View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Archive Ouverte en Sciences de l'Information et de la Communication

Strategic Communication with Side Information at the Decoder

Mael Le Treust, Tristan Tomala

▶ To cite this version:

Mael Le Treust, Tristan Tomala. Strategic Communication with Side Information at the Decoder. 2019. hal-02357222

HAL Id: hal-02357222

https://hal.archives-ouvertes.fr/hal-02357222

Submitted on 9 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Strategic Communication with Side Information at the Decoder

Maël Le Treust* and Tristan Tomala †

* ETIS UMR 8051, Université Paris Seine, Université Cergy-Pontoise, ENSEA, CNRS,

6, avenue du Ponceau, 95014 Cergy-Pontoise CEDEX, FRANCE

Email: mael.le-treust@ensea.fr

† HEC Paris, GREGHEC UMR 2959

1 rue de la Libération, 78351 Jouy-en-Josas CEDEX, FRANCE

Email: tomala@hec.fr

Abstract

We investigate the problem of strategic point-to-point communication with side information at the decoder, in which the encoder and the decoder have mismatched distortion functions. The decoding process is not supervised, it returns the output sequence that minimizes the *decoder's* distortion function. The encoding process is designed beforehand and takes into account the decoder's distortion mismatch. When the communication channel is perfect and no side information is available at the decoder, this problem is referred to as *the Bayesian persuasion game* of Kamenica-Gentzkow in the Economics literature. We formulate the strategic communication scenario as a joint source-channel coding problem with side information at the decoder. The informational content of the source influences the design of the encoding since it impacts differently the two distinct distortion functions. The side information complexifies the analysis since the encoder is uncertain about the decoder's belief on the source statistics. We characterize the single-letter optimal solution by controlling the posterior beliefs induced by the Wyner-Ziv's source coding scheme. This confirms the benefit of sending encoded data bits even if the decoding process is not supervised.

* Maël Le Treust gratefully acknowledges financial support from INS2I CNRS, DIM-RFSI, SRV ENSEA, UFR-ST UCP, The Paris Seine Initiative and IEA Cergy-Pontoise.

† Tristan Tomala gratefully acknowledges the support of the HEC foundation and ANR/Investissements d'Avenir under grant ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX- 0047.

This work was presented in part at the 54th Allerton Conference, Monticello, Illinois, Sept. 2016 [1]; the XXVI Colloque Gretsi, Juan-Les-Pins, France, Sept. 2017 [2]; the International Zurich Seminar on Information and Communication, Switzerland, Feb. 2018 [3]. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01). The authors would like to thank Institute Henri Poincaré (IHP) in Paris, France, for hosting numerous research meetings.

I. INTRODUCTION

What information should be communicated to a receiver who minimizes a mismatched distortion metric? This new question arises in the context of the internet of things (IoT) composed of a variety of devices which are able to interact and coordinate with each other in order to create new applications/services and reach their own goals. In this context, wireless devices may have distinct objectives. For example, adjacent access points in crowded downtown areas, seeking to transmit at the same time, compete for the use of bandwidth; cognitive radio devices mitigate the interference effects by allocating their power budget over several parallel multiple access channels, as in [1, Sec. IV]. Such situations require new efficient techniques to coordinate communication traffic between devices whose objectives are *neither aligned, nor antagonistic*. This question differs from the classical paradigm in Information Theory which assumes that communicating devices are of two types: transmitters who pursue the common goal of transferring information; or opponents who try to mitigate the communication, e.g. the jammer corrupts the information, the eavesdropper infers it, the warden detects the covert transmission. In this work, we characterize the information-theoretic limits of strategic communication between interacting autonomous devices having general distortion functions, not necessarily aligned.

A. Scenario and contributions

We formulate the *strategic communication problem* as a joint source-channel coding problem with decoder's side information. We assume that the decoding process is not supervised, whereas the encoding process is designed in advance and takes into account the decoder's distortion mismatch. The side information partially discloses the source symbols to the decoder and complicates the design of the encoding, due to the mismatch of the distortion functions. Indeed, the encoder is uncertain about the decoder's belief on the source statistics. The closest paper in the literature is [4], in which no side information is available at the decoder. The novel contributions are listed below.

1. We characterize the optimal encoder's distortion for the problem of strategic communication with side information at the decoder; we relate it to Wyner-Ziv's rate-distortion function in [5] and to the separation result by Merhav-Shamai in [6].
2. We determine the posterior beliefs induced by the Wyner-Ziv's source coding, and we show that the sequence of decoder's optimal output is close to the output sequence of Wyner-Ziv's decoding. This confirms the benefit of sending encoded data bits to a non-supervised decoder, since the coding scheme reveals the *exact* amount of information needed.
3. We formulate the optimal solution in terms of a linear program with an information constraint and in terms of a convex closure of an auxiliary distortion function with an entropy constraint.

4. We investigate an example with binary source and side information, and we compute the optimal posterior beliefs for the doubly symmetric binary source of Wyner-Ziv [5, Sec. II].

We point out three essential features of strategic communication problem with decoder's side information.

1. Each source symbol has a different impact on the encoder and the decoder's distortion functions, hence it is optimal to encode each symbol differently.
2. The noiseless version of this problem without decoder's side information corresponds to the Bayesian persuasion game of Kamenica-Gentzkow [7]. In that case, the optimal information disclosure policy requires a fixed amount of information bits. When the channel capacity is larger than this amount, it is optimal not to use all the channel resource.
3. The decoder's side information has two opposite effects on the optimal encoder's distortion: it enlarges the set of decoder's posterior beliefs, so it may decrease the encoder's distortion; it reveals partial information to the decoder, so it forces some decoder's best-reply symbols which might be sub-optimal for the encoder's distortion.

B. Related literature

The problem of "strategic communication" in information-theoretic setting has been formulated by Akyol *et al.* in [8], [9], [10]. The authors characterize the optimal solution for Gaussian source, side information and channel, with the Crawford-Sobel's quadratic cost functions [11]. They prove that the optimal solution in the one-shot problem is also optimal when considering several strategic communication problems. This is not the case for general discrete source, channel and mismatched distortion functions. These results were further extended in [12] for non-identical prior beliefs about the source and the channel. The problem of strategic communication was introduced in the Control Theory literature by Sarıtaş *et al.* in [13] and [14]. The authors extend the model of Crawford-Sobel to multidimentional sources and noisy channels and they determine whether the optimal policies are linear or based on some quantization. The connection to the binary hypothesis-testing problem was pointed out in [15]. Sender-receiver games are also investigated in [16], for the problem of "strategic estimation" involving self-interested sensors; and in [17], [18], for the "network congestion" problem. In [19], [20], [21], the authors investigate the computational aspects of the Bayesian persuasion game, when the signals are noisy. In [22], [23], the interference channel coding problem is formulated as a game in which the users, i.e. the pairs of encoder/decoder, are allowed to use *any* encoding/decoding strategy. The authors compute the set of Nash equilibria for linear deterministic and Gaussian channels. The non-aligned devices' objectives are captured by distinct distortion functions. Coding for several distortion measures is investigated for "multiple descriptions coding" in [24], for the lossy version of "Steinberg's common reconstruction" problem in [25], for the problem of minimax distortion redundancy in [26], for "lossy broadcasting" in [27], for an alternative measure of "secrecy" in [28], [29], [30], [31].

The lossy source coding problem with mismatched distortion functions was formulated by Lapidoth, in [32]. In this model, the decoder attempts to reconstruct a source sequence that was encoded with respect to another distortion metric. The problem of the mismatch channel capacity was studied in [33], [34], [35], [36], [37], in which the decoding metric is not necessarily matched with the channel statistics.

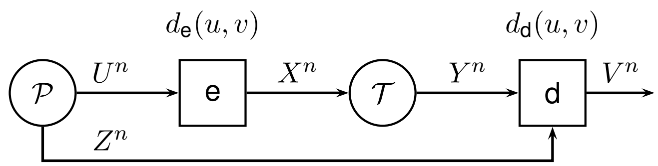

Fig. 1. The information source U and side information Z are drawn i.i.d. according to $\mathcal{P}_{UZ}$ and the channel $\mathcal{T}_{Y|X}$ is memoryless. The encoder e and the decoder d minimize mismatched distortion functions $d_{\mathsf{e}}(u,v) \neq d_{\mathsf{d}}(u,v)$.

The problem of "strategic information transmission" has been well studied in the Economics literature since the seminal paper by Crawford-Sobel [11]. In this model, a better-informed sender transmits a signal to a receiver, who takes an action which impacts both sender and receiver's utility functions. The problem consists in determining the *optimal information disclosure* policy given that the receiver' best-reply action affects the sender's utility, see [38] for a survey. In [7], Kamenica-Gentzkow introduced the Bayesian persuasion game in which the sender *commits* to an information disclosure policy before the game starts. This subtle change of rules of the game induces a very different equilibrium solution related to *Stackelberg equilibrium* [39], instead of *Nash equilibrium* [40]. This problem was later referred to as "information design" in [41], [42], [43] and extended to the setting with "heterogeneous beliefs" in [44] and [45]. In most of the articles in the Economics literature, the transmission between the sender and the receiver is noise-free; except in [46], [47], [48] where the noisy transmission is investigated in a finite block-length with no-error regime. Interestingly, Shannon's mutual information is widely accepted as a *cost of information* for the problem of "rational inattention" in [49] and for the problem of "costly persuasion" in [50], without explicit reference to a coding problem.

Entropy and mutual information appear endogenously in repeated games with finite automata and bounded recall [51], [52], [53], with private observation [54], or with imperfect monitoring [55], [56], [57]. In [58], the authors investigate a sender-receiver game with common interests by formulating a coding problem. They characterize the optimal solution via the mutual information. This result was later refined by Cuff in [59] and referred to as the "coordination problem" in [60], [61], [62], [63].

The paper is organized as follows. The strategic communication problem is formulated in Sec. II. The encoding and decoding strategies and the distortion functions are defined in Sec. II-A. The strategic communication scenario is introduced in Sec. II-B. Our coding result and the four different characterizations are stated in Sec. III. The first one is a linear program under an information constraint, formulated in Sec. III-A. The main Theorem is stated in Sec.III-B, and the sketch of proof is in Sec. III-C. In Sec. III-D, we reformulate the solution in terms of three different convex closures. Sec. IV provides an example based on a binary source, binary side information and binary decoder's actions. The proofs are stated in App A - D.

II. Strategic communication problem

A. *Coding strategies and distortion functions*

We consider the i.i.d. probability distribution of information source $\mathcal{P}_{UZ}$ and the conditional probability distribution $\mathcal{T}_{Y|X}$ of the memoryless channel, depicted in Fig. 1. Uppercase letters U denote the random variables, lowercase letters u denote the realizations and calligraphic fonts $\mathcal{U}$ denote the alphabets. Notations U^n, Z^n, X^n, Y^n, V^n stand for sequences of random variables of information source $u^n = (u_1, \ldots, u_n) \in \mathcal{U}^n$, decoder's side information $z^n \in \mathcal{Z}^n$, channel inputs $x^n \in \mathcal{X}^n$, channel outputs $y^n \in \mathcal{Y}^n$ and decoder's symbols $v^n \in \mathcal{V}^n$, respectively. The sets $\mathcal{U}$, $\mathcal{Z}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{V}$ have finite cardinality and the notation $\Delta(\mathcal{X})$ stands for the set of probability distributions over $\mathcal{X}$, i.e. the probability simplex. For a probability distribution $\mathcal{Q}_X \in \Delta(\mathcal{X})$, we write $\mathcal{Q}(x)$ instead of $\mathcal{Q}_X(x)$ for the probability value assigned to realization $x \in \mathcal{X}$. The notation $\mathcal{Q}_X(\cdot|y) \in \Delta(\mathcal{X})$ denotes the conditional distribution of X given the realization $y \in \mathcal{Y}$ and $\mathcal{Q}_X^{\otimes n} \in \Delta(\mathcal{X}^n)$ denotes the i.i.d. probability distribution. The distance between two probability distributions $\mathcal{Q}_X$ and $\mathcal{P}_X$ is based on L^1 norm, denoted by $||\mathcal{Q}_X - \mathcal{P}_X||_1 = \sum_{x \in \mathcal{X}} |\mathcal{Q}(x) - \mathcal{P}(x)|$. The notation $U -\!\!\circ\!\!- X -\!\!\circ\!\!- Y$ stands for the Markov chain property corresponding to $\mathcal{P}_{Y|XU} = \mathcal{P}_{Y|X}$.

