory has been used by economists in a somewhat different context of rational inattention models, e.g. Sims (2003), (2006). Information theory is concerned with optimal communication when constraints on communication are present, usually in an engineering context of data compression or communication between servers, satellites, or components of a computer. In contrast, this note studies constrained communication between noncooperative economic agents; one sender and *multiple* receivers, for that matter. Thus, it is necessary to cast a standard information theoretic model in such a way that the trade-offs facing the sender can be investigated. For example, one approach in information theory is to study a relationship between some measure of available communication capacity, R, and some measure of distortion at the receiver, D. This results in a function $R(D)$, describing the boundary of feasible distortion-capacity pairs. This note focuses on such functions with multidimensional domain, where each separate dimension is interpreted as a distortion (or disutility) of a separate receiver.

Correlation among the readers. Regarding the effects that mass media may have on readers' actions in this context, one result stands out. Mass media have a striking ability to magnify the correlation of actions across their customers, relative to frictionless environment. That is, when communication constraints are present, the correlation between receivers' actions is typically greater than when communication is frictionless. This effect can be phenomenally strong: if communication constraint is severe enough, all

[4] 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 (2011).

receivers act in complete unison, i.e., the correlation between their actions is perfect.

A correlating effect of mass media has been acknowledged in the literature. There is a stream of thought in sociology or theory of mass communication which views mass media as an institution that socializes individuals, correlates actions and integrates diverse populations (e.g. see McQuail (2010)). Closer to economics, co-movement across diverse industries is one of the defining features of business cycle (Christiano and Fitzgerald (1998)). It has been proposed that frictions in information transmission in mass media are key to explaining such co-movement across industries and geographical regions, e.g. Calvo and Mendoza (2000), Veldkamp (2006a, 2006b). However, in the study below, the receivers are completely isolated, apart from the fact that they see a common message. Hence, there is no underlying strategic complementarity that could be a source of high correlation among their actions.[5]

Flexible product design.[6] The receivers' utility possibility frontier, embedded in the communication constraint, $R(D) \leq R$, says how the newspaper can trade-off distortions across the readers. This utility possibility frontier seems to be a natural starting point for the analysis of product design, which – to my knowledge – has not been used in this context. The uniformly-pricing and profit-maximizing monopolist-sender can quite flexibly design her informational policy not only to create, but also to allocate value across the customers. This flexibility in shaping a demand is closely related to the concept of demand rotation introduced by Johnson and Myatt (2006), but here the demand is derived from the primitive notion of the utility possibility frontier. Whenever there is a chance to equalize the valuations of the buyers,

[5] The study below is also remotely related to models of informational herding, such as in Banerjee (1992), where the signal conveyed through predecessors' actions may be strong enough for an agent to choose the same action regardless of her own signal so that herding may ensue. The key factor is the coarsnes of the action set – otherwise agents would tailor their actions in a flexible manner, thus revealing their private signal to their successors. This resembles the communication constraint in the model below, which – if relaxed a little – would lead to a better guess of the true state of nature.

[6] Mass media are sometimes given as an example in models of product design or product differentiation. Footnote 24 in Johnson and Myatt (2006) mentions radio stations, while Irmen and Thisse (1998) give an example of *Newsweek* and *Time*.

the monopolist would like to do that. The reason is that with uniform prices, the price is equal to the valuation of the buyer with the lowest valuation, so there is an unambiguous benefit from increasing this lowest value, even if that means that values of some other receivers – buyers and non-buyers – are driven down. Consequently, the demand function exhibits a kink at an equilibrium point: values are similar or even equal among all buyers, but drop discontinuously for the non-buyers. In a flexible enough environment, the monopolist can extract full created surplus. This is not because she changes prices to fit the valuation, as in price-discrimination theory, but precisely for the opposite reason, because she can change the valuations to fit some uniform price. In general, not all potential surplus may be created, but in symmetric environments, the uniformly-pricing monopolist may be fully efficient.

Section 2 introduces the model and the concept of constrained communication. The key information theoretic result, allowing the characterization of this constraints, is laid down in Section 3. Section 4 discusses the problem of the monopolist sender and the social planner. Section 5 provides a solution algorithm for the Gaussian-quadratic case. In the context of this case, Section 6 explains the correlating effect of the communication constraint. Some comments on possible extensions are made in Section 7. Proofs are in Appendix A.

2 Constrained communication: the model

Preliminary comments. Any model of communication must have at least two locations, devices or agents, between which communication occurs. The first location, the sender, can inform the other location, the receiver, about the realization of some state of nature. In this note, by constrained communication we mean that the sender cannot inform the receiver perfectly, due to some coarsness of the message space. This could be a result of a combination of two reasons. Firstly, any communication is a physical process that takes time or space, e.g. it takes one minute to read one headline. Secondly, the receivers are impatient or otherwise constrained and they will devote finite time to read the sender's message, e.g. the receiver has 10 minutes. As a consequence, the sender must pack its message in ten headlines.

To be more formal, let the state of nature be represented by a random variable X whose realization is $x \in \mathcal{X}$. If we are given R binary digits (bits) to represent a realization of this random variable, then the available number of messages is 2^R. Let $f(x) \in \{1, ..., 2^R\}$ be a message that is generated from state x; function f is called an encoding function. A decoding function, $g : \{1, ..., 2^R\} \rightarrow \mathcal{X}$, provides an estimate, $\hat{x}$, of the realization, x, as a function of observed messages. That is, $g(f(x)) = \hat{x}$. If the set of states of nature is richer than the number of messages, then the estimator will not be able to replicate the original state of nature perfectly.

One of the aims of information theory is to explain how one can construct f and g so that some measure of expected loss associated with this communication is minimal. Next section applies this structure to an economic model.

2.1 Economic model

Physical environment. There are N *receivers*; each of them has an optimal action which is unknown to her, but is interested in learning it. This optimal action will be called her *state of nature*. The state of nature of receiver n is represented by a random variable X_n, the realization of which is $x_n \in \mathcal{R}$. These states of nature are not necessarily independent across the receivers. A multivariate random variable $X = (X_1, ..., X_N)$ has a joint probability distribution $p(x)$, $x \in \mathcal{R}^N$. Later, we will focus on Gaussian distribution.[7]

There is a *sender*, who knows the state of nature of each receiver. The sender is to inform the receivers about the state of nature by sending a *common message*. Any communication from the sender to the receivers must take a form of a stream of binary digits.

Timing. Many markets and institutional forms can be modelled using

[7]Receivers' heterogeneity can be backed up by a Lancastrian approach to modelling the demand side of the market (Lancaster (1966)). Individuals may be interested in learning the realization of a certain combination of some aspects of reality, e.g. one individual may be interested in a combination of mostly sports news and some weather news, while some other individual may be interested in the half-and-half combination of sports and weather news. In this paper, x_n is assumed to represent this ultimate combination directly.

the following common structure. In the first stage, the sender announces an *encoding* function f, which maps states of nature into the messages. In the second stage, all the transactions occur and – in particular – the audience is decided. It is useful to leave the exact form of this mechanism unspecified at the moment, but various mechanisms can be investigated. For instance, receivers may have unlimited access to the messages; or a monopolistic sender may restrict the audience by charging a positive price, etc. In the third stage, the sender learns the realization of the state of nature and sends the message according to the encoding function. Finally, in the fourth stage, receivers choose their actions, $\hat{x}_n$, and their payoffs are realized. The function that maps the sender's messages into receivers' actions is called a *decoding* function, g_n.

