



Impact of Private Observation in the Bayesian Persuasion Game

Rony Bou Roupheal^(✉) and Maël Le Treust

ETIS UMR 8051, CY Université, ENSEA, CNRS, 6, avenue du Ponceau,
95014 Cergy-Pontoise Cedex, France
{rony.bou-roupheal,mael.le-treust}@ensea.fr

Abstract. In the Bayesian persuasion setting, the sender aims at persuading the decision maker, so called the decoder, to choose a certain action among a set of possible actions. This paper considers two Bayesian persuasion games: one that involves the observation of a private signal by the decoder in addition to the signal transmitted by the encoder, and another version where no private signal is accessible by the decoder. Our goal is to examine the impact of this private signal on the encoder's optimal utility. In order to do so, we investigate an example involving a binary state, a binary private signal and a binary receiver's actions set. We identify the optimal splitting of the decoder's beliefs satisfying the information constraint imposed by the restricted communication channel, and we compute the encoder's optimal utility value, with and without private signal. Varying the parameters such as the prior belief, the precision of the private signal and the channel capacity, we aim at determining which of the two settings is more favorable to the encoder.

Keywords: Bayesian persuasion · Strategic communication · Side information

1 Introduction

In [5], Kamenica-Gentzkow investigate a persuasion game in which the sender observes the realization of a state variable and commits to some signalling mechanism, then the receiver chooses the best-reply action corresponding to its posterior belief. Communication in persuasion games may be constrained by a limited channel's capacity and messages distorted by some source of noise, as in [9]. Moreover, the receiver may privately observe a signal correlated to the state, as in the source coding problem of Slepian-Wolf and Wyner-Ziv, in [12] and [13]. In such settings, the persuasion problem is hard to solve even for simple models. Tools from information theory, involving entropy and mutual information,

M. Le Treust—Gratefully acknowledges financial support from INS2I CNRS, DIM-RFSI, SRV ENSEA, UFR-ST UCP, The Paris Seine Initiative and IEA Cergy-Pontoise. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

provided a solution for certain scenarios of repeated persuasion problems. The optimal solution to the noisy persuasion problem relies on a specific concavification involving an auxiliary utility function for the sender that accounts for the private observation of the receiver as in [8,9] and [10].

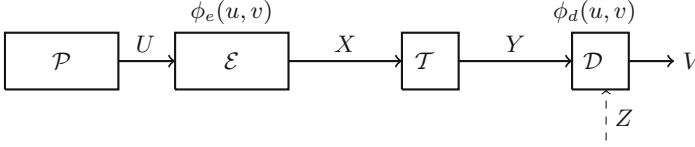


Fig. 1. Bayesian persuasion game with noisy channel $\mathcal{T}(y|x)$, with or without decoder's side information Z . The utility functions of the encoder \mathcal{E} and decoder \mathcal{D} are denoted by $\phi_e(u, v)$ and $\phi_d(u, v)$.

1.1 State of the Art

Channel coding and communication problems originally introduced in [11] have been studied in several settings, particularly with the side information setting as in [1] where a hierarchical communication game is considered to treat information disclosure problems originated in economics and involving different objectives for the encoder and the decoder. In [2], Alonso-Cámara provide necessary and sufficient conditions under which a sender benefits from persuading decoders with distinct prior beliefs. The computational aspects of the persuasion game are considered in [4], where the impact of the channel's capacity on the optimal utility is investigated. Persuasion of a privately informed receiver was also investigated in [6], in which the optimal persuasion mechanisms are characterized. In [7], Laclau-Renou depicted the constraints imposed on the sender when multiple receivers have multiple beliefs.

1.2 Contributions

In this paper, we consider a persuasion game with binary source/state and binary decoder's actions and we investigate the effect of the decoder's side observation on the encoder's optimal utility. We compute numerically the two values of the persuasion problems, with and without decoder's side information, depending on two keys parameters, 1) the channel capacity and 2) the precision of the decoder's side information. Depending on these two parameters, the decoder's side information may increase or decrease the encoder's utility.

The paper is organized as follows. The notations are defined in Sect. 2. In Sect. 3, we formulate the two concavification problems. In Sect. 4, we introduce the example of a binary source and state, and we formulate the optimal solutions for the case with no private observation in Sect. 4.1, and for the case where private observation is available at the decoder Sect. 4.2. In Sect. 5, we provide the results of our numerical simulations.