Definition II.1 (Encoding and decoding strategies)

The encoding strategy σ and the decoding strategy τ are defined by

$$\sigma : \mathcal{U}^n \longrightarrow \Delta(\mathcal{X}^n), \tag{1}$$

$$\tau : \mathcal{Y}^n \times \mathcal{Z}^n \longrightarrow \Delta(\mathcal{V}^n). \tag{2}$$

Both strategies (σ, τ) are stochastic and induce a joint probability distribution $\mathcal{P}_{\sigma,\tau} \in \Delta(\mathcal{U}^n \times \mathcal{Z}^n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{V}^n)$ over the n-sequences of symbols, defined by

$$\mathcal{P}_{\sigma,\tau} = \bigg(\prod_{t=1}^{n} \mathcal{P}_{U_t Z_t}\bigg) \sigma_{X^n|U^n} \bigg(\prod_{t=1}^{n} \mathcal{T}_{Y_t|X_t}\bigg) \tau_{V^n|Y^n Z^n}. \tag{3}$$

The encoding and decoding strategies (σ, τ) are defined in the same way as for the joint source-channel coding problem with side information at the decoder studied in [6], based on Wyner-Ziv's setting in [5]. Unlike these previous works, we assume that the encoder and the decoder minimize distincts distortion functions.

Definition II.2 (Distortion functions) *The single-letter distortion functions of the encoder and decoder are defined by*

$$d_{\mathsf{e}} : \mathcal{U} \times \mathcal{Z} \times \mathcal{V} \longrightarrow \mathbb{R}, \tag{4}$$

$$d_{\mathsf{d}} : \mathcal{U} \times \mathcal{Z} \times \mathcal{V} \longrightarrow \mathbb{R}. \tag{5}$$

The long-run distortion functions $d_{\mathsf{e}}^n(\sigma, \tau)$ and $d_{\mathsf{d}}^n(\sigma, \tau)$ are evaluated with respect to the probability distribution $\mathcal{P}_{\sigma,\tau}$ induced by the strategies (σ, τ)

$$\begin{aligned} d_{\mathsf{e}}^n(\sigma, \tau) =& \mathbb{E}_{\sigma,\tau}\left[\frac{1}{n}\sum_{t=1}^{n} d_{\mathsf{e}}(U_t, Z_t, V_t)\right] \\ =& \sum_{u^n, z^n, v^n} \mathcal{P}_{\sigma,\tau}\big(u^n, z^n, v^n\big) \cdot \left[\frac{1}{n}\sum_{t=1}^{n} d_{\mathsf{e}}(u_t, z_t, v_t)\right], \end{aligned} \tag{6}$$

$$d_{\mathsf{d}}^n(\sigma, \tau) = \sum_{u^n, z^n, v^n} \mathcal{P}_{\sigma,\tau}\big(u^n, z^n, v^n\big) \cdot \left[\frac{1}{n}\sum_{t=1}^{n} d_{\mathsf{d}}(u_t, z_t, v_t)\right]. \tag{7}$$

B. Strategic communication scenario

In this work, the encoder and the decoder are autonomous devices that choose the strategy σ and τ in order to minimize their long-run distortion $d_{\mathsf{e}}^n(\sigma, \tau)$ and $d_{\mathsf{d}}^n(\sigma, \tau)$. We assume that the encoding strategy σ is designed in advance, so that the decoder knows σ before the transmission takes place. The decoder is *not supervised* and is free to choose any decoding strategy τ. This framework corresponds to the Bayesian persuasion game [7], in which the encoder commits to an *information disclosure policy* σ, and announces it to the decoder who chooses a best-reply strategy τ accordingly.

We assume that the strategic communication takes place as follows:

- The encoder chooses and announces the encoding strategy σ to the decoder.
- The sequences (U^n, Z^n, X^n, Y^n) are drawn according to the joint probability distribution $\Big(\prod_{t=1}^{n} \mathcal{P}_{U_t Z_t}\Big) \sigma_{X^n|U^n} \Big(\prod_{t=1}^{n} \mathcal{T}_{Y_t|X_t}\Big)$.

- The decoder knows σ, observes the sequences of symbols (Y^n, Z^n), and draws a sequence of symbols V^n according to a best-reply strategy $\tau_{V^n|Y^nZ^n}$.

Definition II.3 (Decoder's Best-Replies) *For any encoding strategy σ, the set of best-reply decoding strategies $BR_d(\sigma)$ is defined by*

$$BR_d(\sigma) = \text{argmin}_\tau \, d_d^n(\sigma,\tau) = \Big\{\tau, \text{ s.t. } d_d^n(\sigma,\tau) \leq d_d^n(\sigma,\widetilde{\tau}), \, \forall \widetilde{\tau} \neq \tau \Big\}. \tag{8}$$

In case there are several best-reply strategies, we assume that the decoder chooses the one that maximizes the encoder's distortion $\max_{\tau \in \mathsf{BR_d}(\sigma)} d_{\mathsf{e}}^n(\sigma,\tau)$, so that encoder's distortion is robust to the exact specification of decoder's strategy.

The coding problem under investigation consists in minimizing the encoder's long-run distortion

$$\inf_{\sigma} \max_{\tau \in \mathsf{BR_d}(\sigma)} d_{\mathsf{e}}^n(\sigma,\tau). \tag{9}$$

The decoding process τ is not supervised, it is strategic, causing the mismatch of the decoder's output sequence. The design of the encoding strategy σ anticipates this mismatch. Does the "strategic decoder" necessarily decode the coded bits of information? We provide a positive answer in Theorem III.3, by refining the analysis of the Wyner-Ziv's coding scheme [5]. We show that the symbols induced by Wyner-Ziv's decoding τ^{wz} coincide with those induced by any best-reply $\tau \in \mathsf{BR_d}(\sigma^{\mathsf{wz}})$ to Wyner-Ziv's encoding σ^{wz}, for a large fraction of stages.

Remark II.4 (Stackelberg v.s. Nash equilibrium) The optimization problem in (9) corresponds to a Stackelberg equilibrium [39] in which the encoder is the leader and the decoder is the follower, unlike the Nash equilibrium [40] in which the two devices choose their strategy simultaneously.

Remark II.5 (Equal distortion functions) When the encoder and decoder have equal distortion functions $d_{\mathsf{e}} = d_{\mathsf{d}}$, the problem in (9) boils down to the classical approach of Wyner-Ziv [5] and Merhav-Shamai [6], in which both strategies (σ,τ) are chosen *jointly*, in order to minimize a distortion function

$$\inf_{\sigma} \max_{\tau \in \mathsf{BR_d}(\sigma)} d_{\mathsf{e}}^n(\sigma,\tau) = \inf_{\sigma} \min_{\tau} d_{\mathsf{e}}^n(\sigma,\tau) = \min_{(\sigma,\tau)} d_{\mathsf{e}}^n(\sigma,\tau), \tag{10}$$

since by Definition II.3, $\tau \in \mathsf{BR_d}(\sigma) \Longleftrightarrow d_{\mathsf{d}}^n(\sigma,\tau) = \min_{\tau'} d_{\mathsf{d}}^n(\sigma,\tau')$.

III. CHARACTERIZATIONS

A. Linear program with an information constraint

We define the encoder's optimal distortion $D_{\mathsf{e}}^\star$.

Definition III.1 (Target distributions) *We consider an auxiliary random variable $W \in \mathcal{W}$ with $|\mathcal{W}| = \min\big(|\mathcal{U}|+1, |\mathcal{V}|^{|\mathcal{Z}|}\big)$. The set $\mathbb{Q}_0$ of target probability distributions is defined by*

$$\mathbb{Q}_0 = \bigg\{ \mathcal{P}_{UZ}\mathcal{Q}_{W|U}, \quad \textit{s.t.}, \qquad \max_{\mathcal{P}_X} I(X;Y) - I(U;W|Z) \geq 0 \bigg\}. \tag{11}$$

We define the set $\mathbb{Q}_2\big(\mathcal{Q}_{UZW}\big)$ of single-letter best-replies of the decoder

$$\mathbb{Q}_2\big(\mathcal{Q}_{UZW}\big) = \text{argmin}_{\mathcal{Q}_{V|WZ}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW}\\ \mathcal{Q}_{V|WZ}}} \Big[d_{\mathsf{d}}(U,Z,V) \Big]. \tag{12}$$

The encoder's optimal distortion $D_{\mathsf{e}}^{\star}$ is given by

$$D_{\mathsf{e}}^{\star} = \inf_{\mathcal{Q}_{UZW}\in\mathbb{Q}_0} \max_{\substack{\mathcal{Q}_{V|WZ}\in\\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW}\\ \mathcal{Q}_{V|WZ}}} \Big[d_{\mathsf{e}}(U,Z,V) \Big]. \tag{13}$$

We discuss the above definitions.

- The information constraint (11) of the set $\mathbb{Q}_0$ involves the channel capacity $\max_{\mathcal{P}_X} I(X;Y)$ and the Wyner-Ziv's information rate $I(U;W|Z) = I(U;W) - I(W;Z)$, stated in [5]. It corresponds to the separation result by Shannon [64], extended to the Wyner-Ziv setting by Merhav-Shamai in [6].
- For the clarity of the presentation, the set $\mathbb{Q}_2\big(\mathcal{Q}_{UZW}\big)$ contains stochastic functions $\mathcal{Q}_{V|WZ}$, even if for the linear problem (12), some optimal $\mathcal{Q}_{V|WZ}$ are deterministic. If there are several optimal $\mathcal{Q}_{V|WZ}$, we assume the decoder chooses the one that maximize encoder's distortion: $\max_{\substack{\mathcal{Q}_{V|WZ}\in\\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}\big[d_{\mathsf{e}}(U,Z,V)\big]$, so that encoder's distortion is robust to the exact specification of $\mathcal{Q}_{V|WZ}$.
- The infimum over $\mathcal{Q}_{UZW} \in \mathbb{Q}_0$ is not a minimum since the function $\max_{\substack{\mathcal{Q}_{V|WZ}\in\\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}\big[d_{\mathsf{e}}(U,Z,V)\big]$ is not continuous with respect to $\mathcal{Q}_{UZW}$, see Fig. 10, 11.
- In [65, Theorem IV.2], the author shows that the sets $\mathbb{Q}_0$ and $\mathbb{Q}_2$ correspond to the target probability distributions $\mathcal{Q}_{UZW}\mathcal{Q}_{V|WZ}$ that are achievable for the problem of *empirical coordination*, see also [60], [62]. As noticed in [66] and [67], the empirical coordination approach allows us to characterize the "core of the decoder's knowledge", which captures what the decoder is able to infer about the random variables involved in the problem.
- The value $D_{\mathsf{e}}^{\star}$ corresponds to the Stackelberg equilibrium payoff of an auxiliary one-shot game in which the decoder chooses $\mathcal{Q}_{V|WZ}$, knowing in advance that the encoder has chosen $\mathcal{Q}_{W|U} \in \mathbb{Q}_0$ and the distortion functions are $\mathbb{E}\big[d_{\mathsf{e}}(U,Z,V)\big]$ and $\mathbb{E}\big[d_{\mathsf{d}}(U,Z,V)\big]$.

Remark III.2 (Equal distortion functions) When the encoder and the decoder have equal distortion functions $d_{\mathsf{d}} = d_{\mathsf{e}}$, the set $\mathbb{Q}_2\big(\mathcal{Q}_{UZW}\big)$ is equal to $\text{argmin}_{\mathcal{Q}_{V|WZ}} \mathbb{E}\big[d_{\mathsf{e}}(U,Z,V)\big]$. Thus, we have

$$\max_{\substack{\mathcal{Q}_{V|WZ}\in\\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW}\\ \mathcal{Q}_{V|WZ}}} \Big[d_{\mathsf{e}}(U,Z,V) \Big] = \min_{\mathcal{Q}_{V|WZ}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW}\\ \mathcal{Q}_{V|WZ}}} \Big[d_{\mathsf{e}}(U,Z,V) \Big]. \tag{14}$$

Hence, the encoder's optimal distortion $D_{\mathsf{e}}^{\star}$ is equal to:

$$D_{\mathsf{e}}^{\star} = \inf_{\mathcal{Q}_{UZW}\in\mathbb{Q}_0} \max_{\substack{\mathcal{Q}_{V|WZ}\in \\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW} \\ \mathcal{Q}_{V|WZ}}}\Big[d_{\mathsf{e}}(U,Z,V)\Big] \tag{15}$$

$$= \inf_{\mathcal{Q}_{UZW}\in\mathbb{Q}_0} \min_{\mathcal{Q}_{V|WZ}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW} \\ \mathcal{Q}_{V|WZ}}}\Big[d_{\mathsf{e}}(U,Z,V)\Big] \tag{16}$$

$$= \min_{\substack{\mathcal{Q}_{UZW}\in\mathbb{Q}_0, \\ \mathcal{Q}_{V|WZ}}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW} \\ \mathcal{Q}_{V|WZ}}}\Big[d_{\mathsf{e}}(U,Z,V)\Big]. \tag{17}$$

The infimum in (16) is replaced by a minimum in (17) due to the compactness of $\mathbb{Q}_0$ and the continuity of $\min_{\mathcal{Q}_{V|WZ}} \mathbb{E}\Big[d_{\mathsf{e}}(U,Z,V)\Big]$ with respect to $\mathcal{Q}_{UZW}$. We recover the *distortion-rate* function corresponding to [6, Theorem 1].