Receivers' problem. Receiver n cares about estimating the unknown realization of state x_n. Her action is denoted $\hat{x}_n \in \mathcal{R}$, and $\hat{X}_n$ is the action as a random variable induced by a strategy profile. Written as a vector, $\hat{x}$ is the realization of $\hat{X}$. Let $d_n(x_n, \hat{x}_n) \geq 0$ be the measure of distortion or disutility associated with selecting $\hat{x}_n$ when the true state is x_n. Later, we will focus on the case where this distortion is the square distance between $\hat{x}_n$ and x_n. Let $D_n = E_{X,\hat{X}} d_n(x_n, \hat{x}_n)$ be the ex ante distortion of nth receiver, induced by f and g. Let $\bar{D}_n$ be the distortion level of nth receiver induced by her optimal behavior, if she does not receive any messages from the sender. Hence, $0 \leq D_n \leq \bar{D}_n$, and the ex ante willingness to pay for information provided by the sender is the reduction in distortion attributed to reading the sender's message, $\bar{D}_n - D_n$.

The receiver will read the message broadcast by the sender and consisting of a stream of binary digits, but can physically read only at a rate of R bits. It is assumed that this capacity is the same for all receivers.

Sender's problem. Incentives of the sender are described by some non-decreasing function W that maps receivers' expected distortions $D = (D_1, ..., D_N)$ into real numbers. Notice that there is no conflict of interest in communication between the sender and the receivers in the sense that both the sender and receiver n benefit from lowering D_n, if distortions of all the receivers apart from n are given. Alternatively, the seller could benefit from inducing some action of the receiver that the receiver would not want to choose; a case like that is briefly mentioned in Section 7.

The above general model permits many industry structures. For example, the sender may be a monopolist who cannot price-discriminate. In the second stage of the model above, the monopolist posts a single subscription price and then the receivers decide whether to buy the subscription. Since the receivers' willingness to pay for a subscription depends on the profile of distortions D, monopolist's profit will also depend on it. Secondly, the sender may be a price-discriminating monopolist and therefore may be interested in maximizing the total sum of receivers' willingnesses to pay. The same would happen if the sender was a benevolent broadcaster, whose objective function is the minimization of the sum of distortions. Clearly, the sender's objective function again depends directly on D. These examples will be studied in more detail in Section 4 but they are not exhaustive and other interesting cases can be postulated. For example, one can consider a reduced-form model of an advertising newspaper, in which the sender's payoff depends on the number of active readers, subject to a constraint that the gain of an individual receiver is not smaller than some exogenous cost of buying and reading the message. As long as objective of the sender can be expressed as a function $W(D)$, the analysis is relevant.

2.2 Rate distortion function

Different encoding-decoding functions can lead to different expected distortion vectors, for a given rate – or to different rates, if a particular distortion vector is to be obtained. Consider a problem of finding the most efficient encoding-decoding pair, in the following sense: for a given profile of allowed expected distortions D, define $R^c(D)$ to be the minimum rate that is needed if the expected distortion of receiver n is to be at most D_n, where the search is over various encoding-decoding functions. The superscript indicates that this is a "centralized" problem, as one discards the incentive compatibility constraint of the receivers and assumes that they simply behave according to function g. Function $R^c(D)$ is called a *rate distortion function.* This function is key in the analysis below because it describes the boundary of feasible communication.

Let $R^d(D)$ be the minimum rate that is needed if the expected distortion of n is to be at most D_n, where the choice variable is the encoding-decoding functions, subject to the receivers' incentive compatibility constraint – a constraint that says that all receivers are free to select any decoding function. The superscript indicates that this is a "decentralized" problem.

Clearly, $R^c(D) \leq R^d(D)$. Since there is no conflict of interests between the sender and the receivers, when encoding function is already selected, the receivers cannot benefit from deviating to a different decoding function than the one selected by the sender. Hence

Lemma 1 $R^c(D) = R^d(D)$.

3 Information theory: basic implications

The purpose of this section is to present a convenient operational characterization of the rate distortion function $R^c(\cdot)$. Firstly, define mutual information: for a given joint distribution $p_{x,y}(x,y)$ of random variables (X,Y) let the mutual information of X with Y be:

$$I(X;Y) = E_X(-\log p(X)) - E_{X,Y}\left(-\log p_{x|y}(X|Y)\right)$$

Mutual information is the difference between two terms called entropy, measuring the uncertainty pertaining to a certain random variable. The first term is the entropy of X, and the second term is entropy of X conditional on Y. Thus, mutual information can be interpreted as a reduction in uncertainty that follows from learning Y. In our model we are interested in mutual information of X with $\hat{X}$, that is in reduction of uncertainty in X when $\hat{X}$ is recommended to the receiver.

The following definition is key.

Definition 1 *The* information rate distortion function $R(D)$ *for random variable* X *with distortion measure* $d_n(x_n, \hat{x}_n)$ *is*

$$\begin{aligned} R(D) &= \min_{p_{\hat{x}|x}(\hat{x}|x)} I\left(X; \hat{X}\right) \ \ s.t. \\ & E_{X,\hat{X}} d_n(x_n, \hat{x}_n) \leq D_n, \ \textit{for all } n = 1, ..., N \end{aligned} \tag{1}$$

It seems that the definition of $R(D)$ is totally unmotivated and arbitrary. At first glance, it is difficult to judge what the relationship between $R(D)$ defined in (1) and $R^c(D)$ is. However, it can be shown that the value of $R^c(D)$ is not smaller than of $R(D)$ and that these two numbers are equal in a certain limiting sense. This is an incarnation of one of the fundamental results in information theory. Since this result is adequately presented in the information theoretic literature (e.g. Cover and Thomas (2006), Ch. 10), its short discussion is relegated to Appendix B. For the reminder of this paper, *we will use the information rate distortion function defined in (1) as the characterization of the rate distortion function* $R^c(D)$, discarding the superscript in the latter along the way.

It will be sometimes convenient to use an alternative definition of the information rate distortion function. Let $a = (a_1, ..., a_N)$ be any vector of coefficients, such that $a \in \Delta^N$. Define

$$\begin{aligned} \bar{R}(a, B) &= \min_{p_{\hat{x}|x}(\hat{x}|x)} I\left(X; \hat{X}\right) \ \text{s.t.} \\ & \sum_{n=1}^{N} a_n E_{X,\hat{X}} d_n(x_n, \hat{x}_n) \leq B \end{aligned}$$

Namely, the set of constraints in $R(D)$ is replaced by their weighted average. The functions $R(D)$ and $\bar{R}(a, B)$ contain the same information. The relationship between them is

Lemma 2 $\bar{R}(a, B)$ *can be obtained from* $R(D)$

$$\bar{R}(a, B) = \min_{D: \sum_n a_n D_n \leq B} R(D) \tag{2}$$

$R(D)$ can be obtained from $\bar{R}(a,B)$

$$R(D) = \max_{(a,B):\sum_n a_n D_n = B} \bar{R}(a,B)$$

The proof is in Appendix A and it follows from the fact that $R(D)$ is a convex function.