2 Notations

This paper considers a communication model that is illustrated in Fig. 1. Let \mathcal{E} denote the encoder and \mathcal{D} denote the decoder. Notations U , Z , X , Y , and V denote the random variables of information source $u \in \mathcal{U}$, side information $z \in \mathcal{Z}$, channel inputs $x \in \mathcal{X}$, channel outputs $y \in \mathcal{Y}$, and decoder's actions $v \in \mathcal{V}$ respectively. Calligraphic fonts \mathcal{U} , \mathcal{Z} , \mathcal{X} , \mathcal{Y} , and \mathcal{V} , denote the alphabets and lowercase letters u , z , x , y , and v denote the signal realizations. Notation \mathcal{P}_U stands for the probability distribution of the state U of the game. The private observation Z of the receiver is correlated to U according to the conditional probability distribution $\mathcal{P}_{Z|U}$. We will denote the beliefs of the decoder by $p \in \Delta(\mathcal{U})$ whereas $p(u)$ belongs to $[0, 1]$ for each $u \in \mathcal{U}$. The i.i.d. memoryless channel distribution will be denoted by $\mathcal{T}_{Y|X}$. We denote by $\Delta(\mathcal{X})$ the probability simplex, i.e. the set of probability distributions over \mathcal{X} . We denote by \mathcal{Q}_X the probability distribution over \mathcal{X} , i.e. the posterior beliefs of the decoder. The joint probability distribution $\mathcal{Q}_{XV} \in \Delta(\mathcal{X} \times \mathcal{V})$ decomposes as follows, $\mathcal{Q}_{XV} = \mathcal{Q}_X \times \mathcal{Q}_{V|X}$. The channel's capacity will be denoted by C . Notations $H(U)$, $H(U|Z)$ and $I(X; Y)$ refer to Shannon's entropy, conditional entropy and mutual information respectively [3, pp. 12], and are given below.

$$H(U) = \sum_{u \in \mathcal{U}} p(u) \log_2 \frac{1}{p(u)}, \quad H(U|Z) = \sum_{z \in \mathcal{Z}} \sum_{u \in \mathcal{U}} p(z, u) \log_2 \frac{1}{p(u)}, \quad (1)$$

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}, \quad C = \max_{\mathcal{P}(x)} I(X; Y). \quad (2)$$

3 Concavification Problems

Given a capacity value $C \geq 0$, we consider the two concavification problems below, stated in [10, Def. III.1] and [9, Def. 2.4].

$$\Gamma_0 = \sup_{(\lambda_w \cdot p_w)_{w \in \mathcal{W}}} \left\{ \sum_{w \in \mathcal{W}} \lambda_w \cdot \Phi_e(p_w) \text{ s.t. } \sum_{w \in \mathcal{W}} \lambda_w \cdot p_w = \mathcal{P}_U, \right. \\ \left. \sum_w \lambda_w \cdot H(p_w) \geq H(U) - C, |\mathcal{W}| = \min(|\mathcal{U}| + 1, |\mathcal{V}|) \right\}, \quad (3)$$

$$\Gamma = \sup_{(\lambda_w \cdot p_w)_{w \in \mathcal{W}}} \left\{ \sum_{w \in \mathcal{W}} \lambda_w \cdot \Psi_e(p_w) \text{ s.t. } \sum_{w \in \mathcal{W}} \lambda_w \cdot p_w = \mathcal{P}_U, \right. \\ \left. \sum_w \lambda_w \cdot h(p_w) \geq H(U|Z) - C, |\mathcal{W}| = \min(|\mathcal{U}| + 1, |\mathcal{V}|^{|Z|}) \right\}, \quad (4)$$

where

$$\Phi_e(p) = \mathbb{E}_p \left[\phi_e(U, v^*(p)) \right], \quad H(p) = H(U), \quad (5)$$

$$v^*(p) = \arg \min_{v \in \arg \max \mathbb{E}_p[\phi_d(U, v)]} \mathbb{E}_p[\phi_e(U, v)], \tag{6}$$

