# Bayesian Persuasion with signal uncertainty

Gabriel Martnez*

September 8, 2018

**Abstract**

People regularly receive recommendations about how to behave based on some piece of evidence. Whether it is a news report, an article in a magazine or social media, a marketing campaign, a policy evaluation, etc, they usually have data or experiments that support their statements. However, for many decision makers, it is often hard or impossible to fully understand the details of how the study was designed, the power of its tests, its external validity, etc. In general, it is unclear the message informativeness, thus how to respond to its recommendation. I model this uncertainty about the signal that a sender uses to persuade a receiver. I consider both exogenous uncertainty and asymmetric information, and show that compared to the benchmark of Kamenika and Gentskow (2011), this uncertainty may benefit some types of senders and harm others. The amount of information provided increases at moderate levels of uncertainty, but in contrast with the cheap talk literature, large levels of skepticism can lead to completely uninformative signals. Finally, I complement Bergemann and Morris' (2017) framework of information design to connect my setup with existing literature on limited commitment (Frechette et al, 2017 and Min, 2017), and persuasion with multiple senders (Kamenika and Gentzkow, 2018 a,b and Li and Morgan, 2018).

JEL Classification: D82, D83, D91
Keywords: Bayesian persuasion, Skepticism, Bias Uncertainty, Information Design, Multiple Senders, Communication

## 1 Motivation and Related Literature:

This paper fits within the large literature of bayesian persuasion and information design that has been very active during the last decade. As surveyed by Bergermann and Morris (2017) the main problem concerns with the provision of information to influence the decision makers. My contribution will be to consider a two dimensional state of the world: one dimension is payoff relevant and the other one is informationally relevant. The first one is standard in the literature while the second is, to my knowledge, the novel dimension in the bayesian persuasion literature. The second dimension affects the preferences of the sender (its bias) thus the choice of the information

* University of Wisconsin-Madison, Department of Economics; E-mail: gamartinez@wisc.edu.

structure provided. If the sender, or information designer could commit to provide information on both dimensions, this reduces to KG.

In contrast, if the sender cannot commit to provide information on the first dimension, the model will fall under the umbrella of cheap talk. In particular Li and Madarasz (2008) explore the case where there is also no commitment on the informationally relevant dimension. Surprisingly they find that disclosing the bias of the sender can reduce the amount of information communicated in equilibrium. Intuitively, when disclosure is not required and there are two senders with opposed biases. Upon receiving a message, the receiver responds to it as if coming from the average sender whose expected bias will be smaller in absolute value than any of the two senders, thus enhancing communication for reasons similar to Crawford and Sobel (1982). Finally, If they can commit on the second dimension but not on the first one, this would be a model of cheap talk with a sender with state dependent preferences.

Within the bayesian persuasion literature this project is close to Frechette, Lizzeri and Parego (2017) and Min (2017) both consider a sender with partial commitment. In their frameworks, there is an exogenous probability that the sender is actually a cheap-talker[1]. Given the misalignment of preferences between senders and receivers, this possibility makes the receiver skeptical of the information provided. In fact if such probability is large enough, no valuable information is provided in equilibrium. Effectively they also have the double inference problem in their papers, but since a biased cheap-talker will always choose a degenerate distribution over messages, this double inference problem is rather simple.

This model allows for two interpretations, the first one is to think of a single information designer where the state of the world is complex (two dimensional) and the designer can only commit to provide information about one of the dimensions. The second one is to think of a decision maker that reads only the headlines of research results and has to both asses her beliefs about, say the efficacy of a public policy, and the informativeness of the methods used to arrive to this recommendation. Under the second interpretation we can take a step back and think of the problem for platform designers, say news-media channels, or social media platforms like Facebook and Twitter that can affect the frequency with which information consumers get recommendations from each type of information designers. Effectively affecting the distribution over types of senders. Moreover, we can think of disclosure rules, that effectively impose restrictions in the message space so that the posterior distribution over types of senders is degenerate.

Notice that it is not obvious the effect of disclosure requirements. On the one side if disclosure is not required, the double inference problem can make the agents more skeptical about recom-

[1]In their framework, there is an exogenous probability that after committing to an experiment, the the receiver observes, they can secretly revise it.

mendations, possibly obfuscating persuasion and thus incentives to provide information. On the other hand, information providers might need to choose more informative experiments in order to counteract this effect and persuade the receivers.