B. Main result

We denote $\mathbb{N}^{\star} = \mathbb{N} \setminus \{0\}$ and characterize the encoder's long-run optimal distortion (9).

Theorem III.3 (Main result) *The long-run optimal distortion of the encoder satisfies:*

$$\forall \varepsilon > 0,\ \exists \bar{n} \in \mathbb{N}^{\star},\ \forall n \geq \bar{n}, \qquad \inf_{\sigma} \max_{\tau \in BR_d(\sigma)} d_e^n(\sigma,\tau) \leq D_e^{\star} - \varepsilon, \tag{18}$$

$$\forall n \in \mathbb{N}^{\star}, \qquad \inf_{\sigma} \max_{\tau \in BR_d(\sigma)} d_e^n(\sigma,\tau) \geq D_e^{\star}. \tag{19}$$

When removing the decoder's side information $\mathcal{Z} = \emptyset$ and changing the infimum to a supremum, we recover our previous result [4, Theorem 3.1]. Theorem III.3 and Fekete's lemma for a sub-additive sequence in [68], show that the long-run encoder's distortion converges to its infimum.

$$D_{\mathsf{e}}^{\star} = \lim_{n\to+\infty} \inf_{\sigma} \max_{\tau\in\mathsf{BR}_{\mathsf{d}}(\sigma)} d_{\mathsf{e}}^n(\sigma,\tau) = \inf_{n\in\mathbb{N}^{\star}} \inf_{\sigma} \max_{\tau\in\mathsf{BR}_{\mathsf{d}}(\sigma)} d_{\mathsf{e}}^n(\sigma,\tau). \tag{20}$$

C. Sketch of proof of Theorem III.3

We provide some intuitions for the main arguments of the proofs, which are given in App. B and C.

Proof of the achievability result (18). We analyse the posterior beliefs induced by the random coding scheme obtained by concatenating Wyner-Ziv's source coding [5] with Shannon's channel coding [64]. We consider a probability distribution $\mathcal{Q}_{UZW}$ such that 1) the capacity is strictly positive, 2) the information constraints are satisfied with strict inequalities, 3) the set of best-reply symbol $\mathcal{V}^{\star}\big(z, \mathcal{Q}_U(\cdot|z,w)\big)$ of

Definition III.4, is a singleton for all $(z,w)\in\mathcal{Z}\times\mathcal{W}$. We introduce the Wyner-Ziv's coding rates R, $\mathsf{R_L}$ for the messages (M,L) and denote $\eta>0$ the parameter such that

$$\mathsf{R}+\mathsf{R_L} = I(U;W)+\eta, \tag{21}$$

$$\mathsf{R_L} \leq I(Z;W)-\eta, \tag{22}$$

$$\mathsf{R} \leq \max_{\mathcal{P}_X} I(X;Y)-\eta. \tag{23}$$

We denote by $E_\delta=0$, the event on which the messages (M,L) are recovered by the decoder and the sequences are jointly typical $\big(U^n,Z^n,W^n,X^n,Y^n\big)\in A_\delta$ with tolerance $\delta>0$.

The major step is to show that the posterior beliefs $\mathcal{P}_{\sigma,U_t}(\cdot|y^n,z^n,E_\delta=0)$ induced by coding scheme σ regarding U_t at stage $t\in\{1,\dots,n\}$, are close on average to the target probability distribution $\mathcal{Q}_U(\cdot|w,z)$:

$$\mathbb{E}_\sigma\left[\frac{1}{n}\sum_{t=1}^{n} D\bigg(\mathcal{P}_{\sigma,U_t}(\cdot|Y^n,Z^n,E_\delta=0)\bigg|\bigg|\mathcal{Q}_{U_t}(\cdot|W_t,Z_t)\bigg)\right]$$
$$\leq 2\delta+\eta+\frac{2}{n}+2\log_2|\mathcal{U}|\cdot\mathcal{P}_\sigma\big(E_\delta=1\big):=\epsilon. \tag{24}$$

This is the purpose of the proof of Proposition B.1, stated in Appendix B-C.

Proof of the converse result (19). For any encoding strategy σ of length $n\in\mathbb{N}^\star$, we introduce an auxiliary random variable $W=(Y^n,Z^{-T},T)$, where T is the uniform random variable over $\{1,\dots,n\}$ and Z^{-T} stands for $(Z_1,\dots,Z_{t-1},Z_{t+1},\dots Z_n)$, where Z_T has been removed. We show that the Markov chain $W -\!\!\circ\!\!- U_T -\!\!\circ\!\!- Z_T$ is satisfied and that the probability distribution $\mathcal{P}_{UZW}$ satisfy

$$\mathcal{P}(u,z,w)=\frac{1}{n}\cdot\mathcal{P}_\sigma\big(u_t,z_t,y^n,z^{-t}\big),\quad\forall(u,w,z,u^n,z^n,y^n). \tag{25}$$

We define $\tilde{\tau}_{V|WZ}=\tau_{V_T|Y^nZ^n}$ and we prove that for both encoder and decoder, the long-run distortion writes

$$d^n_{\mathsf{e}}(\sigma,\tau)=\sum_{u,z,w}\mathcal{P}(u,z,w)\sum_{v}\tilde{\tau}(v|w,z)\cdot d_{\mathsf{e}}(u,z,v), \tag{26}$$

hence

$$\tau\in\text{argmin}_{\tau'_{V^n|Y^nZ^n}}\mathbb{E}_{\sigma\tau'}\left[\frac{1}{n}\sum_{t=1}^{n}\phi_{\mathsf{d}}(u_t,z_t,v_t)\right]\Longleftrightarrow\tilde{\tau}_{V|WZ}\in\mathbb{Q}_2\big(\mathcal{P}_{UZW}\big). \tag{27}$$

Well known arguments from [5] and [6] show that

$$0\leq\max_{\mathcal{P}_X}I(X;Y)-I(U;W|Z), \tag{28}$$

therefore, for any encoding strategy σ and all n, we have

$$\max_{\tau\in\mathsf{BR}_{\mathsf{d}}(\sigma)}d^n_{\mathsf{e}}(\sigma,\tau)\geq\inf_{\mathcal{Q}_{UZW}\in\mathbb{Q}_0}\max_{\substack{\mathcal{Q}_{V|WZ}\in\\\mathbb{Q}_2(\mathcal{Q}_{UZW})}}\mathbb{E}_{\substack{\mathcal{Q}_{UZW}\\\mathcal{Q}_{V|WZ}}}\Big[d_{\mathsf{e}}(U,Z,V)\Big]=D^\star_{\mathsf{e}}. \tag{29}$$

D. Convex closure formulation

The convex closure of a function f is the largest convex function $\text{vex}\, f : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ everywhere smaller than f on X. In this section, we reformulate the encoder's optimal distortion $D_{\mathsf{e}}^{\star}$ in terms of a convex closure, similarly to [7, Corollary 1] and [4, Definition 2.4]. This alternative approach may simplify the optimization problem in (13), by plugging the decoder's posterior beliefs and best-reply symbols into the encoder's distortion function. The goal of the strategic communication is to *control the posterior beliefs* of the decoder, knowing it will choose a best-reply symbol afterwards.

Before the transmission, the decoder holds a prior belief corresponding to the source's statistics $\mathcal{P}_U \in \Delta(\mathcal{U})$. After observing the pair of symbols $(w,z) \in \mathcal{W} \times \mathcal{Z}$, the decoder updates its posterior belief $\mathcal{Q}_U(\cdot|z,w) \in \Delta(\mathcal{U})$ according to Bayes rule $\mathcal{Q}(u|z,w) = \frac{\mathcal{P}(u,z)\mathcal{Q}(w|u)}{\sum_{u'}\mathcal{P}(u',z)\mathcal{Q}(w|u')}$, for all $(u,z,w) \in \mathcal{U}\times\mathcal{W}\times\mathcal{Z}$.

Lemma 1 (Markov property on posterior belief) *The Markov chain property $Z -\!\!\circ\!\!- U -\!\!\circ\!\!- W$ implies that the* posterior beliefs $\mathcal{Q}_U(\cdot|z,w) \in \Delta(\mathcal{U})$ *can be expressed from the* interim beliefs $\mathcal{Q}_U(\cdot|w) \in \Delta(\mathcal{U})$,

$$\mathcal{Q}(u|w,z) = \frac{\mathcal{Q}(u,z,w)}{\mathcal{Q}(z,w)} = \frac{\mathcal{Q}(u,z|w)}{\sum_{u'}\mathcal{Q}(u',z|w)} = \frac{\mathcal{Q}(u|w)\mathcal{P}(z|u)}{\sum_{u'}\mathcal{Q}(u'|w)\mathcal{P}(z|u')}, \qquad \forall (u,z,w) \in \mathcal{U}\times\mathcal{Z}\times\mathcal{W}. \tag{30}$$

Definition III.4 (Best-reply) *For each symbol $z \in \mathcal{Z}$ and belief $p \in \Delta(\mathcal{U})$, the decoder chooses the best-reply $v^{\star}(z,p)$ that belongs to the set $\mathcal{V}^{\star}(z,p)$, defined by*

$$\mathcal{V}^{\star}(z,p) = \text{argmax}_{v \in \text{argmin}\, \mathbb{E}_p\left[d_{\mathsf{d}}(U,z,v)\right]} \mathbb{E}_p\Big[d_{\mathsf{e}}(U,z,v)\Big]. \tag{31}$$

If several symbols are best-reply to $z \in \mathcal{Z}$ and belief $p \in \Delta(\mathcal{U})$, the decoder chooses the worst one for encoder's distortion. This is a reformulation of the maximum in (13).

Definition III.5 (Robust distortion) *For a symbol $z \in \mathcal{Z}$ and a belief $p \in \Delta(\mathcal{U})$, the encoder's* robust distortion function *is defined by*

$$\psi_{\mathsf{e}}(z,p) = \mathbb{E}_p\Big[d_{\mathsf{e}}\big(U,z,v^{\star}(z,p)\big)\Big], \tag{32}$$

where the best-reply $v^{\star}(z,p)$ belongs to the set defined by (31).

Definition III.6 (Average distortion and average entropy) *For each belief $p \in \Delta(\mathcal{U})$, we define the average distortion function $\Psi_{\mathsf{e}}(p)$ and average entropy function $h(p)$ by*

$$\Psi_{\mathsf{e}}(p) = \sum_{u,z} p(u)\cdot\mathcal{P}(z|u)\cdot\psi_{\mathsf{e}}\big(z, p_U(\cdot|z)\big), \quad \text{where} \quad p_U(u|z) = \frac{p(u)\cdot\mathcal{P}(z|u)}{\sum_{u'}p(u')\cdot\mathcal{P}(z|u')}\ \forall (u,z), \tag{33}$$

$$h(p) = H(p) + \sum_u p(u)\cdot H\Big(\mathcal{P}_Z(\cdot|u)\Big) - H\Big(\sum_u p(u)\cdot\mathcal{P}_Z(\cdot|u)\Big). \tag{34}$$

The conditional probability distribution $\mathcal{P}_Z(\cdot|u)$ is given by the information source. Note that $h(\mathcal{P}_U) = H(U|Z)$.