and

$$\forall z \in \mathcal{Z}, \quad q_z \in \Delta(\mathcal{U}), \quad q_z(u) = \frac{p(u) \cdot \mathcal{P}(z|u)}{\sum_{u'} p(u') \cdot \mathcal{P}(z|u')}, \quad \forall u \in \mathcal{U}, \tag{7}$$

$$\Psi_e(p) = \sum_{u, z} p(u) \cdot \mathcal{P}(z|u) \cdot \Phi_e(q_z), \tag{8}$$

$$h(p) = \sum_{u, z} p(u) \cdot \mathcal{P}(z|u) \cdot \log_2 \frac{\sum_{u'} p(u') \cdot \mathcal{P}(z|u')}{p(u) \cdot \mathcal{P}(z|u)}. \tag{9}$$

The notation $v^*(p) \in V$ stands for the decoder’s best reply action with respect to its posterior belief $p \in \Delta(\mathcal{U})$. If several actions maximize the utility of the decoder, we assume that it chooses the one that minimizes the encoder’s utility. Thus, the encoder’s expected utility $\Phi_e(p)$ is evaluated with respect to the decoder’s belief $p \in \Delta(\mathcal{U})$. In the presence of side information $z \in \mathcal{Z}$, the decoder’s belief is denoted by $q_z \in \Delta(\mathcal{U})$. As a consequence, the encoder’s expected utility $\Psi_e(p)$ is a convex combination between the utilities $\Phi_e(q_z)$ evaluated at different possible beliefs $(q_z)_{z \in \mathcal{Z}}$. The supremum in (4) and (3) are taken over the set of splittings $(\lambda_w, p_w)_{w \in \mathcal{W}}$ of the prior probability distribution $\mathcal{P}_U \in \Delta(\mathcal{U})$, that satisfy the cardinality bound, either $|\mathcal{W}| = \min(|\mathcal{U}| + 1, |\mathcal{V}|)$ or $|\mathcal{W}| = \min(|\mathcal{U}| + 1, |\mathcal{V}|^{|\mathcal{Z}|})$.

Formulas (3) and (4) are solutions to the persuasion game with noisy channel. The value Γ corresponds to the persuasion problem in which the decoder has a private observation Z correlated with the state U according to the conditional probability distribution $P_{Z|U}$, whereas the value Γ_0 corresponds to the persuasion problem in which the decoder has no access to a side information, or equivalently, has a private observation Z that is independent from the state U . When removing the entropy-based constraints in Γ_0 , the concavification problem boils down to the optimal solution provided by Kamenica-Gentzkow in [5].

4 Example with Binary Source and State

In this section, we will illustrate a particular scenario of a strategic communication involving binary source and state and decoder’s action. Let $\mathcal{U} = \{u_0, u_1\}$ the state space, $\mathcal{V} = \{v_0, v_1\}$ the action space, and $p_0 = P(U = u_1) \in [0, 1]$ the decoder’s prior belief. We consider a binary symmetric noisy channel where $\mathcal{X} = \{x_0, x_1\}$ denotes the set of channel inputs, $\mathcal{Y} = \{y_0, y_1\}$ denotes the set of channel outputs. The channel’s capacity for noise level $\epsilon \in [0, \frac{1}{2}]$ is given by $C = 1 - H_b(\epsilon)$ where $H_b(p)$ denotes the binary entropy. Utility functions of both encoder and decoder are given in Tables 1 and 2.

Table 1. Encoder’s utility

	v_0	v_1
u_0	0	1
u_1	0	1

Table 2. Decoder’s utility

	v_0	v_1
u_0	9	0
u_1	4	10

As shown in the decoder’s expected utility graph Fig. 2a, the red lines represent the decoder’s best reply action. Therefore, the action of the decoder will only change from v_0 to v_1 depending on the utility threshold γ . In this example, we consider the prior $p_0 = 0.4$ and the utility threshold $\gamma = 0.6$.