Lemma 3 *$R(D)$ given in equation (1) is a non-increasing, continuous and convex function of D.*

4 Market form

Define the Utility Possibility Frontier as the set of feasible distortions, with a property that lowering distortion of one receiver implies that some other receiver's distortion must increase,

$$UPF = \{D : R(D) \leq R \text{ and } D' \leq D \Rightarrow R(D) > R\}$$

If there are two receivers then one can illustrate the situation graphically, as on Figure 1, where the axes are inverted, so that they indicate the receivers' values $v_n = \bar{D}_n - D_n$. The set of feasible designs, satisfying $R(D) \leq R$, is marked by the shaded area.[8] The North-East frontier of the feasible set, indicated by a thicker curve, is the set of designs that are Pareto efficient for the receivers; this is the UPF. For example, point A shows a newspaper that is the best for receiver 1, while point C indicates a design that is the best for receiver 2. By Lemma 3 the set $\{D : R(D) \leq R\}$ is convex.

The purpose of this section is to show how UPF can be used in the analysis of a number of different versions of the above model. One has to specify the precise nature of interactions in the second stage and, in particular, the form of function W. One observation that is important later in this paper

[8] To be precise, the shaded area is the set of designs that satisfy $R(D) \leq R$, but this set may include designs that are *not* feasible. Suppose that the set of trully feasible designs is the convex set $OGCAH$. The set characterised by $R(D) \leq R$ will include the entire shaded area, because of the inequalities in the constraints of problem (1). This distinction is inconsequential as far as the set UPF is concerned.

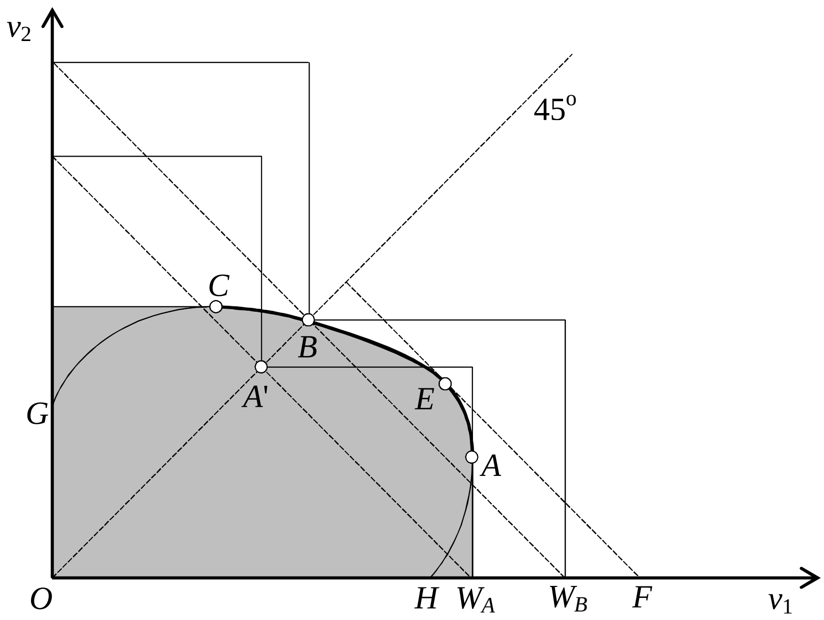

Figure 1: Utility Possibility Frontier

is the following: in industry structures considered in this paper, an optimal newspaper design is a point in the UPF.

4.1 Monopolist setting uniform prices

Suppose that the sender is a profit-maximizing monopolist who cannot price discriminate. That is, in the second stage of the model above, the monopolist posts a single subscription price, which can be accepted or rejected by the receivers. If receiver n rejects, her loss is $\bar{D}_n$.

Take as given the set $\mathcal{N}^*$ of receivers who are to buy the message. Each receiver is willing to pay at most $\bar{D}_n - D_n$ for the subscription, so the maximal price that the monopolist can charge and still expect all receivers in set $\mathcal{N}^*$ to buy the newspaper is

$$\min_{n \in \mathcal{N}^*} \left\{ \bar{D}_n - D_n \right\}.$$

It is assumed that costs of newspaper production are zero. The round brackets in the following expression contains the revenue associated with set $\mathcal{N}^*$, and the whole expression defines the profit associated with a given informational policy D.

$$W(D) = \max_{\mathcal{N}^* \subseteq \mathcal{N}} \left(N^* \min_{n \in \mathcal{N}^*} \left\{ \bar{D}_n - D_n \right\} \right)$$

If there are two receivers, then graphical presentation is possible. The iso-profit curves for the uniformly-pricing monopolist are also drawn on Figure 1. These are step-like line segments with kinks on the 45^o line. For example, newspapers A and A' achieve the same profit, although through a different pricing policy. Newspaper A sells only to receiver 1 for a price W_A, while newspaper A' sells to both receivers for a price $W_A/2$. The total profit is clearly W_A in both cases. In this example, the profit-maximizing design, indicated by point B, involves selling to both receivers and the monopolist's profit is W_B.

In this two-receiver case, it is easy to observe that the only candidate optimal designs are either the most extreme points, such as A or C, or the most symmetric point, such as point B on the 45^o line. None of the remaining points on the UPF, such as E, is optimal. To see this, suppose that E is proposed. If the monopolist sells to one receiver, then E is not optimal because there exists a better design for this buyer and so higher price can be extracted. If the monopolist sells to both receivers, then the price is equal to the lowest valuation among the two. But this lowest valuation can be increased by lowering higher valuation a little, that is, by moving along the UPF towards the 45^o line. In both cases the price can be increased without reducing the quantity demanded, so E is not optimal.

This observation is closely related to the result by Johnson and Myatt (2006). They postulate that a monopolist can select policies that rotate the demand and conclude that this monopolist would either create a niche market (that is, to select the policy that makes the demand as vertical as possible to charge a high price and sell low quantity), or a mass market (that is to select a policy that makes the demand as flat as possible, to sell as many units as possible at a low price). In the two-receiver example above, designs A and C are niche products and B is a mass product.

The obvious general result is the following

Proposition 1 *If D is an optimal newspaper design for the uniformly pricing monopolist, with $\mathcal{N}^*$ being the set of buyers, then D maximizes the Leontief function* $\min_{n \in \mathcal{N}^*} \left\{ \bar{D}_n - D_n \right\}$.

Therefore, the monopolist behaves as a Rawlsian welfare maximizer vis-a-vis the receivers to whom it plans to sell the newspaper.

In a classical model of a monopolist who faces a given demand, there is no sense in which the product can be designed – this situation is captured by assuming that the UPF is just a single point. It is useful to contrast this model to another polar case, the one of full flexibility, in which the set UPF extends to every axis.

Proposition 2 *Assume that for every receiver n, there is a distortion profile with D_n such that $\left(0, ..., 0, \bar{D}_n - D_n, 0, ..., 0\right) \in UPF$. Then in equilibrium, all receivers are left with zero net surplus.*

This result resembles the one that emerges in perfect price discrimination. However, full extraction of created surplus occurs not because the monopolist can tailor prices to fit the buyers' valuations, but precisely the opposite reason – the monopolist can flexibly affect the valuations in order to fit them to some uniform price. The resulting (endogenous) demand is a perfect step function. All receivers who buy have the same valuation, equal to price, and all receivers who abstain from buying have valuation equal to zero. However, this very particular demand function is not assumed, but is endogenously generated by optimal product design. In particular, full surplus extraction does not mean that the monopolist maximizes the utilitarian welfare (as will be seen in the next subsection).