Lemma 2 (Concavity) *The average entropy $h(p)$ is concave in $p \in \Delta(\mathcal{U})$.*

Proof. [Lemma 2] The average entropy $h(p)$ in (34) is equal to the conditional entropy $H(U|Z)$ evaluated with respect to the probability distribution $p \cdot \mathcal{P}_{Z|U} \in \Delta(\mathcal{U} \times \mathcal{Z})$. The mutual information $I(U;Z)$ is convex in $p \in \Delta(\mathcal{U})$ (see [69, pp. 23]), and the entropy $H(U)$ is concave in $p \in \Delta(\mathcal{U})$. Hence the conditional entropy $h(p) = H(U|Z) = H(U) - I(U;Z)$ is concave in $p \in \Delta(\mathcal{U})$. □

Theorem III.7 (Convex closure) *The solution $D_{\mathsf{e}}^{\star}$ of (13) is the convex closure of $\Psi_{\mathsf{e}}(p)$ evaluated at the prior distribution $\mathcal{P}_U$, under an information constraint,*

$$D_{\mathsf{e}}^{\star} = \inf \Bigg\{ \sum_w \lambda_w \cdot \Psi_{\mathsf{e}}(p_w) \quad \textit{s.t.} \quad \sum_w \lambda_w \cdot p_w = \mathcal{P}_U \in \Delta(\mathcal{U}), \quad \textit{and} \quad \sum_w \lambda_w \cdot h(p_w) \geq H(U|Z) - \max_{\mathcal{P}_X} I(X;Y) \Bigg\}, \tag{35}$$

where the infimum is taken over $\lambda_w \in [0,1]$ summing up to 1 and $p_w \in \Delta(\mathcal{U})$, for each $w \in \mathcal{W}$ with $|\mathcal{W}| = \min\big(|\mathcal{U}|+1, |\mathcal{V}|^{|\mathcal{Z}|}\big)$.

The proof of Theorem III.7, stated in App. A, rely on the Markov chain property $Z -\!\!\circ\!\!- U -\!\!\circ\!\!- W$. When removing, the decoder's side information, e.g. $|\mathcal{Z}| = 1$, and changing the infimum into a supremum, we recover the value of the optimal splitting problem of [4, Definition 2.4]. The "splitting Lemma" by Aumann and Maschler [70], also called "Bayes plausibility" in [7], ensures that the strategy $\mathcal{Q}(w|u) = \frac{\lambda_w \cdot p_w(u)}{\mathcal{P}(u)}$ induces the collection of posterior beliefs $(\lambda_w, p_w)_{w \in \mathcal{W}}$, also referred to as "the splitting of the prior belief". Formulation (35) provides an alternative point of view on the encoder's optimal distortion (13).

- The optimal solution $D_{\mathsf{e}}^{\star}$ can be found by adapting the concavification method [70], to the minimization problem. In Sec IV, we investigate an example involving binary source and side information and we compute explicitly the optimal strategy for the Wyner-Ziv's example with a *doubly symmetric binary source* (DSBS), in [5, Sec. II, pp. 3].
- When the channel is perfect and has a large input alphabet $|\mathcal{X}| \geq \min(|\mathcal{U}|, |\mathcal{V}|^{|\mathcal{Z}|})$, the strategic communication problem is equivalent to several i.i.d. copies of the one-shot problem, whose optimal solution is obtained by removing the information constraint (35). This noise-free setting is related to the problem of persuasion with heterogeneous beliefs, investigated in [44] and [45].

- The information constraint $\sum_w \lambda_w \cdot h(p_w) \geq H(U|Z) - \max_{\mathcal{P}_X} I(X;Y)$ in (35) is a reformulation of $I(U;W|Z) \leq \max_{\mathcal{P}_X} I(X;Y)$ in (11), since

$$\sum_w \lambda_w \cdot h(p_w) = \sum_w \lambda_w \cdot H(U|Z, W = w) = H(U|Z, W). \tag{36}$$

- The dimension of the problem (35) is $|\mathcal{U}|$. Caratheodory's Lemma (see [71, Corollary 17.1.5, pp. 157] and [4, Corollary A.2, pp. 26]) provides the cardinality bound $|\mathcal{W}| = |\mathcal{U}| + 1$.
- The cardinality of $\mathcal{W}$ is also restricted by the vector of recommended symbols $|\mathcal{W}| = |\mathcal{V}|^{|\mathcal{Z}|}$, telling to the decoder which symbol to return for each side information. Otherwise assume that two posteriors p_{w_1} and p_{w_2} induce the same vectors of symbols $v^1 = (v^1_1, \ldots, v^1_{|\mathcal{Z}|}) = v^2 = (v^2_1, \ldots, v^2_{|\mathcal{Z}|})$. Then, both posteriors p_{w_1} and p_{w_2} can be replaced by their average:

$$\widetilde{p} = \frac{\lambda_{w_1} \cdot p_{w_1} + \lambda_{w_2} \cdot p_{w_2}}{\lambda_{w_1} + \lambda_{w_2}}, \tag{37}$$

without changing the distortion and still satisfying the information constraint:

$$h(\widetilde{p}) \geq \frac{\lambda_{w_1} \cdot h(p_{w_1}) + \lambda_{w_2} \cdot h(p_{w_2})}{\lambda_{w_1} + \lambda_{w_2}} \tag{38}$$

$$\Longrightarrow \sum_{\substack{w \neq w_1, \\ w \neq w_2}} \lambda_w \cdot h(p_w) + (\lambda_{w_1} + \lambda_{w_2}) \cdot h(\widetilde{p}) \geq H(U|Z) - \max_{\mathcal{P}_X} I(X;Y). \tag{39}$$

Inequality (38) comes from the concavity of $h(p)$, stated in Lemma 2.

The splitting under information constraint of Theorem III.7 can be reformulated in terms of Lagrangian and in terms of the convex closure of $\tilde{\Psi}_{\mathsf{e}}(p, \nu)$ defined by

$$\tilde{\Psi}_{\mathsf{e}}(p, \nu) = \begin{cases} \Psi_{\mathsf{e}}(p), \text{ if } \nu \leq h(p), \\ +\infty, \text{ otherwise.} \end{cases} \tag{40}$$

Theorem III.8 *The optimal solution $D^\star_{\mathsf{e}}$ reformulates as:*

$$D^\star_{\mathsf{e}} = \sup_{t \geq 0} \left\{ \text{vex} \left[\Psi_{\mathsf{e}} + t \cdot h \right] \left(\mathcal{P}_U \right) - t \cdot \left(H(U|Z) - \max_{\mathcal{P}_X} I(X;Y) \right) \right\} \tag{41}$$

$$= \text{vex}\, \tilde{\Psi}_{\mathsf{e}} \Big(\mathcal{P}_U, H(U|Z) - \max_{\mathcal{P}_X} I(X;Y) \Big). \tag{42}$$

Equation (41) is the convex closure of a Lagrangian that integrates the information constraint and equation (42) corresponds to the convex closure of a a bi-variate function where the information constraint requires an additional dimension. The proof follows directly from [4, Theorem 3.3, pp. 37], by replacing concave closure by convex closure.

IV. Example with binary source and side information

We consider a binary source $U \in \{u_0, u_1\}$ with probability distribution $\mathcal{P}(u_1) = p_0 \in [0,1]$. The binary side information $Z \in \{z_0, z_1\}$ is drawn according to the conditional probability distribution $\mathcal{P}(z|u)$ with parameter $\delta_0 \in [0,1]$ and $\delta_1 \in [0,1]$. The cardinality bound in Definition III.1 is $|\mathcal{W}| = \min\big(|\mathcal{U}|+1, |\mathcal{V}|^{|\mathcal{Z}|}\big) = 3$, hence the random variable W is drawn according to the conditional probability distribution $\mathcal{Q}(w|u)$ with parameters $(\alpha_k, \beta_k)_{k\in\{1,2,3\}} \in [0,1]^6$ such that $\sum_k \alpha_k = \sum_k \beta_k = 1$. The joint probability distribution $\mathcal{P}(u,z)\mathcal{Q}(w|u)$ is depicted in Fig. 2.

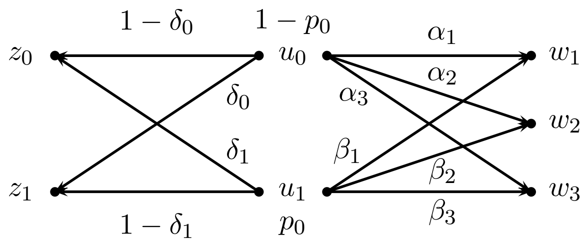

Fig. 2. Joint probability distribution $\mathcal{P}(u,z)\mathcal{Q}(w|u)$ depending on parameters p_0, δ_0, δ_1, $(\alpha_k, \beta_k)_{k\in\{1,2,3\}}$.

The encoder minimizes the Hamming distortion $d_{\mathsf{e}}(u,v)$ given by Fig. 3. The decoder's distortion in Fig. 4 includes an extra cost $\kappa \in [0,1]$ when it returns the symbol v_1 instead of the symbol v_0. The extra cost κ may capture a computing cost, an energy cost, or the fact that an estimation error for the symbol v_1 is more harmful than an estimation error for the symbol v_0.

	v_0	v_1
u_0	0	1
u_1	1	0

Fig. 3. Encoder's distortion function $d_{\mathsf{e}}(u,v)$.

	v_0	v_1
u_0	0	$1+\kappa$
u_1	1	κ

Fig. 4. Decoder's distortion $d_{\mathsf{d}}(u,v)$ with extra cost $\kappa \in [0,1]$.

A. *Decoder's best-reply*

After receiving the pair of symbols (w,z), the decoder updates its *posterior belief* $\mathcal{Q}_U(\cdot|w,z) \in \Delta(\mathcal{U})$, according to Bayes rule. We denote by $p = \mathcal{Q}(u_1|w,z) \in [0,1]$ the parameter of the posterior belief. Given the extra cost is $\kappa = \frac{3}{4}$ and we denote by $\gamma = \frac{1+\kappa}{2} = \frac{7}{8}$ the *belief threshold* at which the decoder changes its symbol, as in Fig. 5. When the decoder's belief is exactly equal to the threshold $p = \gamma = \frac{7}{8}$, the decoder is indifferent between the two symbols $\{v_0, v_1\}$, by convention we assume that it chooses

v_0, i.e. the worst symbol for the encoder. Hence the decoder chooses a best-reply $v_0^\star$ or $v_1^\star$ depending on the interval $[0,\gamma]$ or $(\gamma,1]$ in which lies the belief parameter $p \in [0,1]$, see Fig. 5.

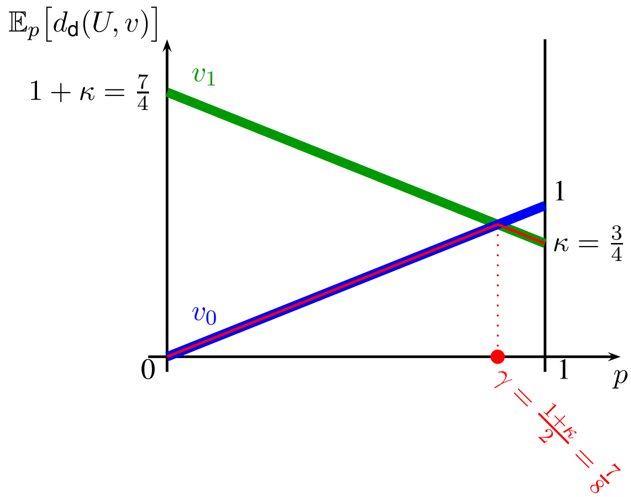

Fig. 5. Decoder's expected distortion $\mathbb{E}_p\big[d_{\mathsf{d}}(U,v)\big] = (1-p)\cdot d_{\mathsf{d}}(u_0,v) + p \cdot d_{\mathsf{d}}(u_1,v)$ for $v \in \{v_0, v_1\}$, depending on the belief parameter $p = \mathcal{Q}(u_1|w,z) \in [0,1]$. For an extra cost $\kappa = \frac{3}{4}$, the decoder's best-reply is the symbol $v_0^\star$ if the posterior belief $p \in [0,\gamma]$ and $v_1^\star$ if $p \in (\gamma, 1]$ with $\gamma = \frac{7}{8}$.

B. Interim and posterior belief

The correlation between random variables (U,Z) is fixed whereas the correlation between random variables (U,W) is selected by the encoder. This imposes a strong relationship between the *iterim belief* $\mathcal{Q}_{U|W}$ induced by the encoder, and the *posterior belief* $\mathcal{Q}_{U|WZ}$ that determine the decoder's best-reply symbol v. For a symbol $w \in \mathcal{W}$, we denote the *interim belief* by $q = \mathcal{Q}(u_1|w) \in [0,1]$. Lemma 1 ensures that the posterior belief depending on the side information z_0 or z_1, are given by

$$\mathcal{Q}(u_1|w,z_0) = \frac{\mathcal{Q}(u_1|w)\cdot\mathcal{P}(z_0|u_1)}{\mathcal{Q}(u_0|w)\cdot\mathcal{P}(z_0|u_0) + \mathcal{Q}(u_1|w)\cdot\mathcal{P}(z_0|u_1)} = \frac{q\cdot\delta_1}{(1-q)\cdot(1-\delta_0) + q\cdot\delta_1} =: p_0(q), \quad (43)$$

$$\mathcal{Q}(u_1|w,z_1) = \frac{\mathcal{Q}(u_1|w)\cdot\mathcal{P}(z_1|u_1)}{\mathcal{Q}(u_0|w)\cdot\mathcal{P}(z_1|u_0) + \mathcal{Q}(u_1|w)\cdot\mathcal{P}(z_1|u_1)} = \frac{q\cdot(1-\delta_1)}{(1-q)\cdot\delta_0 + q\cdot(1-\delta_1)} =: p_1(q). \quad (44)$$

The posterior beliefs after receiving the side information z_0 or z_1 are related to the *interim belief* $q \in [0,1]$ through the two functions $p_0(q)$, $p_1(q)$, depicted on Fig. 6.