4.1 Persuasion Without Side Information (Equation for Γ_0)

The optimal number of posterior beliefs when no side information is available at the decoder is two [9, lemma 6.1]. These posterior beliefs of the decoder need to satisfy the splitting condition and information constraint

$$\lambda q_1 + (1 - \lambda)q_2 = p_0 \iff \lambda = \frac{p_0 - q_2}{q_1 - q_2} \iff 1 - \lambda = \frac{q_1 - p_0}{q_1 - q_2}, \tag{10}$$

$$\lambda H_b(q_1) + (1 - \lambda)H_b(q_2) \geq H_b(p_0) - C. \tag{11}$$

Assuming the information constraint is binding at the optimal, we get

$$\lambda H_b(q_1) + (1 - \lambda)H_b(q_2) = H_b(p_0) - C \tag{12}$$

$$\iff H_b(q_1) = \frac{p_0 H_b(q_2) - q_2 (H_b(p_0) - C)}{(p_0 - q_2)} + q_1 \frac{(-H_b(q_2) + H_b(p_0) - C)}{(p_0 - q_2)} \tag{13}$$

The encoder’s expected utility function Φ_e depicted in Fig. 2b is defined over $[0, 1]$ by $\Phi_e(q) = \mathbb{1}_{q \in [\gamma, 1]}$. For each $q_2 \in [p_0, 1]$, we denote by $q_1(q_2)$ the unique solution of (13) for a given pair (p_0, C) . We assume that the decoder’s threshold $\gamma > p_0$, hence at the optimum $q_2 = \gamma$, thus

$$\Gamma_0 = \sup_{q_2 \in [0, 1]} \left(\lambda \Phi_e(q_1(q_2)) + (1 - \lambda) \Phi_e(q_2) \right) = \frac{q_1(\gamma) - p_0}{q_1(\gamma) - \gamma}. \tag{14}$$

Figure 2b shows an unrestricted communication without decoder’s side information. The green dotted line is the concavification of the encoder’s expected utility function represented in the red lines. The optimal utility value corresponds to the evaluation of this concavification at the prior belief p_0 .

4.2 Persuasion with Side Information (Equation for Γ)

When side information $\mathcal{Z} = \{z_0, z_1\}$ is observed by the decoder, then [9, Lemma 6.3] ensures that the optimal number of posterior beliefs is three. The posterior distributions (q_1, q_2, q_3) from observing the message delivered by the encoder, must satisfy the information constraint given by

$$\lambda_1 \cdot h(q_1) + \lambda_2 \cdot h(q_2) + \lambda_3 \cdot h(q_3) \geq H(U|Z) - \max_{\mathcal{P}(x)} I(X; Y) \tag{15}$$

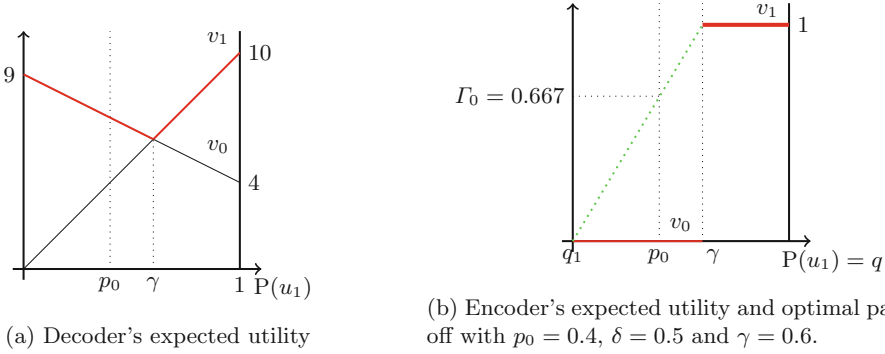


Fig. 2. Encoder and Decoder's Expected Utilities

Thus $(\lambda_1, \lambda_2, \lambda_3)$ can be computed from the above information constraint, the splitting lemma $\lambda_1 q_1 + \lambda_2 q_2 + \lambda_3 q_3 = p_0$ and the fact that $\lambda_1 + \lambda_2 + \lambda_3 = 1$. We assume that the information constraint is binding. From [10, Eq. (57)-(59)], we have

$$\lambda_1 = \frac{IC \cdot (q_2 - q_3) + h(q_2) \cdot (q_3 - p_0) + h(q_3) \cdot (p_0 - q_2)}{h(q_1) \cdot (q_2 - q_3) + h(q_2) \cdot (q_3 - q_1) + h(q_3) \cdot (q_1 - q_2)}, \quad (16)$$