If the monopolist faces a situation somewhere between a fully inflexible classical monopolist and fully flexible monopolist in Proposition 2, then some surplus may be created for non-buyers and some net surplus may be left with buyers, but still there is likely to be some "gap" in valuations between buyers and non-buyers, as the monopolist wants to increase the valuations of the buyer with the lowest valuation at the expense of the valuations of non-buyers. Flexibility in product design gives them some opportunity to create this gap.

4.2 Utilitarian welfare

Suppose that the sender is interested in maximizing the utilitarian welfare function

$$W(D) = \sum_{n} \left(\bar{D}_n - D_n\right)$$

This may be because the sender is a profit-maximizing monopolist who can price discriminate, or because it is a non-for-profit mass medium whose objective is to maximize social welfare.

In the example depicted on Figure 1, the iso-welfare lines are lines with the slope -1, such as $\overleftrightarrow{BW_B}$ or $\overleftrightarrow{EF}$. Point E maximizes utilitarian welfare and the resulting total social welfare is represented by the distance form zero to F. Clearly, the optimal newspaper of uniform-pricing monopolist does not have to be efficient in this sense. Even if the uniform-pricing monopolist extracts all created surplus, not all potential surplus is created.

Note also that it is difficult to observe the level of inefficiency associated with a monopolist, even if an investigator observed the valuations of the receivers. One can observe that both receivers are served by the monopolist, so there is no reduction in quantity supplied known from the classical monopolist facing a fixed demand. Inefficiency comes from the fact that actual valuations of the receivers are endogenously shaped by the monopolist facing a constraint that price must be uniform.

5 Gaussian-quadratic case

Next we assume that the source is Gaussian, $X \sim N(0, K)$, where $K \succeq 0$ is a covariance matrix,[9] and the distortion is square distance, $d_n(x_n, \hat{x}_n) = (x_n - \hat{x}_n)^2$.

Let H be the covariance matrix of a random variable $Z = X - \hat{X}$.

The Gaussian-quadratic example is convenient because, still being rich enough to give interesting results, it simplifies the analysis in three ways. Firstly, it can be shown that the distribution solving the problem in Defi-

[9] $K \succeq 0$ means that matrix K is positive semidefinite.

nition 1 must be Gaussian, and so the only free variable is the covariance matrix H. Secondly, the objective in the definitions of the rate distortion functions is relatively tractable, because there is a simple expression for mutual information if the underlying distributions are Gaussian. Thirdly, the constraint that for every n the expected distortion cannot exceed D_n translates into $H_{nn} \leq D_n$. In the Gaussian-quadratic case, the functions $R(D)$ and $\bar{R}(a,B)$ can be computed by solving simpler problems:

Lemma 4 *Under the Gaussian-quadratic assumption*

$$\begin{aligned} R(D) &= \min_H (1/2) \log(\det K / \det H) \;\; \textit{s.t.} \\ & \quad H \succeq 0,\ K - H \succeq 0,\ H_{nn} \leq D_n \textit{ for all } n \end{aligned} \tag{3}$$

$$\begin{aligned} \bar{R}(a,B) &= \min_H (1/2) \log(\det K / \det H) \;\; \textit{s.t.} \\ & \quad H \succeq 0,\ K - H \succeq 0,\ \sum_{n=1}^{N} a_n H_{nn} \leq B \end{aligned} \tag{4}$$

5.1 Solution

The aim of this section is to provide an algorithm to derive $\bar{R}(a,B)$; then Lemma 2 can be used to recover $R(D)$.

Let $A = \operatorname{diag} a$ be a matrix with a on the main diagonal and zero elsewhere. After changing variables $K_a = A^{1/2} K A^{1/2}$ and $H_a = A^{1/2} H A^{1/2}$, we can write (4) as[10]

$$\begin{aligned} \bar{R}(a,B) &= \min_{H_a} (1/2) \log(\det K_a / \det H_a) \;\; \text{s.t.} \\ & \quad H_a \succeq 0,\ K_a - H_a \succeq 0,\ \operatorname{tr}(H_a) \leq B \end{aligned} \tag{5}$$

To solve (5), let $K_a = Q \Lambda Q'$ be the eigendecompostion of K_a, where Q is a matrix containing orthogonal eigenvectors, and $\Lambda = \operatorname{diag} \lambda$ is a diagonal

[10] Note that (i) $\det K / \det H = \det\left(A^{1/2} K A^{1/2}\right) / \det\left(A^{1/2} H A^{1/2}\right)$, (ii) $H \succeq 0 \Leftrightarrow A^{1/2} H A^{1/2} \succeq 0$, and $K - H \succeq 0 \Leftrightarrow A^{1/2} K A^{1/2} - A^{1/2} H A^{1/2} \succeq 0$, (iii) $\sum_{n=1}^{N} a_n H_{nn} \leq B \Leftrightarrow \operatorname{tr}\left(A^{1/2} H A^{1/2}\right) \leq B$.

matrix with eigenvalues $\lambda = (\lambda_1, ..., \lambda_N)$. Once we write $\tilde{H} = Q' H_a Q$, the problem (5) becomes[11]

$$\bar{R}(a, B) = \min_{\tilde{H}} (1/2) \log \left(\det \Lambda / \det \tilde{H} \right) \text{ s.t.} \tag{6}$$
$$\tilde{H} \succeq 0, \; \Lambda - \tilde{H} \succeq 0, \; \operatorname{tr}\left(\tilde{H}\right) \leq B$$

This representation is useful because some properties of optimal $\tilde{H}$ become immediately apparent. In particular, if an off-diagonal element of a feasible matrix $\tilde{H}$ is strictly positive, lowering it increases the determinant in the objective function, and does not violate any of the constraints. Hence an optimal $\tilde{H}$ must be diagonal, and $\det \tilde{H}$ is just the product of its diagonal elements. Thus, the problem simplifies further to

$$\bar{R}(a, B) = \min_{\tilde{H}_{11}, ..., \tilde{H}_{NN}} (1/2) \log \left(\det \Lambda / \left(\tilde{H}_{11} \times ... \times \tilde{H}_{NN} \right) \right) \text{ s.t.}$$
$$0 \leq \tilde{H}_{nn} \leq \lambda_n \text{ for all } n, \; \sum_n \tilde{H}_{nn} \leq B$$

Setting up a constrained maximization leads to solution

$$\tilde{H}_{nn} = \begin{cases} \lambda^* & \text{if} \quad \lambda_n > \lambda^* \\ \lambda_n & \text{if} \quad \lambda_n \leq \lambda^* \end{cases} \tag{7}$$

where λ^* is selected so that $\sum_n \tilde{H}_{nn} = B$.[12]

Let $\tilde{H}(a, B)$ be the solution of this problem and let $H(a, B)$ be the solution to problem (4). We have

Proposition 3 *The solution to problem (4) is*

$$H(a, B) = A^{-1/2} Q \tilde{H}(a, B) Q' A^{-1/2}$$

where $\tilde{H}(a, B)$ is defined by (7).