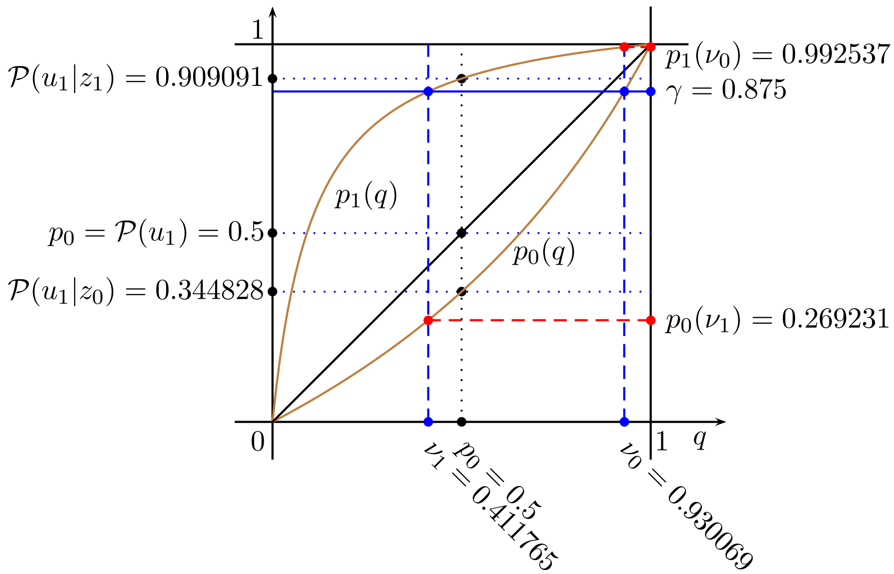

Fig. 6. The posterior beliefs functions $p_0(q)$ and $p_1(q)$ defined in (43) and (44), depending on the *interim belief* $q \in [0,1]$, for $p_0 = 0.5$, $\delta_1 = 0.05$, $\delta_2 = 0.5$ and $\gamma = 0.875$.

C. Encoder's average distortion function

Given the *belief threshold* $\gamma = \frac{7}{8}$ at which the decoder changes its symbol, we define the parameters ν_0 and ν_1 such that $p_0(\nu_0) = \gamma$ and $p_1(\nu_1) = \gamma$.

$$\gamma = p_0(\nu_0) \iff \nu_0 = \frac{\gamma \cdot (1-\delta_0)}{\delta_1 \cdot (1-\gamma) + \gamma \cdot (1-\delta_0)}, \tag{45}$$

$$\gamma = p_1(\nu_1) \iff \nu_1 = \frac{\gamma \cdot \delta_0}{\gamma \cdot \delta_0 + (1-\delta_1)\cdot(1-\gamma)}. \tag{46}$$

Remark IV.1 We have the equivalence $\nu_1 < \nu_0 \iff \delta_0 + \delta_1 < 1$.

Since the distortion functions of Fig. 3, 4 do not depend on the side information z, we denote by $\psi_{\mathsf{e}}(p)$ the *robust distortion function* of Definition III.5, given by

$$\psi_{\mathsf{e}}(p) = \min_{\substack{v \in \operatorname{argmin} \\ (1-p)\cdot d_{\mathsf{d}}(u_0,v) + p\cdot d_{\mathsf{d}}(u_1,v)}} (1-p)\cdot d_{\mathsf{e}}(u_0,v) + p \cdot d_{\mathsf{e}}(u_1,v), \tag{47}$$

$$= \mathbb{1}\big(p \le \gamma\big) \cdot \Big((1-p)\cdot d_{\mathsf{e}}(u_0,v_0) + p \cdot d_{\mathsf{e}}(u_1,v_0)\Big)$$

$$+ \mathbb{1}\big(p > \gamma\big) \cdot \Big((1-p)\cdot d_{\mathsf{e}}(u_0,v_1) + p \cdot d_{\mathsf{e}}(u_1,v_1)\Big) \tag{48}$$

$$= p \cdot \mathbb{1}\big(p \le \gamma\big) + (1-p)\cdot \mathbb{1}\big(p > \gamma\big). \tag{49}$$

Without loss of generality, we assume that $\delta_0 + \delta_1 < 1$, hence $\nu_1 < \nu_0$. The average distortion function $\Psi_{\mathsf{e}}(q)$ of Definition III.6 depends on the interim belief parameter $q \in [0, 1]$ as follows

$$\Psi_{\mathsf{e}}(q) = \mathbb{P}_q(z_0) \cdot \psi_{\mathsf{e}}\big(p_0(q)\big) + \mathbb{P}_q(z_1) \cdot \psi_{\mathsf{e}}\big(p_1(q)\big) \tag{50}$$

$$= \Big((1-q)\cdot(1-\delta_0) + q\cdot\delta_1\Big) \cdot \Big(p_0(q)\cdot \mathbb{1}\big(p_0(q) \le \gamma\big) + (1-p_0(q))\cdot \mathbb{1}\big(p_0(q) > \gamma\big)\Big)$$
$$+ \Big((1-q)\cdot\delta_0 + q\cdot(1-\delta_1)\Big) \cdot \Big(p_1(q)\cdot \mathbb{1}\big(p_1(q) \le \gamma\big) + (1-p_1(q))\cdot \mathbb{1}\big(p_1(q) > \gamma\big)\Big) \tag{51}$$

$$= q \cdot \mathbb{1}\big(q \le \nu_1\big) + \big(q\cdot\delta_1 + (1-q)\cdot\delta_0\big)\cdot \mathbb{1}\big(\nu_1 < q \le \nu_0\big) + (1-q)\cdot \mathbb{1}\big(q > \nu_0\big). \tag{52}$$

In Fig. 7, 8 and 11, the average distortion function $\Psi_{\mathsf{e}}(q)$ are represented by the orange lines, whereas the black curve is the average entropy $h(q)$ defined by

$$h(q) = H_b(q) + (1-q)\cdot H_b(\delta_0) + q\cdot H_b(\delta_1) - H_b\Big((1-q)\cdot\delta_0 + q\cdot(1-\delta_1)\Big). \tag{53}$$

D. Optimal splitting with three posteriors

Since the cardinality bound is $|\mathcal{W}| = \min\big(|\mathcal{U}|+1, |\mathcal{V}|^{|\mathcal{Z}|}\big) = 3$, we consider a splitting of the prior p_0 in three posteriors $(q_1, q_2, q_3) \in [0,1]^3$ with respective weights $(\lambda_1, \lambda_2, \lambda_3) \in [0,1]^3$, defined by (54), (55).

$$1 = \lambda_1 + \lambda_2 + \lambda_3, \tag{54}$$

$$p_0 = \lambda_1\cdot q_1 + \lambda_2\cdot q_2 + \lambda_3\cdot q_3, \tag{55}$$

$$H(U|Z) - C = \lambda_1\cdot h(q_1) + \lambda_2\cdot h(q_2) + \lambda_3\cdot h(q_3), \tag{56}$$

Equation (56) is satisfied when the information constraint is binding. By inverting the system (54)-(56), we obtain

$$\lambda_1 = \frac{\big(H(U|Z)-C\big)\cdot(q_2-q_3) + h(q_2)\cdot(q_3-p_0) + h(q_3)\cdot(p_0-q_2)}{h(q_1)\cdot(q_2-q_3) + h(q_2)\cdot(q_3-q_1) + h(q_3)\cdot(q_1-q_2)}, \tag{57}$$

$$\lambda_2 = \frac{\big(H(U|Z)-C\big)\cdot(q_3-q_1) + h(q_3)\cdot(q_1-p_0) + h(q_1)\cdot(p_0-q_3)}{h(q_1)\cdot(q_2-q_3) + h(q_2)\cdot(q_3-q_1) + h(q_3)\cdot(q_1-q_2)}, \tag{58}$$

$$\lambda_3 = \frac{\big(H(U|Z)-C\big)\cdot(q_1-q_2) + h(q_1)\cdot(q_2-p_0) + h(q_2)\cdot(p_0-q_1)}{h(q_1)\cdot(q_2-q_3) + h(q_2)\cdot(q_3-q_1) + h(q_3)\cdot(q_1-q_2)}. \tag{59}$$

The triple of posteriors (q_1, q_2, q_3) is feasible if and only if the weights $(\lambda_1, \lambda_2, \lambda_3)$ belong to the interval $[0,1]^3$. The average distortion $\Psi_{\mathsf{e}}(q)$ is piece-wise linear, hence the optimal triple of posteriors may belong to distinct intervals $q_1 \in [0, \nu_1)$, $q_2 \in [\nu_1, \nu_2)$, $q_3 \in [\nu_2, 1]$. Otherwise, take the average of two posteriors of the same interval which provides the same distortion value and has a larger entropy, due to the strict concavity of entropy function.

Aumann and Maschler's splitting lemma [70] or Kamenica and Gentzkow's Bayes plausibility [7] claim that the splitting $p_0 = \lambda_1 \cdot q_1 + \lambda_2 \cdot q_2 + \lambda_3 \cdot q_3$ is implemented by the following strategy

$$\mathcal{Q}(w_k|u_0) = \mathcal{Q}(w_k) \cdot \frac{1 - \mathcal{Q}(u_1|w_k)}{1 - \mathcal{P}(u_1)} = \lambda_k \cdot \frac{1 - q_k}{1 - p_0} =: \alpha_k, \qquad k \in \{, 1, 2, 3\} \tag{60}$$

$$\mathcal{Q}(w_k|u_1) = \mathcal{Q}(w_k) \cdot \frac{\mathcal{Q}(u_1|w_k)}{\mathcal{P}(u_1)} = \lambda_k \cdot \frac{q_k}{p_0} =: \beta_k, \qquad k \in \{, 1, 2, 3\}. \tag{61}$$

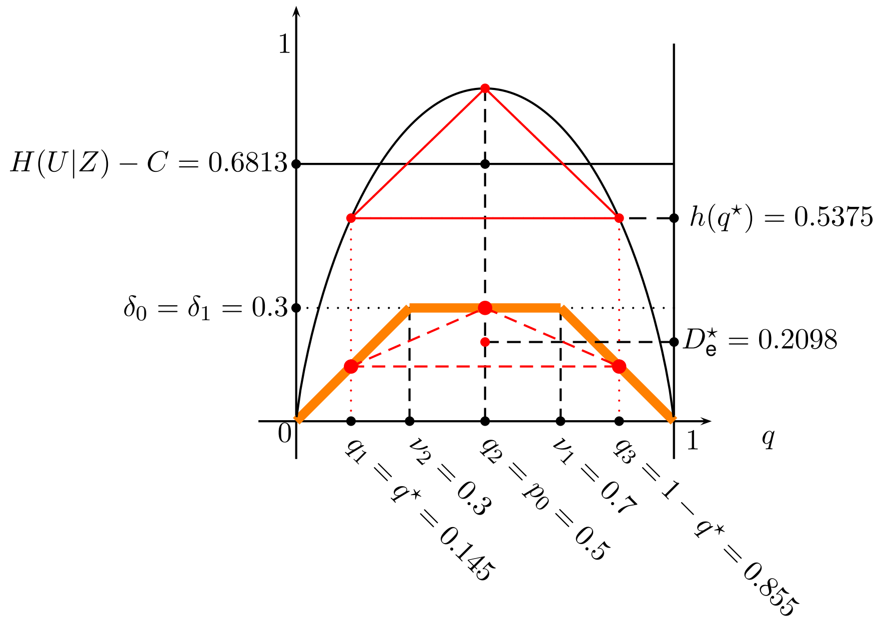

Fig. 7. Wyner-Ziv's example for DSBS. When $C \in [0, H(U|Z) - h(q^\star)]$ the optimal splitting is $(q_1, q_2, q_3) = (q^\star, \frac{1}{2}, 1 - q^\star)$. For the parameters $p_0 = 0.5$, $\delta_0 = \delta_1 = 0.3$, $C = 0.2$, $\kappa = 0$, the optimal encoder's distortion is $D_e^\star = 0.2098$.