$$\lambda_2 = \frac{IC \cdot (q_3 - q_1) + h(q_3) \cdot (q_1 - p_0) + h(q_1) \cdot (p_0 - q_3)}{h(q_1) \cdot (q_2 - q_3) + h(q_2) \cdot (q_3 - q_1) + h(q_3) \cdot (q_1 - q_2)}, \quad (17)$$

$$\lambda_3 = \frac{IC \cdot (q_1 - q_2) + h(q_1) \cdot (q_2 - p_0) + h(q_2) \cdot (p_0 - q_1)}{h(q_1) \cdot (q_2 - q_3) + h(q_2) \cdot (q_3 - q_1) + h(q_3) \cdot (q_1 - q_2)}. \quad (18)$$

Given a *interim* belief parameter $q \in [0, 1]$, the decoder's side information might be z_0 or z_1 , thus inducing the two following posterior beliefs

$$p_1(q) = \frac{q \cdot \delta}{(1 - q) \cdot (1 - \delta) + q \cdot \delta}, \quad p_2(q) = \frac{q \cdot (1 - \delta)}{(1 - q) \cdot \delta + q \cdot (1 - \delta)}. \quad (19)$$

The decoder's threshold γ induces the two corresponding threshold ν_1 and ν_2 for the *interim* belief parameter $q \in [0, 1]$

$$\nu_1 = \frac{\gamma \cdot (1 - \delta)}{\delta \cdot (1 - \gamma) + \gamma \cdot (1 - \delta)}, \quad \nu_2 = \frac{\gamma \cdot \delta}{\gamma \cdot \delta + (1 - \delta) \cdot (1 - \gamma)}. \quad (20)$$

Thus the encoder's utility function $\Psi_e(q)$ represented by the red lines in Fig. 3 and the conditional entropy $h(q)$ reformulate as

$$\Psi_e(q) = 0 \cdot \mathbb{1}_{\{q \in [0, \nu_2]\}} + ((1 - q) \cdot \delta + q \cdot (1 - \delta)) \cdot \mathbb{1}_{\{q \in [\nu_2, \nu_1]\}} + 1 \cdot \mathbb{1}_{\{q \in [\nu_1, 1]\}}, \quad (21)$$

$$h(q) = ((1 - q) \cdot (1 - \delta) + q \cdot \delta) \cdot H_b(p_1(q)) + ((1 - q) \cdot \delta + q \cdot (1 - \delta)) H_b(p_2(q)). \quad (22)$$

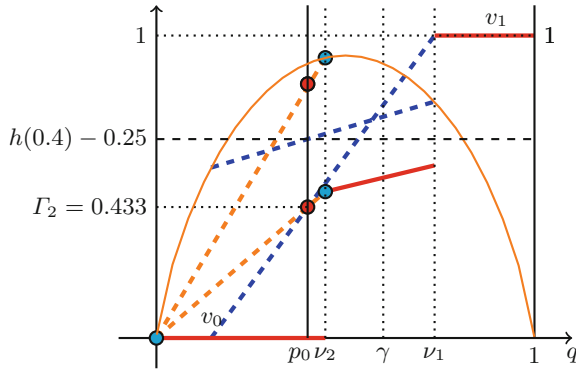


Fig. 3. Splitting over 2 posteriors ($q_1 = 0; q_2 = 0.4468$) with $C = 0.25$, $p_0 = 0.4$, $\delta = 0.35$, $\gamma = 0.6$.