Finally, given the interpretation of multiple senders, the project relates to the competitive persuasion literature (GK, 2016 and 2017; Li & Norman, 2017, for example) with the crucial difference that the receiver dos not observe the signal realizations of all senders, it just observes one of such signals. Under this framework the methods used in Bergemann and Morris (2017) via obedience constraints are feasible, a version of the revelation principle in mechanism design. Moreover, suffices to restrict attention to a choice of independent experiments chosen by each sender. Therefore, It is not obvious that the result of GK (2016) will hold here. This is that the collusive equilibrium (basically equivalent to assuming commitment in the informationally relevant dimension), is no more informative than the competitive equilibrium.

## 2 Model

Lets consider a single sender (female) and a receiver (male). The receiver has preferences, $u(a, \theta)$, over actions and the state of the world. Let's assume that the action space, $A$ is compact, the utility continuous and the state space $\Theta$ finite. The sender chooses a signal that is a function from the state of the world to a distribution over messages, $\pi_t : \Theta \to \Delta(M)$, the subindex $t$ is explained next.[2]. The receiver is uncertain about this mapping from the state of the world to messages, so we endow him with beliefs about it. This is, he believes that the signal is one among a family of possible signals indexed by $t \in \mathcal{T}$, and he has a prior on the join distribution over $\Theta$ and $\mathcal{T}$, denoted $\mu_0 \in \Delta(\Theta \times \mathcal{T})$. Exogenous uncertainty and asymmetric information can be easily captured by thinking of $\mathcal{T}$ as a type space and partitioning the space into $\mathcal{T}_n$ and $\mathcal{T}_s$ with $\mathcal{T}_s \neq \emptyset$ and $|\mathcal{T}| \geq 2$. The first subset denotes noise types for which $\pi_t$ is exogenously pre-specified. While the second subset denotes proper (strategic) types that choose $\pi_t$ to maximize their utility $v(a, \omega, t)$ that, in general, may depend on the realization of $t$. The sender has no commitment power to reveal her type[3].

To illustrate this setting, lets consider some examples:

- Noisy Persuasion: a single strategic sender chooses a signal whose recommendation is delivered to the receiver with probability $1 - \epsilon$ and with $\epsilon$ probability the recommendation is swapped to some other action chosen at random; i.e. $|\mathcal{T}_s| = |\mathcal{T}_n| = 1$. The noise type $t_n$, is associated with the signal $\pi_{t_n}$ which is a pre-defined permutation of the signal chosen by the strategic type, $\pi_{t_s}$. Le Treust and Tomala (2018) and Tsakas *et al.* (2018) study various settings that can be directly mapped into this language and have equivalent analysis.

[2] The symbol $\Delta(M)$ is used to denote the set of all probability distributions with support in $M$.
[3] If she could condition the signal on its type, this will be equivalent to KG11.

- Privately informed sender: all the types of senders are strategic, but their utility is constant over their type. However, types are correlated with the state, i.e. $\Delta(\Theta \times \mathcal{T})$ is not a product distribution. Information is privately observed, and senders cannot commit to reveal it. Perez-Richet (2014), Hedlung (2016), Piermont (2016) and Kosenko (2018) consider different settings with these properties.

- Sender with a hidden agenda: Strategic types of senders do not have private information about the state of the world, i.e. $\Theta$ and $\mathcal{T}$ are independent random variables, but $v(a, \theta, t)$ is not constant over $t$. For example, the sender might have perfectly aligned preferences with the receiver, might be biased or might have perfectly misaligned preferences. Likewise, the receiver might me unsure about the action towards which the sender is biased. To my knowledge this is the first paper to allow for this kind of asymmetric information.

A key distinction with the standard model is that the receiver does not observe the experiment, $\pi_{\mathcal{T}}$, directly, they only observe the recommendation[4]. In KG11, this distinction is irrelevant. Through common knowledge of the environment and of rationality, the sender can infer the exact signal that produced a recommendation and act accordingly to a recommendation. This would also be the case here if $\mathcal{T}$ was a singleton. Knowledge of the preferences of the sender suffices to infer the optimal signal that would be chosen. However, if the receiver believes that $\mathcal{T}$ is not a singleton, there is a difference between simply observing the recommendation and understanding the experiment that generated it. In this setting we have a decision maker who believes that a recommendation is not simply cheap talk, but is uncertain the exact properties of the signal that produced it, this special kind of uncertainty is what i call skepticism. In general, they have a double inference problem: to learn about the state of the world, and about the quality of the signal.