E. Wyner-Ziv's example for DSBS with $p_0 = 0.5$, $\delta_0 = \delta_1 = 0.3$, $\kappa = 0$

We investigate the example of *doubly symmetric binary source* (DSBS) with $\delta_0 = \delta_1 = 0.3$ whose solution is characterized in [5, Sec. II, pp. 3]. In this example, both encoder and decoder minimize the Hamming distortion, hence $\kappa = 0 \Longleftrightarrow \gamma = \frac{1}{2}$. We introduce the notation $q \star \delta := (1 - q) \cdot \delta + q \cdot (1 - \delta)$, the average distortion and average entropy write

$$\Psi_e(q) = q \cdot \mathbb{1}\big(q \leq \delta\big) + \delta \cdot \mathbb{1}\big(\delta < q \leq 1 - \delta\big) + (1 - q) \cdot \mathbb{1}\big(q > 1 - \delta\big), \tag{62}$$

$$h(q) = H(U|Z) + H_b(q) - H_b\big(q \star \delta\big). \tag{63}$$

We remark that $H(U|Z) - h(q) = H_b\big(q \star \delta\big) - H_b(q)$.

Proposition IV.2 *We denote by $q^\star$ the unique solution of*

$$h'(q) = \frac{H(U|Z) - h(q)}{\delta - q}. \tag{64}$$

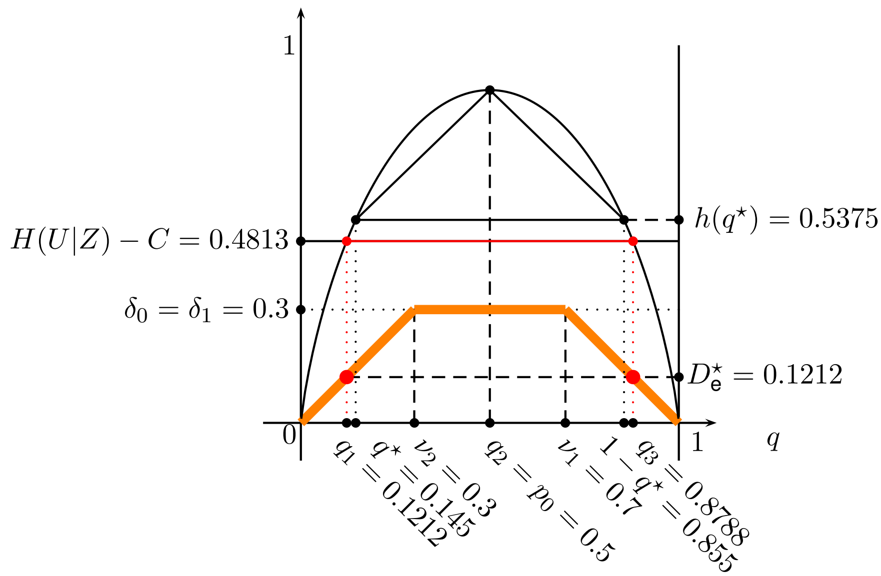

Fig. 8. Wyner-Ziv's example for DSBS. When $C \in [H(U|Z) - h(q^\star), H(U|Z)]$ the optimal splitting has two posteriors $(q_1, q_3) = \Big(h^{-1}\big(H(U|Z) - C\big), 1 - h^{-1}\big(H(U|Z) - C\big)\Big)$. For parameters $p_0 = 0.5$, $\delta_0 = \delta_1 = 0.3$, $C = 0.4$, $\kappa = 0$, the optimal encoder's distortion is $D_e^\star = 0.1212$.

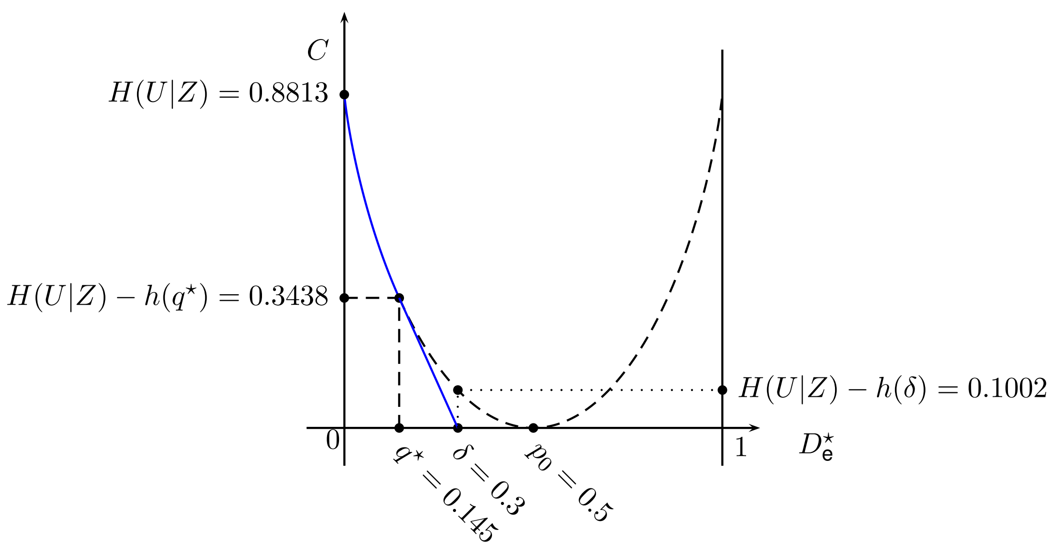

Fig. 9. Optimal trade-off between the capacity C and the optimal distortion $D_e^\star$ for the DSBS with parameters $p_0 = 0.5$, $\delta_0 = \delta_1 = 0.3$, $\kappa = 0$.

1) If $C \in [0, H(U|Z) - h(q^\star)]$, then the optimal splitting (Fig. 7) has three posterior beliefs

$q_1 = q^\star$	$q_2 = \frac{1}{2}$	$q_3 = 1 - q^\star$
$\lambda_1 = \frac{1}{2} \cdot \frac{C}{H(U|Z) - h(q^\star)}$	$\lambda_2 = 1 - \frac{C}{H(U|Z) - h(q^\star)}$	$\lambda_3 = \frac{1}{2} \cdot \frac{C}{H(U|Z) - h(q^\star)}$

corresponding to the strategies

$\alpha_1 = (1-q^\star)\cdot\frac{C}{H(U\|Z)-h(q^\star)}$	$\alpha_2 = 1-\frac{C}{H(U\|Z)-h(q^\star)}$	$\alpha_3 = q^\star\cdot\frac{C}{H(U\|Z)-h(q^\star)}$
$\beta_1 = q^\star\cdot\frac{C}{H(U\|Z)-h(q^\star)}$	$\beta_2 = 1-\frac{C}{H(U\|Z)-h(q^\star)}$	$\beta_3 = (1-q^\star)\cdot\frac{C}{H(U\|Z)-h(q^\star)}$

and the optimal distortion is

$$D_e^\star = \delta - C\cdot\frac{\delta-q^\star}{H(U|Z)-h(q^\star)} \tag{65}$$

2) If $C\in[H(U|Z)-h(q^\star),H(U|Z)]$, then the optimal splitting (Fig. 8) has two posterior beliefs

$q_1 = h^{-1}\big(H(U\|Z)-C\big)$	$q_2=\frac{1}{2}$	$q_3 = 1-h^{-1}\big(H(U\|Z)-C\big)$
$\lambda_1=\frac{1}{2}$	$\lambda_2 = 0$	$\lambda_3=\frac{1}{2}$

corresponding to the strategies

$\alpha_1 = 1-h^{-1}\big(H(U\|Z)-C\big)$	$\alpha_2=0$	$\alpha_3 = h^{-1}\big(H(U\|Z)-C\big)$
$\beta_1 = h^{-1}\big(H(U\|Z)-C\big)$	$\beta_2=0$	$\beta_3 = 1-h^{-1}\big(H(U\|Z)-C\big)$

and the optimal distortion is

$$D_e^\star = h^{-1}\big(H(U|Z)-C\big), \tag{66}$$

where the notation $h^{-1}\big(H(U|Z)-C\big)$ stands for the unique solution of equation $h(q)=H(U|Z)-C$.

3) If $C > H(U|Z)$, then the optimal splitting rely on the two extreme posterior beliefs $(0,1)$ and $D_e^\star = 0$.

The proof of Proposition IV.2 is given in the Appendix D. When $C\leq H(U|Z)-h(q^\star)$, the optimal strategy consists of a time-sharing between the operating point $(D_{\mathsf{e}}^\star, C) = \big(q^\star, H(U|Z)-h(q^\star)\big)$ and the zero rate point $(\delta,0)$, as depicted in Fig. 9.

F. Distinct distortions without side information, $p_0=0.5$, $\delta_0=0.5$, $\delta_1=0.5$, $C=0.2$, $\kappa=\frac{3}{4}$

We consider that the parameters $\delta_1=\delta_2=0.5$ so that the side information Z is independent of the source U. This corresponds to the problem studied in [4], when replacing the minimization by the maximization. We have $H_b(\delta_0)=H_b(\delta_1)=H_b\Big((1-q)\cdot\delta_0+q\cdot(1-\delta_1)\Big)=1$ and $\nu_1=\nu_2=\gamma=\frac{7}{8}$, the average entropy and average distortion write

$$h(q) = H_b(q), \tag{67}$$

$$\Psi_{\mathsf{e}}(q) = \psi_{\mathsf{e}}(q) = p\cdot\mathbb{1}\big(p\leq\gamma\big)+(1-p)\cdot\mathbb{1}\big(p>\gamma\big). \tag{68}$$

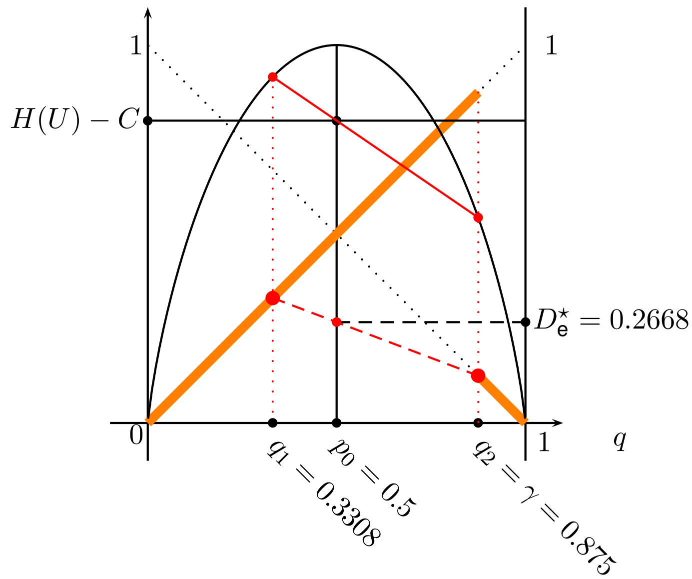

Fig. 10. For the parameters $p_0 = 0.5$, $\delta_1 = \delta_2 = 0.5$, $C = 0.2$, $\kappa = \frac{3}{4}$, the optimal encoder's distortion $D_{\mathsf{e}}^{\star} = 0.2668$.

Applying [4, Corollary 3.5, pp. 15], the optimal splitting has two posteriors $|\mathcal{W}| = 2$ and must satisfy the information constraint

$$\frac{p_0 - q_2}{q_1 - q_2} \cdot H_b(q_1) + \frac{q_1 - p_0}{q_1 - q_2} \cdot H_b(q_2) \geq H(U) - C. \tag{69}$$

The distortion function has two piece-wise linear components, hence the optimal splitting involves $q_1 \in [0, p_0]$ and $q_2 \in [\gamma, 1]$. For each $q_2 \in [\gamma, 1]$, we denote by $q_1(q_2)$ the function that returns the posterior which satisfies (69) with equality. From [4, Fig. 5, pp. 19], the function $q_1(q_2)$ is strictly increasing, hence its derivative $q_1'(q_2)$ is strictly positive. The encoder's distortion function reformulates in terms of q_2 as

$$\Phi_{\mathsf{e}}(q_2) = \frac{p_0 - q_2}{q_1(q_2) - q_2} \cdot q_1(q_2) + \frac{q_1 - p_0}{q_1(q_2) - q_2} \cdot (1 - q_2). \tag{70}$$

Its derivative writes

$$\Phi_{\mathsf{e}}'(q_2) = \frac{1}{(q_1(q_2) - q_2)^2} \cdot \bigg(q_1'(q_2) \cdot \Big(q_2 \cdot (2 \cdot q_2 - 1) + p_0 \Big) - (p_0 - q_1(q_2)) \cdot (1 - 2 \cdot q_1(q_2)) \bigg). \tag{71}$$

Since $q_2 \geq \gamma > \frac{1}{2}$, the sign of the derivative is negative if and only if

$$0 < q_1'(q_2) \leq \frac{(p_0 - q_1(q_2)) \cdot (1 - 2 \cdot q_1(q_2))}{q_2 \cdot (2 \cdot q_2 - 1) + p_0}. \tag{72}$$

By numerical optimization, the above inequality is satisfied for $p_0 = 0.5$, $\delta_1 = \delta_2 = 0.5$, $C = 0.2$, $\kappa = \frac{3}{4}$, hence the optimal distortion is achieved by using $q_2 = \gamma$, as depicted on Fig. 10.