The encoder’s optimal utility value is given by

$$\begin{aligned}
 \Gamma &= \sup_{\substack{q_1 \in [0, \nu_2], q_2 \in [\nu_2, \nu_1], \\ q_3 \in [\nu_1, 1]}} \left(\lambda_1 \cdot \Psi_e(q_1) + \lambda_2 \cdot \Psi_e(q_2) + \lambda_3 \cdot \Psi_e(q_3) \right) & (23) \\
 &= \sup_{\substack{q_1 \in [0, \nu_2], q_2 \in [\nu_2, \nu_1], \\ q_3 \in [\nu_1, 1]}} \left(\frac{(h(p_0) - C)((q_3 - q_1) \cdot (q_2 \cdot (1 - 2\delta) + \delta) + (q_1 - q_2))}{h(q_1) \cdot (q_2 - q_3) + h(q_2) \cdot (q_3 - q_1) + h(q_3) \cdot (q_1 - q_2)} \right. \\
 &+ \frac{(h(q_3) \cdot (q_1 - p_0) + h(q_1) \cdot (p_0 - q_3)) \cdot (q_2 \cdot (1 - 2\delta) + \delta)}{h(q_1) \cdot (q_2 - q_3) + h(q_2) \cdot (q_3 - q_1) + h(q_3) \cdot (q_1 - q_2)} \\
 &\left. + \frac{h(q_1) \cdot (q_2 - p_0) + h(q_2) \cdot (p_0 - q_1)}{h(q_1) \cdot (q_2 - q_3) + h(q_2) \cdot (q_3 - q_1) + h(q_3) \cdot (q_1 - q_2)} \right) & (24)
 \end{aligned}$$

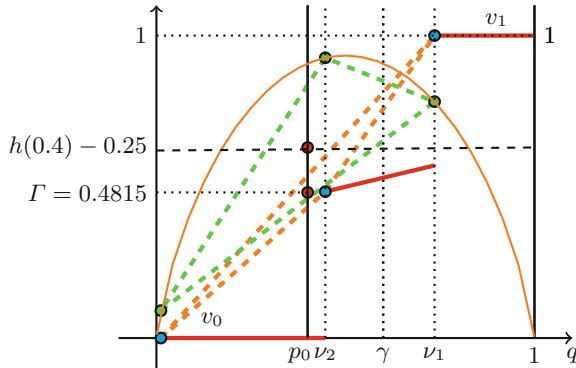


Fig. 4. Optimal Splittings over 3 posteriors ($q_1 = 0.012; q_2 = 0.4468; q_3 = 0.7358$) with $C = 0.25$, $p_0 = 0.4$, $\delta = 0.35$, $\gamma = 0.6$.

In some cases, the optimal splitting has only two posterior instead of three. Fig. 3 and Fig. 4 represent the optimal utility of the encoder depending on belief parameter q over a constrained communication channel with capacity $C = 0.25$ and with decoder’s private observation $\delta = 0.35$. Splitting over three posteriors instead of two, improves the encoder’s optimal payoff.

5 Numerical Simulations

In this section we investigate the impact of the private observation on the encoder’s optimal utility. Numerical simulations over (C, δ) regions are performed for both concavification problems Γ_0 and Γ , revealing the encoder’s optimal payoff values with and without decoder’s private observation.

5.1 Encoder’s Optimal Payoff Values

The optimal splitting of the prior over 3 posterior beliefs results in the encoder’s optimal payoff values shown in Fig. 5 with respect to the (C, δ) regions.

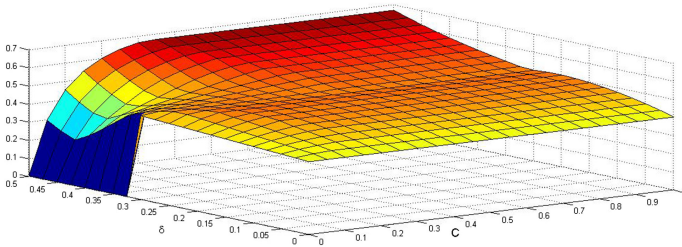


Fig. 5. Encoder’s optimal payoff evaluated with three posteriors w.r.t. δ and C for $p_0 = 0.4$ and $\gamma = 0.6$.

As the channel’s capacity increases, the encoder’s utility is improved without decoder’s side information. This is due to the fact that more capacity allows the transmission of more information and hence information can be optimally disclosed. However; with low capacity, the decoder’s side observation can enhance the utility of the encoder until the encoder has no capacity at all, it becomes optimal to have private information up to some threshold δ^* evaluated in Proposition 1 below.