[11] Note that (i) $\det K_a / \det \hat{H} = \det \Lambda / \det Q' \hat{H} Q$, (ii) $\hat{H} \succeq 0 \Leftrightarrow Q' \hat{H} Q \succeq 0$, and $K_a - \hat{H} \succeq 0 \Leftrightarrow \Lambda - Q' \hat{H} Q \succeq 0$, (iii) $\operatorname{tr}\left(\hat{H}\right) = \operatorname{tr}\left(Q' \hat{H} Q\right)$.

[12] This solution method is called the water-filling algorithm.

The following result may be of some interest

Lemma 5 *Suppose $X \sim N(0, K)$ has a rate distortion function $\bar{R}_X(A, B)$. Consider a re-scaled random variable $Y = CX$, where C is an $n \times n$ diagonal matrix of weights (hence, Y has the same correlations as X). Then the rate distortion function of Y satisfies $\bar{R}_Y(A, B) = \bar{R}_X(CAC, B)$.*

5.2 Example: Independent dimensions

Assume that receivers' states of nature are independent. The solution is particularly simple in this case. Matrix K is diagonal, so is K_a. Eigenvalues are equal to variances, $\Lambda = K_a$. To find the UPF, as characterized by $R = R(D)$, use equation (3) directly and note that H must be diagonal too. Choose a dimension i that will be determined in terms of all remaining distortions D_{-i} and level R. Choose $H_{nn} = D_n$ for all $n \neq i$. Then it must be that H_{ii} is such that $H_{ii}\Pi_{n\neq i}D_{nn} = 2^{-2R}(\Pi_n\sigma_n^2)$, provided that such H_{ii} can be found in $(0, \sigma_i^2)$. That is, the UPF has a Cobb-Douglass feel to it.

If the receivers are independent, then the monopolist is fully flexible in trading off valuations among the receivers, in the sense that the set UPF extends to every axis; the conclusion of Proposition 2 holds. Moreover,

Proposition 4 *Suppose that the receivers are independent and symmetric ($\sigma_i^2 = \sigma^2$ for every i). Then the uniformly pricing monopolist sells the message to all receivers.*

From the outset, this result is not obvious, because it is not entirely clear why the best way to inform two receivers whose states are completely independent involves creating a message that is somewhat informative to both. It turns out that trying to inform a receiver runs into a familiar phenomenon of diminishing marginal returns. That is why it is better to start reducing a distortion of another receiver, before investing more in distortion reduction of a receiver whose distortion is already low. If symmetric receivers are somewhat correlated, then intuitively there are even more incentives to sell to all of them; compare the UPF on the first two pictures of Figure 2.

5.3 Example: Two dimensions

Consider the case of two dimensions, $N = 2$. Let $\sigma_n^2 > 0$ be the variance of X_n and let $\rho \in (-1, 1)$ be the correlation coefficient. Matrix K is

$$K = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

One can follow the algorithm of section 5.1, but in this case it is easier to solve problem (3) directly. Focus on the part of the solution in which inequalities in the last constraint are satisfied with equalities; in this case, the only remaining variable to pin down is the covariance H_{12}. The problem involves the maximization of $\det H$, so clearly, H_{12} should be as close to zero as possible, while obeying both remaining semipositiveness constraints. Matrix H is positive semidefinite if and only if its determinant is nonnegative, or

$$H_{12} \in \left[0 \pm \sqrt{D_1 D_2}\right]$$

Likewise, matrix $K - H$ is positive semidefinite if and only if its determinant is nonnegative, or

$$H_{12} \in \left[\rho\sigma_1\sigma_2 \pm \sqrt{\left(\sigma_1^2 - D_1\right)\left(\sigma_2^2 - D_2\right)}\right] \tag{8}$$

Consider the case of non-negative correlation $\rho \geq 0$. Provided that the two intervals are not disjoint, the solution is this: if the lower bound of the interval in (8) is negative, then optimal H_{12} should be zero. Otherwise, it should be equal to this lower bound. Hence

$$H_{12} = \max\left\{0, \rho\sigma_1\sigma_2 - \sqrt{\left(\sigma_1^2 - D_1\right)\left(\sigma_2^2 - D_2\right)}\right\}$$

Figure 2 shows the level curves of the function $R(D_1, D_2)$ for three different cases (with equality constraint, $H_{nn} = D_{nn}$), where the axis are flipped so that they measure receiver's valuation, $v_n = \sigma_n^2 - D_n$, rather than her distortion D_n.

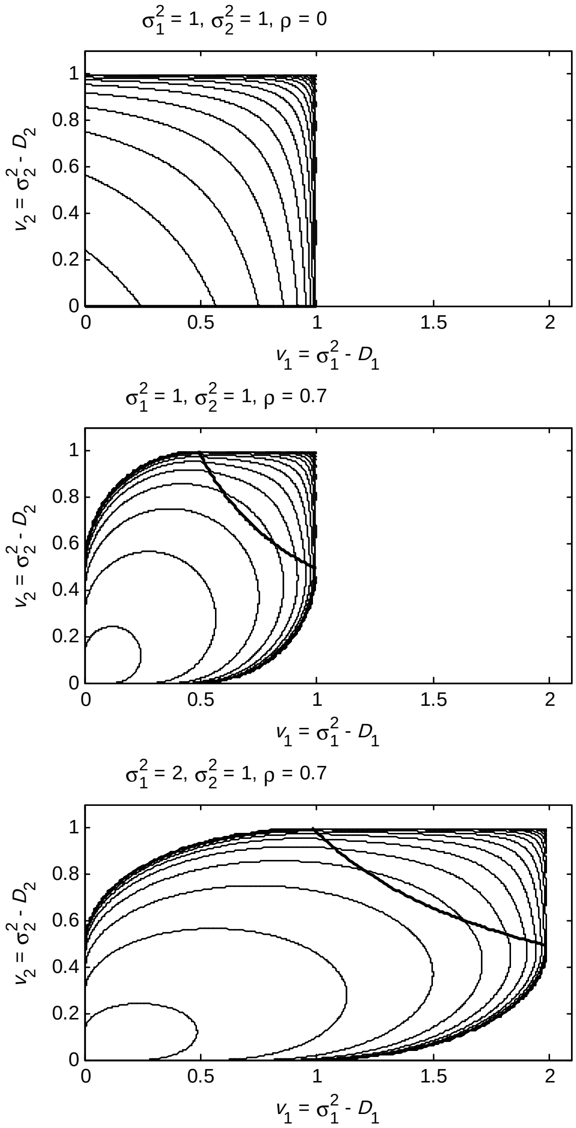

Figure 2: Level curves of $R\,(D_1, D_1)$

6 Correlation among receivers

The second example above leads to an interesting observation. The correlation of the states of nature among the receivers (optimal but unknown actions) is $\rho \geq 0$, while the correlation of the actual actions, if $\sigma_n^2 > D_n$, is

$$\begin{aligned} r_{12} &= (K_{12} - H_{12}) / \sqrt{(\sigma_1^2 - D_1)(\sigma_2^2 - D_2)} \\ &= \min\left\{ \rho\sigma_1\sigma_2 / \sqrt{(\sigma_1^2 - D_1)(\sigma_2^2 - D_2)}, 1 \right\} \end{aligned}$$

That is, the only case when these two correlations are equal is when the dimensions are independent, $\rho = 0$. Otherwise, communication constraint works as a magnifier of correlation in the following sense: if the correlation of unknown optimal actions is positive, $\rho > 0$ then

1. $r_{12}(D)$ is increasing in D, and hence $r_{12}(D) > \rho$.
2. There is a finite $R^* > 0$ such that for all $R < R^*$, optimal communication has $r_{12}(D) = 1$.