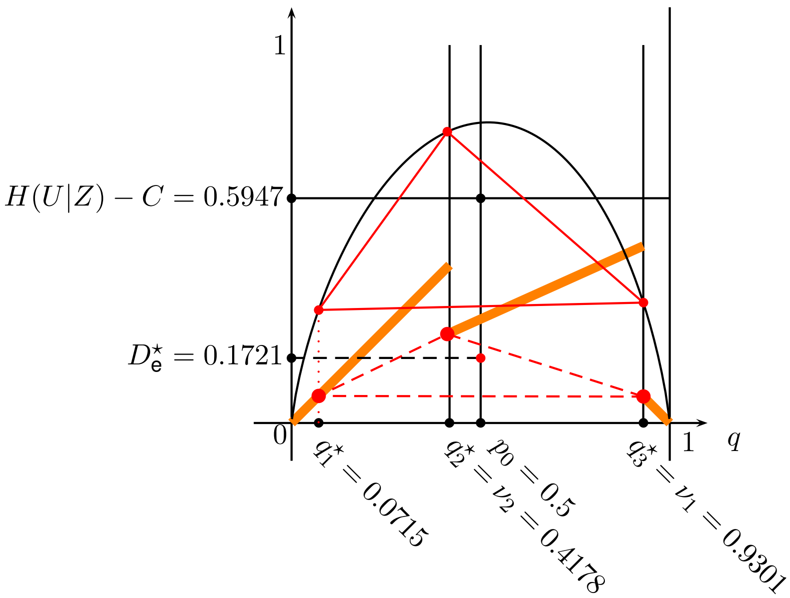

Fig. 11. For the parameters $p_0 = 0.5$, $\delta_0 = 0.05$, $\delta_1 = 0.5$, $C = 0.2$, $\kappa = \frac{3}{4}$, the optimal encoder's distortion is $D_{\mathsf{e}}^{\star} = 0.1721$.

G. Distinct distortions with side information, $p_0 = 0.5$, $\delta_0 = 0.05$, $\delta_1 = 0.5$, $C = 0.2$, $\kappa = \frac{3}{4}$

We consider an example with distinct distortion functions, i.e. with $\kappa = \frac{3}{4}$, with decoder's side information. By numerical simulation, we determine the optimal triple of posteriors (q_1, q_2, q_3) represented by the red dots in Fig. 11, that corresponds to the minimal distortion $D_{\mathsf{e}}^{\star} = 0.1721$.

$q_1 = 0.0715$	$q_2 = 0.4118$	$q_3 = 0.9301$
$\lambda_1 = 0.1288$	$\lambda_2 = 0.6165$	$\lambda_3 = 0.2548$

The parameters of the optimal strategy in Fig. 2, are given by

$\alpha_1 = 0.2392$	$\alpha_2 = 0.7252$	$\alpha_2 = 0.0356$
$\beta_1 = 0.0184$	$\beta_2 = 0.5077$	$\beta_3 = 0.4739$

Appendix A
Proof of Theorem III.7

We consider a joint probability distribution $\mathcal{Q}_{UW} \in \Delta(\mathcal{U} \times \mathcal{W})$, we identify the parameters $\lambda_w = \mathcal{Q}(w)$ and $p_w = \mathcal{Q}_U(\cdot|w) \in \Delta(\mathcal{U})$. The average distortion writes

$$\sum_w \lambda_w \cdot \Psi_{\mathsf{e}}(p_w) = \sum_w \mathcal{Q}(w) \cdot \Psi_{\mathsf{e}}\big(\mathcal{Q}_U(\cdot|w)\big) \tag{73}$$

$$= \sum_w \mathcal{Q}(w) \cdot \sum_{u,z} \mathcal{Q}(u|w) \cdot \mathcal{P}(z|u) \cdot \psi_{\mathsf{e}}\Big(z, \mathcal{Q}_U(\cdot|w,z)\Big) \tag{74}$$

$$= \sum_{w,z} \mathcal{Q}(w) \cdot \mathcal{Q}(z|w) \cdot \mathbb{E}_{\mathcal{Q}_U(\cdot|w,z)}\bigg[d_{\mathsf{e}}\Big(U, z, v^{\star}\big(z, \mathcal{Q}_U(\cdot|w,z)\big)\Big)\bigg] \tag{75}$$

$$= \mathbb{E}_{\mathcal{Q}_{UZW}}\bigg[d_{\mathsf{e}}\Big(U, Z, V^{\star}\big(Z, \mathcal{Q}_U(\cdot|W,Z)\big)\Big)\bigg] \tag{76}$$

$$= \max_{\substack{\mathcal{Q}_{V|ZW} \in \\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW} \\ \mathcal{Q}_{V|ZW}}}\bigg[d_{\mathsf{e}}(U, Z, V)\bigg]. \tag{77}$$

Equations (74), (75) and (77) come from Definitions III.6, III.5 and III.4.

Equations (73) and (76) are reformulations.

The average entropy writes

$$\sum_w \lambda_w \cdot h(p_w) = \sum_w \mathcal{Q}(w) \cdot h\big(\mathcal{Q}_U(\cdot|w)\big) \tag{78}$$

$$= \sum_w \mathcal{Q}(w) \cdot \bigg(H\big(\mathcal{Q}_U(\cdot|w)\big) + \sum_u \mathcal{Q}(u|w) \cdot H\Big(\mathcal{P}_Z(\cdot|u)\Big) - H\Big(\sum_u \mathcal{Q}(u|w) \cdot \mathcal{P}_Z(\cdot|u)\Big)\bigg) \tag{79}$$

$$= \sum_w \mathcal{Q}(w) \cdot \bigg(H\big(\mathcal{Q}_U(\cdot|w)\big) + \sum_u \mathcal{Q}(u|w) \cdot H\Big(\mathcal{Q}_Z(\cdot|u,w)\Big) - H\Big(\mathcal{Q}_Z(\cdot|w)\Big)\bigg) \tag{80}$$

$$= H(U|W) + H(Z|U,W) - H(Z|W) = H(U|W,Z). \tag{81}$$

Equation (79) come from Definition III.6.

Equation (80) come from Markov chain property $Z -\!\!\circ\!\!- U -\!\!\circ\!\!- W$ that implies $\mathcal{P}_Z(\cdot|u) = \mathcal{Q}_Z(\cdot|u,w)$ and $\mathcal{Q}_Z(\cdot|w) = \sum_u \mathcal{Q}(u|w) \cdot \mathcal{P}_Z(\cdot|u)$.

Equations (78) and (81) are reformulations.

Hence, equation (35) reformulates

$$\inf_{\substack{\lambda_w\in[0,1],\\ p_w\in\Delta(\mathcal{U})}} \Bigg\{ \sum_w \lambda_w \cdot \Psi_{\mathsf{e}}(p_w) \quad \text{s.t.} \quad \sum_w \lambda_w \cdot p_w = \mathcal{P}_U \in \Delta(\mathcal{U}),$$

$$\text{and} \quad \sum_w \lambda_w \cdot h(p_w) \geq H(U|Z) - \max_{\mathcal{P}_X} I(X;Y) \Bigg\} \tag{82}$$

$$= \inf_{\mathcal{Q}_W, \mathcal{Q}_{U|W}} \Bigg\{ \max_{\substack{\mathcal{Q}_{V|ZW}\in\\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW}\\ \mathcal{Q}_{V|ZW}}}\Big[d_{\mathsf{e}}(U,Z,V)\Big] \quad \text{s.t.} \quad \sum_w \mathcal{Q}(w)\cdot\mathcal{Q}_U(\cdot|w) = \mathcal{P}_U \in \Delta(\mathcal{U}),$$

$$\text{and} \quad H(U|W,Z) \geq H(U|Z) - \max_{\mathcal{P}_X} I(X;Y) \Bigg\} \tag{83}$$

$$= \inf_{\mathcal{Q}_{UZW}\in\mathbb{Q}_0} \max_{\substack{\mathcal{Q}_{V|ZW}\in\\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW}\\ \times\mathcal{Q}_{V|ZW}}}\Big[d_{\mathsf{e}}(U,Z,V)\Big] = D_{\mathsf{e}}^{\star}. \tag{84}$$

This concludes the proof of Theorem III.7.

Appendix B

Achievability proof of Theorem III.3

We refine the analysis of the Wyner-Ziv's source coding scheme [5], in order to control the posterior beliefs of a large fraction of stages, as stated in Proposition B.1 and in (96)-(97). Corollary B.3 shows that the best-reply strategy of the decoder performs similarly as the Wyner-Ziv's decoding scheme.

A. Zero capacity

We first investigate the special case of zero capacity.

Lemma 3 *If the channel has zero capacity* $\max_{\mathcal{P}_X} I(X;Y)=0$*, then we have:*

$$\forall n \in \mathbb{N}^{\star},\ \forall \sigma, \qquad \max_{\tau\in BR_d(\sigma)} d_e^n(\sigma,\tau) = D_e^{\star}. \tag{85}$$

Proof. [Lemma 3] When capacity is zero $\max_{\mathcal{P}_X} I(X;Y) = 0$, then the probability distribution $\mathcal{P}_{UZ}\mathcal{Q}_{W|U} \in \mathbb{Q}_0$ must satisfy $I(U;W|Z)=0$, hence the Markov chain property $U -\!\!\circ\!\!- Z -\!\!\circ\!\!- W$, i.e. $\mathcal{Q}_{U|ZW} = \mathcal{P}_{U|Z}$.

$$D_{\mathsf{e}}^{\star} = \inf_{\mathcal{Q}_{UZW}\in\mathbb{Q}_0} \max_{\substack{\mathcal{Q}_{V|WZ}\in\\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW}\\ \mathcal{Q}_{V|WZ}}}\Big[d_{\mathsf{e}}(U,Z,V)\Big] \tag{86}$$

$$= \inf_{\mathcal{Q}_{UZW}\in\mathbb{Q}_0} \mathbb{E}_{\mathcal{Q}_{UZW}}\bigg[d_{\mathsf{e}}\Big(U,Z,V^{\star}\big(Z,\mathcal{Q}_U(\cdot|W,Z)\big)\Big)\bigg] \tag{87}$$

$$= \inf_{\mathcal{Q}_{UZW}\in\mathbb{Q}_0} \mathbb{E}_{\mathcal{Q}_{UZW}}\bigg[d_{\mathsf{e}}\Big(U,Z,V^{\star}\big(Z,\mathcal{P}_U(\cdot|Z)\big)\Big)\bigg] \tag{88}$$

$$=\mathbb{E}_{\mathcal{P}_{UZ}}\bigg[d_{\mathsf{e}}\Big(U,Z,V^{\star}\big(Z,\mathcal{P}_U(\cdot|Z)\big)\Big)\bigg]. \tag{89}$$

Equation (87) is a reformulation by using the best-reply $v^{\star}\big(z,p\big)$ of Definition III.4 for symbol $z \in \mathcal{Z}$ and the belief $\mathcal{Q}_U(\cdot|w,z)$.

Equation (88) comes from Markov chain property $U -\!\!\!\circ\!\!\!- Z -\!\!\!\circ\!\!\!- W$ that allows to replace the belief $\mathcal{Q}_U(\cdot|w,z)$ by $\mathcal{P}_U(\cdot|z)$.

Equation (89) comes from removing the random variable W since it has no impact on the distortion function $d_{\mathsf{e}}\Big(u,z,v^{\star}\big(z,\mathcal{P}_U(\cdot|z)\big)\Big)$.