**Example 1** *To illustrate this setup, lets consider the simple example of a binary state, with an identical binary action space, say $\Omega = \{0, 1\}$. The receiver wants to match her action with the state of the world, so $a = 0$ iff $\mu_1 < 0.5$. She has a prior belief that the sate is 1 of 0.2, say $\mu_0 = 0.2$. Likewise she believes that with equal probability a recommendation may come from a biased sender that wants action 1 to be taken regardless or by noise - this is an experiment that yields either recommendation with equal probability. Perhaps not surprisingly, if the biased sender type were to choose an experiment as prescribed by KG11, the skeptic receiver would not be convinced to choose a different action than the default: $a = 0$.*

To see this, let $\mu_1$ denote the posterior belief that the state of the world is 1. KG11 prescribe in this situation for the biased sender, $t = b$, to choose $\pi_b(1|1) = 1, \pi_b(0|1) = \mu_0/(1 - \mu_0) = 1/4$ where $\pi_t(a|\omega)$ is the probability that action $a$ is recommended when the state of the world is $\omega$ by type $t \in \{b, n\}$. Then

[4] In many settings it is easier for the decision maker to make a belief about the intentions and reliability of the source, than to fully understand the methodology used derive a recommendation

$$\begin{aligned}\mu_1 &= \frac{(0.5\pi_b(1|1) + 0.5\pi_n(1|1))\,\mu_0}{(0.5\pi_b(1|1) + 0.5\pi_n(1|1))\,\mu_0 + (0.5\pi_b(1|0) + 0.5\pi_n(1|0)) * (1-\mu_0)} \\ &= \frac{0.15}{0.15+0.3} = \frac{1}{3} < 0.5.\end{aligned}$$

However, perhaps more surprisingly, $\mu_1 < 0.5$ for all possible choices of $\pi_b$. Even if the biased sender chooses a perfectly informative signal, $\mu_1 = 3/7 < 0.5$. This skepticism of the receiver prevents a biased sender to effectively communicate any message that will influence the action of the receiver. Furthermore, we can dispense of the non-strategic noise type, $t \in \mathcal{T}$, and replace it by a biased sender that wants action 0 to be chosen regardless. In equilibrium, this sender is happy with the default action, so choosing a noisy signal is weakly dominant for him, since it ensure that action 0 will always be chosen, and the sender biased towards action 1 is indifferent among all actions, so it has no profitable deviation. This is to say that if the decision maker listens with high enough probability, uninformative messages, or alternatively only an information provider that is happy with the status quo, no information structure can convince her to change her action!

Indeed, in general, lets denote by $\tau_0$ the probability that the message is noise, then the best response for a biased sender is:

$$\pi_b(1|1)^* = 1, \text{ and } \pi_b(1|0)^* = \frac{\mu_0}{1-\mu_0} - \left(1 - \frac{\mu_0}{1-\mu_0}\right)\frac{\tau_0}{2(1-\tau_0)} \quad \text{iff } \tau_0 \leq 2\mu_0;$$

otherwise, any signal implements beliefs strictly lower than 1/2, so the sender is indifferent among them all.

This is, the receiver must believe that the message is noisy with low enough probability. In fact notice that $\pi_b(1|0)$ is the probability of sending the wrong recommendation when the state of nature is zero. A fully informative signal would have $\pi_b(1|1) = 1$ as in the best response and $\pi_b(1|0)$ would be equal to zero. This example illustrates that low levels of skepticism are good: they incentivize a biased sender to send a more informative signal in order to influence the action ($\frac{\partial \pi_b(1|0)}{\partial \tau_0} < 0$ whenever $\tau_0 \leq 2\mu_0$)[5]. (Insert picture here).