The same holds for the case of negative ρ, but the correlation increases in absolute terms and becomes negative one.

The second of these points partially generalizes to the case $N > 2$. Let $K > 0$ mean that all elements of K are strictly positive.

Proposition 5 *Assume Gaussian-quadratic case and let $K > 0$. There exists $R^* > 0$ such that for all $R < R^*$, optimal communication implies that the correlation among receivers' actions is one.*

The first observation above does not necessarily hold if $N > 2$; that is, it is not always true that function $r_{nm}(.)$ is monotonic. In practice, it often is. In particular, it is if all correlations are the same, $\rho_{nm} = \rho$, for all n, m.

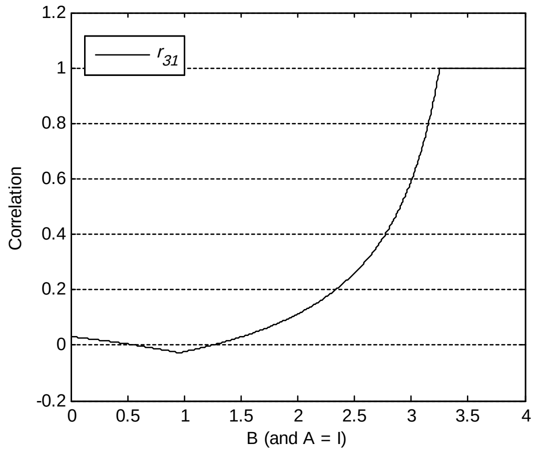

Figure 3: Correlation r_{31} as a function of B (for $A = I$)

However, consider an example in which

$$K = \begin{bmatrix} 1 & & & \\ .8 & 1 & & \\ .03 & .03 & 1 & \\ .2 & .5 & .8 & 1 \end{bmatrix}.$$

Let us express the correlations as a function of (a, B). The correlation between actions of the first and the third receivers, $r_{31}(a, B)$, may be decreasing in B. In fact, on a certain interval this correlation becomes negative (and obviously less than $\rho_{31} = .03$), see Figure 3.

Secondly, the requirement $K > 0$ in the Proposition is needed because if some correlations ρ_{nm} are negative, then the conclusion of the Proposition does not have to hold. Suppose that

$$K = \begin{bmatrix} 1 & & \\ .1 & 1 & \\ .1 & -.1 & 1 \end{bmatrix}.$$

In this example, $r_{nm}(a, B)$ reaches ± 0.5 and stays there for all B above some cutoff level and a equal to ones. In other words, the correlation among receivers' action is not perfect no matter how bad the communication is. This, however is a nongeneric phenomenon; a small change of K or a means that the values of $r_{nm}(a, B)$ will attach themselves to plus or minus one, for all B large enough.

In short, the poorer the communication, the more correlated actions of the receivers in absolute terms – at least eventually and generically.

7 Extensions and future research

Suppose that different receivers face different communication constraints. For example, let receiver one be more impatient relative to receiver two, $R_1 < R_2$. One can solve the resulting distortion profiles (D_1, D_2) for a given "joint" capacity, $R = R_1 = \min(R_1, R_2)$, exactly as in the analysis above. That is, for any D_1, one can find D_2 that can be achieved if only capacity R for both receivers was available; then one can use the remaining capacity of receiver two, $(R_2 - R_1)$, to further reduce her distortion to some $\tilde{D}_2 < D_2$. By varying D_1, one obtains different values of $\tilde{D}_2$, and thus the whole Utility Possibility Frontier.

The capacity constraint in the model above can be interpreted as a special form of cost function in which the cost of reading R' digits of the message is zero if $R' \leq R$ and the cost is infinity otherwise. Alternatively, one can have a cost function that is smoother; let the cost of reading at rate R be denoted by $c(R)$, e.g. with a constant average cost $c(R)/R = c$. Then, the size of R in equilibrium would be endogenous, a result of receiver's decision, for a given informational policy of the sender.

Non-aligned interests of the sender and the receiver. Suppose that there is only one receiver. In our discussion up to now, the boundary of feasible communication was characterized by function $R(D)$, but it can also be characterized by its inverse, $D(R)$, defined as

$$\begin{aligned} D(R) &= \min_{p_{\hat{x}|x}} E_{X,\hat{X}} d\left(X, \hat{X}\right) \text{ s.t.} \\ & I\left(X, \hat{X}\right) \leq R \end{aligned}$$

So far we assumed that the sender was interested in minimizing some combination of receivers' expected distortions, but was not interested in their actions per se. This is often justified, like in the model of mass media who want to reduce readers' uncertainty. However, in many cases newspapers want to persuade the readers to vote for a party, to change their lifestyle, or to buy a particular product. So suppose sender's incentives are not aligned with the receiver's, e.g. there is one sender, whose loss function is $c(x, \hat{x})$ and one receiver, whose loss function is $d(x, \hat{x})$. Now c does not have to be equal to d.

The problem of the sender (who is assumed to be the designer of encoding-decoding functions) is this

$$\begin{aligned} C(R) &= \min_{p_{\hat{x}|x}} E_{X,\hat{X}} c\left(X, \hat{X}\right) \text{ s.t.} \\ & \begin{cases} I\left(X, \hat{X}\right) \leq R \\ \hat{x} \in \arg\min_y \sum_x p_{x|\hat{x}}(x|\hat{x})\, d(x, y) \end{cases} \end{aligned}$$

The main difference is that there is an extra constraint – the Incentive Compatibility. It is convenient to imagine that the communication does not involve receiving a message $f(x)$, but rather a recommendation what to do, $\hat{x} = g(f(x))$. The term under the argmin is the expected distortion of the receiver if $\hat{x}$ has been recommended but y was selected. Incentive Compatibility then states that the recommendation received from the sender is actually the minimizer of receiver's expected distortion. In general, Lemma 1 would not hold anymore. This is an interesting model and its analysis is left for future research.

8 Appendix A: Proofs

8.1 Proof of Lemma 2

Plugging in $R(D)$ on the right-hand side of equation (2) we obtain

$$\min_{p_{\hat{x}|x}(\hat{x}|x)} \min_{D} I\left(X;\hat{X}\right) \text{ s.t.} \tag{9}$$
$$\sum_n a_n D_n \leq B,\ E_{X,\hat{X}} d_n\left(x_n,\hat{x}_n\right) \leq D_n$$

Note that for any $p_{\hat{x},x}\left(\hat{x},x\right)$ that satisfies

$$\sum_n a_n E_{X,\hat{X}} d_n\left(x_n,\hat{x}_n\right) \leq B \tag{10}$$

there exists a vector D that satisfies the constraints of problem (9). Replacing them with (10) we obtain a definition of $\bar{R}\left(a,B\right)$.