5.2 Impact of the Decoder’s Private Signal

Proposition 1. *Let $C = 0$.*

- *If $p_0 < \gamma$ and $\delta \in [0, \frac{p_0 \cdot (\gamma - 1)}{p_0 \cdot (-1 + 2\gamma) - \gamma}] \cup [\frac{\gamma \cdot (1 - p_0)}{p_0 \cdot (1 - 2\gamma) + \gamma}, 1]$, then $\Gamma > \Gamma_0$.*
- *If $p_0 \geq \gamma$ then $\Gamma_0 \geq \Gamma$.*

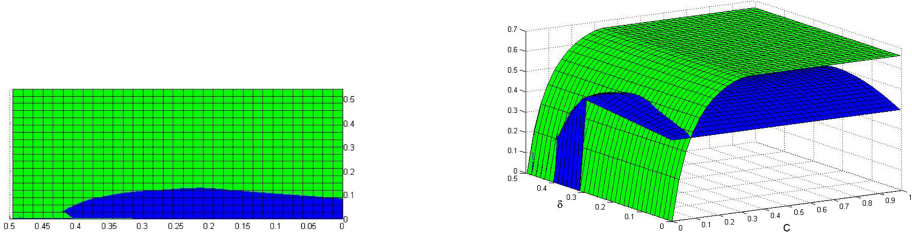


Fig. 6. (δ, C) regions for encoder's optimal utility with (blue) and without (green) decoder's private observation for $p_0 = 0.4$ and $\gamma = 0.6$. (Color figure online)

5.3 Impact of the Number of Posteriors

The encoder could potentially achieve a greater payoff by splitting the prior over three posterior beliefs instead of splitting over two posteriors only (Fig. 7).

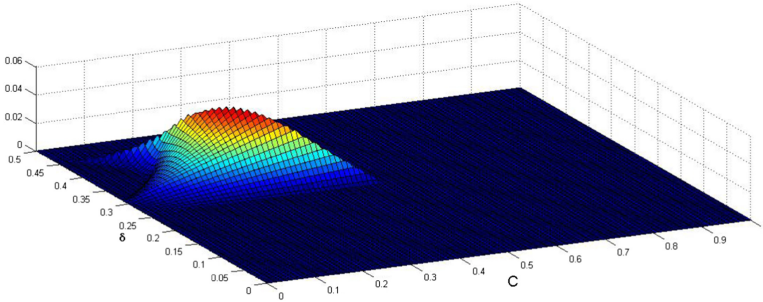


Fig. 7. Difference between optimal utility values obtained by splitting with three posteriors and two posteriors.

References

1. Akyol, E., Langbort, C., Başar, T.: Information-theoretic approach to strategic communication as a hierarchical game. *Proc. IEEE* **105**(2), 205–218 (2016)
2. Alonso, R., Câmara, O.: Bayesian persuasion with heterogeneous priors. *J. Econ. Theory* **165**(C), 672–706 (2016)
3. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. Wiley, Hoboken (1991)
4. Dughmi, S., Kempe, D., Qiang, R.: Persuasion with limited communication. In: *Proceedings of the 2016 ACM Conference on Economics and Computation. EC 2016*, pp. 663–680. Association for Computing Machinery, New York (2016). <https://doi.org/10.1145/2940716.2940781>
5. Kamenica, E., Gentzkow, M.: Bayesian persuasion. *Am. Econ. Rev.* **101**, 2590–2615 (2011)

6. Kolotilin, A., Mylovanov, T., Zapechelnyuk, A., Li, M.: Persuasion of a privately informed receiver. Discussion Papers 2016–21, School of Economics, The University of New South Wales (2016)
7. Laclau, M., Renou, L.: Public Persuasion (2016). <https://hal-pse.archives-ouvertes.fr/hal-01285218>
8. Le Treust, M., Tomala, T.: Information-theoretic limits of strategic communication. Preliminary Draft (2018). <https://arxiv.org/abs/1807.05147>
9. Le Treust, M., Tomala, T.: Persuasion with limited communication capacity. *J. Econ. Theory* **184**, 104940 (2019)
10. Le Treust, M., Tomala, T.: Strategic communication with side information at the decoder. arXiv preprint [arXiv:1911.04950](https://arxiv.org/abs/1911.04950) (2019)
11. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948)
12. Tuncel, E.: Slepian-wolf coding over broadcast channels. *IEEE Trans. Inf. Theory* **52**(4), 1469–1482 (2006)
13. Wyner, A.D., Ziv, J.: The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **22**(1), 1–11 (1976)