# 3 General framework

In this framework, we have restricted to a message space equal to the action space. This assumption, restricts attention to situations where the different types of senders, share the support of their strategies. Thus, effectively ruling out situations where every message uniquely identifies the type of sender providing information. This assumption is in line with a receiver that understands what a study is recommending to do, but not the incentives or methodology of the people providing the

[5] This result is similar to the limited commitment setup in Frechette et al (2017) and Min (2017). In their work, one type is a cheap talker and is equivalent to a type that always recommends an action equal to 1. However, their setup is sequential with a hidden move, which deals aways with a hurdle regarding the message space, that iIwill expand in the next section. They also find an increase in informativeness and no information after some threshold.

recommendation. Worth noting here, that the receiver in this setting responds to the average signal, so contrary to Bergeman and Morris (2015) comment on Kamenica and Gentzkow (2017), allowing for mixed strategies has no bite here. Similarly, so long as only one message is being observed, the ability to correlate signals with other type's signals is immaterial. Let's think of the type space $\mathcal{T}$ as a finite space with $\mathcal{T}^* \subseteq \mathcal{T}$ strategic types of senders. Bergeman and Morris (2016) showed that with a single sender, the problem of bayesian persuasion is equivalent to maximizing the sender's utility over the set of bayes-nash outcomes of the underlying game of uncertainty complemented with any information that is congruent with the receiver's prior (the bayesian plausibility constraint). They show how this is equivalent to the set of rationalizable outcomes. I show that existence of multiple types of senders impose further restrictions on the information structure that a sender can supplement the game. The information provided, needs now to also be congruent with the information provided by other senders. The setup is now equivalent to the set of Delta-rationalizable outcomes. So for any strategic senders, their strategy maximizes over this restricted set, which restrictions depend on the action of other senders.

The shape of this restrictions depend on 4 main characteristics:

1. How many messages are seen by the receiver?
2. Of those, what probability the receiver assigns to it coming from each possible sender?
3. When multiple messages are received, can they be correlated?
4. Can senders choose mixed strategies?

KG (2017) consider a setting where the number of messages seen coincides with the cardinality of $\mathcal{T}$ and all senders are identified (this is equivalent, to assume that the support of messages they use, is pair-wise disjoint). Finally, they assume messages can be perfectly correlated, and mixed strategies are ruled out. The first two assumptions imply that the receiver responds to the unique (from point 2) join signal. The perfect correlation assumption coupled with the focus on pure strategies only, imply that weakly more informative join messages are always feasible, and no less informative ones are (as illustrated by Bergemann and Morris, 2015). Hence, the result that competitive equilibrium cannot be less informative than a collusive one.

# 4 characterization (existance)

# 5 Competitive interpretation

# 6 Extensions

## References

Austen-Smith, D. (1990). Information transmission in debate. American Journal of political science, 124-152.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. Econometrica: Journal of the Econometric Society, 1431-1451.

Battaglini, M. (2002). Multiple referrals and multidimensional cheap talk. Econometrica, 70(4), 1379-1401.

Bergemann, D., & Morris, S. (2017). Information design: A unified perspective.

Benabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. The Quarterly Journal of Economics, 117(3), 871-915.

Dimitrakas, V., & Sarafidis, Y. (2005). Advice from an expert with unknown motives.

Gentzkow, M., & Kamenica, E. (2016). Competition in persuasion. The Review of Economic Studies, 84(1), 300-322.

Gentzkow, M., & Kamenica, E. (2017). Bayesian persuasion with multiple senders and rich signal spaces. Games and Economic Behavior, 104, 411-429.

Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. American Economic Review, 101(6), 2590-2615.

Krishna, V., & Morgan, J. (2001). A model of expertise. The Quarterly Journal of Economics, 116(2), 747-775.

Li, M., & Madarasz, K. (2008). When mandatory disclosure hurts: Expert advice and conflicting interests. Journal of Economic Theory, 139(1), 47-74.

Li, F., & Norman, P. (2017). On Bayesian persuasion with multiple senders.

Min, D. (2001). Bayesian Persuasion under Partial Commitment. Accountability in research, 8, 245-260.

Morgan, J., & Stocken, P. C. (2003). An analysis of stock recommendations. RAND Journal of economics, 183-203.

Morris, S. (2001). Political correctness. Journal of political Economy, 109(2), 231-265.

Perego, G., F. A. & L. J. (2017). Rules and Commitment in Communication.

Portugal, H. P. R. (2012). Cheap Talk With Global Unknown Misalignment.

Stein, J. C. (1989). Cheap talk and the Fed: A theory of imprecise policy announcements. The American Economic Review, 32-42.