For the second result, note that for any configuration of (a,B) and D such that $\sum_n a_n D_n = B$ we have $R\left(D\right) \geq \bar{R}\left(a,B\right)$. This is because the minimization in $R\left(D\right)$ involves a smaller constraint set. It remains to show that it is impossible that $R\left(D\right) > \max_a \bar{R}\left(a,\sum_n a_n D_n\right)$.

By Lemma 3 the set of distortions that generate strictly lower rate than vector D, $\{y : R\left(D\right) > R\left(y\right)\}$, is convex. By supporting hyperplane theorem, there is a vector $a \neq 0$ such that $\sum_n a_n D_n \leq \sum_n a_n y_n$ for every vector y such that $R\left(D\right) > R\left(y\right)$. Since $R\left(.\right)$ is non-increasing, $a \geq 0$; one can assume that $a \in \Delta^N$. Since $R\left(.\right)$ is continuous, for this a and for every vector x such that $\sum_n a_n D_n = \sum_n a_n x_n$ we have $R\left(D\right) \leq R\left(x\right)$. Take that a and let $B = \sum_n a_n D_n$. The value of $\bar{R}\left(.\right)$ evaluated at this (a,B) must not be lower than $R\left(D\right)$. This implies the result.

8.2 Proof of Lemma 3

The fact that $R\left(D\right)$ is non-decreasing and continuous is standard. The proof of convexity is a direct extension of CT (Lemma 10.4.1, p. 316) to the case of multidimensional domain, and follows from the convexity of mutual information.

Consider a distortion tuple D'. Let $p'_{\hat{x}|x}$ be the marginal distribution that is a solution at D' and let I' be the value of mutual information achieved at

$p'_{\hat{x}|x}$. Similarly, for another D'' define $p''_{\hat{x}|x}$ and I''. For any $\alpha \in (0,1)$ take a convex combination of distortions $D^{\alpha} = \alpha D' + (1-\alpha) D''$ and now consider an minimization problem for this D^{α}. Notice first that a corresponding convex combination of solutions $p'_{\hat{x}|x}$ and $p''_{\hat{x}|x}$, denoted by $p^{\alpha}_{\hat{x}|x}$, is also in the feasibility set in the new problem. Namely, if for all i

$$\begin{aligned}
\int_{X,\hat{X}} p'_{\hat{x}|x}\left(\hat{x}|x\right) p\left(x\right) d_i\left(x_i, \hat{x}_i\right) dx d\hat{x} &\leq D'_i \\
\int_{X,\hat{X}} p''_{\hat{x}|x}\left(\hat{x}|x\right) p\left(x\right) d_i\left(x_i, \hat{x}_i\right) dx d\hat{x} &\leq D''_i
\end{aligned}$$

then for all i

$$\int_{X,\hat{X}} p^{\alpha}_{\hat{x}|x}\left(\hat{x}|x\right) p\left(x\right) d_i\left(x_i, \hat{x}_i\right) dx d\hat{x} \leq D^{\alpha}_i$$

Secondly, mutual information is a convex function of the conditional distribution, for a given marginal distribution. That is, if I^{α} is a mutual information for a distribution $p^{\alpha}_{\hat{x}|x}$ we have $\alpha I' + (1-\alpha) I'' \geq I^{\alpha}$.

Together we can write

$$\begin{aligned}
\alpha R\left(D'\right) + (1-\alpha) R\left(D''\right) &= \alpha I' + (1-\alpha) I'' \\
&\geq I^{\alpha} \\
&\geq R\left(D^{\alpha}\right)
\end{aligned}$$

8.3 Proof of Lemma 4

The entropy of the multivariate Gaussian random variable $X \sim N\left(0, K\right)$ can be expressed as

$$h\left(X\right) = (1/2) \log \left(2\pi e\right)^N \det K$$

Mutual information can be bounded below

$$
\begin{aligned}
I\left(X;\hat{X}\right) &= h(X) - h\left(X|\hat{X}\right) \\
&= h(X) - h\left(X-\hat{X}|\hat{X}\right) \\
&\geq h(X) - h\left(X-\hat{X}\right) \\
&\geq h(N(0,K)) - h(N(0,H)) \\
&= (1/2)\log(2\pi e)^N \det K - (1/2)\log(2\pi e)^N \det H \\
&= (1/2)\log(\det K/\det H)
\end{aligned}
$$

where the first inequality follows from the fact that conditioning reduces entropy, and the second follows from the fact that Gaussian distribution maximizes entropy for a given covariance.

To see that this bound is achievable by some distribution, consider a distribution $X|\hat{x}$ of the following form: let $\hat{X} \sim N(0, K-H)$ and let $Z \sim N(0,H)$ be independent. Let $X = \hat{X} + Z$. Then $X \sim N(0,K)$, as desired. The first inequality above is satisfied with equality because of independence of $Z = X - \hat{X}$ and $\hat{X}$, while the second by the fact that $Z = X - \hat{X}$ is distributed normally.

Hence $I\left(X;\hat{X}\right) \geq (1/2)\log(\det K/\det H)$, with equality as long as $\hat{X} \sim N(0, K-H)$ and $Z \sim N(0,H)$ are independent.

8.4 Proof of Proposition 5

Lemma 6 *If $K > 0$, then the greatest eigenvalue of K_a is simple, $\lambda_1^{K_a} > \lambda_2^{K_a}$, and the normalized eigenvector that belongs to it has all elements positive.*

Proof. Suppose that all correlations are strictly positive, $K > 0$. Then, by Perron–Frobenius theorem, there exists the greatest simple eigenvalue of matrix K, $\lambda_1 > \max_{n\neq 1}\lambda_n = \lambda_2$. Also, its normalized eigenvector has all elements positive. Likewise, there exists the greatest simple eigenvalue of a re-scaled matrix $K_a = A^{1/2}KA^{1/2}$, for any $a \geq 0$. This is true even if some diagonal elements of A are zero. To see this, suppose that only first i of the diagonal elements of A are strictly positive. Then K_a can be written in a block form as

$$K_a = \begin{bmatrix} K_a^i & 0 \\ 0 & 0 \end{bmatrix}$$

where K_a^i is the part of matrix K_a consisting of first $i \times i$ of its elements, all strictly positive. Eigenvalues of K_a are solutions of the characteristic equation

$$\begin{aligned} 0 &= \det\left(\begin{bmatrix} K_a^i & 0 \\ 0 & 0 \end{bmatrix} - \lambda I_N\right) \\ &= \det\left(K_a^i - \lambda I_i\right)\det\left(-\lambda I_{N-i}\right) \end{aligned}$$

So they are equal to i eigenvalues of K_a^i and $N-i$ zeros. By Perron–Frobenius theorem there exists the greatest simple eigenvalue of K_a^i, which, clearly, is also the greatest simple eigenvalue of K_a itself. ■

Now fix a. As B gets closer to its upper bound $\mathrm{tr}\left(K_a\right)$, the water-level λ^* eventually falls into an interval $\lambda_2^{K_a} < \lambda^* < \lambda_1^{K_a}$, which is non-empty, by the Lemma. Then for any two receivers n, m,

$$cov\left(\hat{X}_n^a, \hat{X}_m^a\right) = \left(\lambda_1^{K_a} - \lambda^*\right) q_{n1}^a q_{m1}^a$$

and, noting that all elements of the first eigenvector are positive, $q_{n1}^a > 0$ and $q_{m1}^a > 0$ we have that the correlation between $\hat{X}_n^a$ and $\hat{X}_m^a$ is

$$r\left(\hat{X}_n^a, \hat{X}_m^a\right) = \frac{\left(\lambda_1^{K_a} - \lambda^*\right) q_{n1}^a q_{m1}^a}{\sqrt{\left(\lambda_1^{K_a} - \lambda^*\right)\left(q_{n1}^a\right)^2}\sqrt{\left(\lambda_1^{K_a} - \lambda^*\right)\left(q_{m1}^a\right)^2}} = 1$$

Define $R^o\left(a\right) > 0$ to be a cutoff rate, such that $R \leq R^o\left(a\right)$ implies $r\left(\hat{X}_n^a, \hat{X}_m^a\right) = 1$.

Contrary to the statement of this Proposition, suppose that for any positive R^*, we continue to find a, such that $R^o\left(a\right) < R^*$. In this sequence of vectors a, there is a convergent subsequence. Let a^* be a convergence point, that is, there is a sequence of a such that $\lim_{a \to a^*} R^o\left(a\right) = 0$. Eigenvalues of K_a are a continuous function of A, so must be $R^o\left(\cdot\right)$. Hence $R^o\left(a^*\right) = 0$.

But, by the previous paragraph we have $R^o(a^*) > 0$. This is contradiction.

9 Appendix B: Rate distortion function

As was mentioned earlier, the values $R(D)$ and $R^c(D)$ are equal in a certain limiting sense. The purpose of this section is to state what limit this is (but not prove the equality).

Suppose that the designer of the encoding-decoding functions is allowed R bits to communicate the realization of a random variable X, which has a distribution $p(x)$. The distortion function, which measures the loss when x is represented by some $\hat{x}$, is given too, $d(x, \hat{x})$. Suppose that the optimal encoding-decoding functions achieves some expected distortion D_1. Now, consider a different experiment. Suppose that instead of having one realization of X, the designer has two i.i.d realizations of this random variable, (X_1, X_2). At the same time the designer is allowed to use twice as many bits, $2R$. In other words, there is both more information to transmit and a greater capacity, but still the number of bits per one sample point is constant at R. What is the associated distortion? Clearly, the designer can replicate the previous design independently for two realizations of X and will end up paying the distortion cost twice, or the same distortion per sample point, D_1. However, this, in general, will not be optimal. The optimal design will encode and decode the two sample points *together*. That way, the distortion per sample point will be lower, $D_2 < D_1$. This is true even if the sample points are statistically independent of each other.[13] As the sample size (number of these experiments) goes to infinity, this sequence will converge down to some D^*. It turns out that this limit can be expressed using only the primitives: the distribution p and the distortion function d. Namely, it is exactly the (inverse of) information rate distortion function defined in (1).

Formally speaking, let $X_1, ..., X_T$ be an i.i.d. sample of size T from distribution $p(x)$, $x \in \mathcal{X}$. The encoding function for the whole sample is denoted

[13] CT, p.301: "One of the most intriguing aspects of this theory is that joint descriptions are more efficient than individual descriptions. It is simpler do describe an elephant and a chicken with one dscription than to describe each alone"

$f_T : \mathcal{X}^T \to \left\{1, ..., 2^{TR}\right\}$ and a decoding function is $g_T : \left\{1, ..., 2^{TR}\right\} \to \mathcal{X}^T$. Since we allow X, $\mathcal{X}$, x etc. to be multidimensional, g_T may also be multidimensional, and for a dimension n we may write $g_{nT} : \left\{1, ..., 2^{TR}\right\} \to (\mathcal{X}_n)^T$. Given a distortion function, $d_n(x_{tn}, \hat{x}_{tn})$ for dimension n and a single sample point $x_{tn} \in \mathcal{X}_n$, let the dimension n distortion between the entire samples $x_n^T = (x_{1n}, ..., x_{Tn})$ and $\hat{x}_n^T = (\hat{x}_{1n}, ..., \hat{x}_{Tn})$ be

$$d_n^{\mathrm{avg},T}\left(x_n^T, \hat{x}_n^T\right) = (1/T)\sum_{t=1}^{T} d_n(x_{tn}, \hat{x}_{tn})$$

The expected distortion associated with (f_T, g_T) at dimension n is the expected value of the above, $Ed_n^{\mathrm{avg},T}\left(x_n^T, g_{nT}\left(f_T\left(x_n^T\right)\right)\right)$.

The rate-distortion (R, D) is achievable if, for a given average rate R, there is a sequence (f_T, g_T), such that

$$\lim_{T\to\infty} Ed_n^{\mathrm{avg},T}\left(x_n^T, g_{nT}\left(f_T\left(x_n^T\right)\right)\right) \leq D_n$$

The rate distortion function $R^c(D)$ is the infimum of rates R such that (R, D) is achievable.

The following theorem is the main result: $R^c(D) = R(D)$, where $R(D)$ is specified in Definition (1).

10 Literature

Banerjee, A., (1992), A simple model of herd behavior, *Quarterly Journal of Economics*, Vol. 107, pp. 797-817.

Calvo, Guillermo A. and Enrique G. Mendoza, (2000), Capital-Markets Crises and Economic Collapse in Emerging Markets: An Informational Frictions Approach, *The American Economic Review*, Vol. 90, No. 2, Papers and Proceedings of the One Hundred Twelfth Annual Meeting of the American Economic Association, pp. 59-64.

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 Wiley & Sons Inc, Hoboken, New Jersey.

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

Christiano, Lawrence J. and Terry J. Fitzgerald, (1998), The business cycle: It's still a puzzle, *Economic Perspectives*, Vol.22, pp. 56-83.

Gill, David, Daniel Sgroi, (2011), The Optimal Choice of Pre-Launch Reviewer, *Journal of Economic Theory*, forthcoming.

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

Irmen, Andreas, and Jacques-Francois Thisse, (1998), Competition in Multi-characteristics Spaces: Hotelling Was Almost Right, *Journal of Economic Theory*, 78, 76 102

Johnson, Justin P. and David P. Myatt, (2006), On the Simple Economics of Advertising, marketing, and Product Design, *American Economic Review*, Vol. 96 No. 3.

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

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

Lancaster, Kelvin J., (1966), A New Approach to Consumer Theory, *Journal of Political Economy*, 74(2), pp. 132-57.

McQuail, Denis, (2010), *McQuail's Mass Communication Theory*, Sage Publications Ltd; Sixth Edition.

Prat, Andrea, and David Strömberg, (2011), The Political Economy of Mass Media, mimeo.

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.

Shannon, Claude E., (1948), A Mathematical Theory of Communication, *Bell System Technical Journal*, vol. 27, pp. 379–423, 623-656.

Veldkamp, Laura L., (2006a), Media Frenzies in Markets for Financial Information, *American Economic Review*, Vol. 96, No. 3, pp. 577-601.

Veldkamp, Laura L., (2006b), Information Markets and the Comovement of Asset Prices, *Review of Economic Studies*, Vol. 73, No. 3, pp. 823-845