For any n and for any encoding strategy σ, the encoder's long-run distortion is given by

$$\max_{\tau\in\mathsf{BR}_{\mathsf{d}}(\sigma)} d_{\mathsf{e}}^n(\sigma,\tau) = \max_{\tau\in\mathsf{BR}_{\mathsf{d}}(\sigma)} \sum_{\substack{u^n,z^n,x^n,\\ y^n,v^n}} \prod_{t=1}^{n}\mathcal{P}\big(u_t,z_t\big)\sigma\big(x^n\big|u^n\big)\prod_{t=1}^{n}\mathcal{T}\big(y_t\big)\tau\big(v^n\big|y^n,z^n\big)\cdot\left[\frac{1}{n}\sum_{t=1}^{n}d_{\mathsf{e}}(u_t,z_t,v_t)\right] \tag{90}$$

$$= \max_{\tau\in\mathsf{BR}_{\mathsf{d}}(\sigma)} \sum_{u^n,z^n,v^n}\prod_{t=1}^{n}\mathcal{P}\big(u_t,z_t\big)\tau\big(v^n\big|z^n\big)\cdot\left[\frac{1}{n}\sum_{t=1}^{n}d_{\mathsf{e}}(u_t,z_t,v_t)\right] \tag{91}$$

$$= \frac{1}{n}\sum_{t=1}^{n}\left[\sum_{u_t,z_t,v_t}\mathcal{P}\big(u_t,z_t\big)\cdot\mathbb{1}\Big(v_t = v^{\star}\big(z_t,\mathcal{Q}_U(\cdot|z_t)\big)\Big)\cdot d_{\mathsf{e}}(u_t,z_t,v_t)\right] \tag{92}$$

$$= \mathbb{E}_{\mathcal{P}(u,z)}\left[d_{\mathsf{e}}\Big(U,Z,V^{\star}\big(Z,\mathcal{P}_U(\cdot|Z)\big)\Big)\right]. \tag{93}$$

Equation (90) comes from the zero capacity which imposes that the channel outputs Y^n are independent from the channel inputs X^n.

Equation (91) comes from removing the random variables (X^n,Y^n) and noting that the decoder's best-reply $\tau\big(v^n\big|z^n\big)$ does not depend on y^n anymore, since y^n is independent of (u^n,z^n).

Equation (92) is a reformulation based on the best-reply $v^{\star}\big(z,\mathcal{P}_U(\cdot|z)\big)$ of Definition III.4, for the symbol $z \in \mathcal{Z}$ and the belief $\mathcal{P}_U(\cdot|z)$.

Equation (93) comes from the i.i.d. property of (U,Z) and concludes the proof of Lemma 3. □

B. Positive capacity

We now assume that the channel capacity is strictly positive $\max_{\mathcal{P}_X} I(X;Y) > 0$. We define a specific convex closure in which the information constraint is satisfied with *strict* inequality and the sets of decoder's best-reply symbols are always *singletons*.

$$\widehat{D_{\mathsf{e}}} = \inf \Bigg\{ \sum_w \lambda_w \cdot \Psi_{\mathsf{e}}(p_w) \quad \text{s.t.} \quad \sum_w \lambda_w \cdot p_w = \mathcal{P}_U \in \Delta(\mathcal{U}), \\ \text{and} \quad \sum_w \lambda_w \cdot h(p_w) > H(U|Z) - \max_{\mathcal{P}_X} I(X;Y), \\ \text{and} \quad \forall (z,w) \in \mathcal{Z} \times \mathcal{W}, \;\; \mathcal{V}^\star\big(z, \mathcal{Q}_U(\cdot|z,w)\big) \;\text{ is a singleton } \Bigg\}. \tag{94}$$

Lemma 4 *If* $\max_{\mathcal{P}_X} I(X;Y) > 0$*, then* $\widehat{D_{\mathsf{e}}} = D^\star_{\mathsf{e}}$*.*

For the proof of Lemma 4, we refers to the similar proof of [4, Lemma A.5, pp. 32]. We denote by $\mathcal{P}^\star_X$ the probability distribution that maximizes the mutual information $I(X;Y)$, we denote by Q^n_{UZW} the empirical distribution of the sequence (u^n, z^n, w^n) and we denote by A_δ the set of typical sequences with tolerance $\delta > 0$, defined by

$$A_\delta = \bigg\{ (u^n, z^n, w^n, x^n, y^n), \quad \text{s.t.} \quad ||Q^n_{UZW} - \mathcal{P}_{UZ}\mathcal{Q}_{W|U}||_1 \leq \delta, \\ \text{and} \quad ||Q^n_{XY} - \mathcal{P}^\star_X \mathcal{T}_{Y|X}||_1 \leq \delta \bigg\}. \tag{95}$$

We denote by $\mathcal{P}_{\sigma,U_t}(\cdot|y^n,z^n) \in \Delta(\mathcal{U})$ the posterior belief induced by the strategy σ on U_t at stage t, given (y^n, z^n). We define $T_\alpha(w^n, y^n, z^n)$ and $B_{\alpha,\gamma,\delta}$ depending on parameters $\alpha > 0$ and $\gamma > 0$:

$$T_\alpha(w^n,y^n,z^n) = \bigg\{ t \in \{1,\ldots,n\}, \quad \text{s.t.} \quad D\Big(\mathcal{P}_{\sigma,U_t}(\cdot|y^n,z^n) \Big|\Big| \mathcal{Q}_{U_t}(\cdot|w_t,z_t)\Big) \leq \frac{\alpha^2}{2\ln 2} \bigg\}, \tag{96}$$

$$B_{\alpha,\gamma,\delta} = \bigg\{ (w^n,y^n,z^n), \quad \text{s.t.} \quad \frac{|T_\alpha(w^n,y^n,z^n)|}{n} \geq 1-\gamma \quad \text{and} \quad (w^n,y^n,z^n) \in A_\delta \bigg\}. \tag{97}$$

The notation $B^c_{\alpha,\gamma,\delta}$ stands for the complementary set of $B_{\alpha,\gamma,\delta} \subset \mathcal{W}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$. The sequences (w^n, y^n, z^n) belong to the set $B_{\alpha,\gamma,\delta}$ if, 1) they are typical, 2) the corresponding posterior belief $\mathcal{P}_{\sigma,U_t}(\cdot|y^n,z^n)$ is close in K-L divergence to the target belief $\mathcal{Q}_{U_t}(\cdot|w_t,z_t)$, for a large fraction of stages $t \in \{1,\ldots,n\}$.

The cornerstone of this achievability proof is Proposition B.1, which refines the analysis of Wyner-Ziv's source coding by controlling the posterior beliefs of a large fraction of stages.

Proposition B.1 (Wyner-Ziv's Posterior Beliefs) *If the probability distribution* $\mathcal{P}_{UZ}\mathcal{Q}_{W|U}$ *satisfies:*

$$\begin{cases} \max_{\mathcal{P}_X} I(X;Y) - I(U;W|Z) > 0, \\ \mathcal{V}^\star\big(z, \mathcal{Q}_U(\cdot|z,w)\big) \text{ is a singleton } \forall (z,w) \in \mathcal{Z} \times \mathcal{W}, \end{cases} \tag{98}$$

then

$$\forall \varepsilon > 0,\ \forall \alpha > 0,\ \forall \gamma > 0,\ \exists \bar{\delta} > 0,\ \forall \delta < \bar{\delta},\ \exists \bar{n} \in \mathbb{N}^\star,\ \forall n \geq \bar{n}, \exists \sigma,\ \text{s.t. } \mathcal{P}_\sigma(B^c_{\alpha,\gamma,\delta}) \leq \varepsilon. \tag{99}$$

The proof of proposition B.1 is stated in App. B-C.

Proposition B.2 *For any encoding strategy σ, we have:*

$$\left| \max_{\tau \in BR_d(\sigma)} d_e^n(\sigma,\tau) - \widehat{D_e} \right| \leq (\alpha + 2\gamma + \delta) \cdot \bar{\phi_e} + (1 - \mathcal{P}_\sigma(B_{\alpha,\gamma,\delta})) \cdot \bar{\phi_e}, \tag{100}$$

where $\bar{\phi_e} = \max_{u,z,v} \left|\phi_e(u,z,v)\right|$ is the largest absolute value of encoder's distortion.

For the proof of Proposition B.2, we refers directly to the similar proof of [4, Lemma A.8, pp. 33].

Corollary B.3 *For any $\varepsilon > 0$, there exists $\bar{n} \in \mathbb{N}^\star$ such that for all $n \geq \bar{n}$ there exists an encoding strategy σ such that:*

$$\left| \max_{\tau \in BR_d(\sigma)} d_e^n(\sigma,\tau) - \widehat{D_e} \right| \leq \varepsilon. \tag{101}$$

The proof of Corollary B.3 comes from combining Proposition B.1 with Proposition B.2 and choosing parameters α, γ, δ small and $n \in \mathbb{N}^\star$ large. The decoder's best-reply performs similarly as the Wyner-Ziv's decoding scheme. It concludes the achievability proof of Theorem III.3.

C. Proof of Proposition B.1

We assume that the probability distribution $\mathcal{P}_{UZ}\mathcal{Q}_{W|U}$ satisfies the two following conditions:

$$\begin{cases} \max_{\mathcal{P}_X} I(X;Y) - I(U;W|Z) > 0, \\ \mathcal{V}^\star\big(z, \mathcal{Q}_U(\cdot|z,w)\big) \text{ is a singleton } \forall (z,w) \in \mathcal{Z} \times \mathcal{W}, \end{cases} \tag{102}$$

The strict information constraint ensures that there exists a small parameter $\eta > 0$ and rates $\mathsf{R} \geq 0$, $\mathsf{R_L} \geq 0$, such that

$$\begin{aligned} \mathsf{R} + \mathsf{R_L} &= I(U;W) + \eta, & (103) \\ \mathsf{R_L} &\leq I(Z;W) - \eta, & (104) \\ \mathsf{R} &\leq \max_{\mathcal{P}_X} I(X;Y) - \eta. & (105) \end{aligned}$$

We now recall the random coding construction of Wyner-Ziv [5] and we investigate the corresponding posterior beliefs. We note by Σ the random coding scheme, described as follows.

- *Random codebook.* We introduces the indices $m \in \mathcal{M}$ with $|\mathcal{M}| = 2^{n\mathsf{R}}$ and $l \in \mathcal{M}_\mathsf{L}$ with $|\mathcal{M}_\mathsf{L}| = 2^{n\mathsf{R_L}}$. We draw $|\mathcal{M} \times \mathcal{M}_\mathsf{L}| = 2^{n(\mathsf{R}+\mathsf{R_L})}$ sequences $W^n(m,l)$ with the i.i.d. probability distribution $\mathcal{Q}_W^{\otimes n}$, and $|\mathcal{M}| = 2^{n\mathsf{R}}$ sequences $X^n(m)$, with the i.i.d. probability distribution $\mathcal{P}_X^{\star\otimes n}$ that maximizes the channel capacity in (105).

- *Encoding function.* The encoder observes the sequence of symbols of source $U^n \in \mathcal{U}^n$ and finds a pair of indices $(m,l) \in \mathcal{M} \times \mathcal{M}_{\mathsf{L}}$ such that the sequences $\big(U^n, W^n(m,l)\big) \in A_\delta$ are jointly typical. It sends the sequence $X^n(m)$ corresponding to the index $m \in \mathcal{M}$.
- *Decoding function.* The decoder observes the sequence of channel output $Y^n \in \mathcal{Y}^n$. It returns an index $\hat{m} \in \mathcal{M}$ such that the sequences $\big(Y^n, X^n(\hat{m})\big) \in A_\delta$ are jointly typical. Then it observes the sequence of side information $Z^n \in \mathcal{Z}^n$ and returns an index $\hat{l} \in \mathcal{M}_{\mathsf{L}}$ such that the sequences $\big(Z^n, W^n(\hat{m},\hat{l})\big) \in A_\delta$ are jointly typical.
- *Error Event.* We introduce the event of error $E_\delta \in \{0,1\}$ defined as follows:

$$E_\delta = \begin{cases} 0 & \text{if} \qquad (M,L) = (\hat{M},\hat{L}) \quad \text{and} \quad \big(U^n, Z^n, W^n, X^n, Y^n\big) \in A_\delta, \\ 1 & \text{otherwise.} \end{cases} \tag{106}$$

Expected error probability of the random encoding/decoding Σ. For all $\varepsilon_2 > 0$, for all $\eta > 0$, there exists a $\bar{\delta} > 0$, for all $\delta \leq \bar{\delta}$ there exists $\bar{n}$ such that for all $n \geq \bar{n}$, the expected probability of the following error events are bounded by ε_2:

$$\mathbb{E}_\Sigma\bigg[\mathcal{P}\bigg(\forall (m,l), \quad \big(U^n, W^n(m,l)\big) \notin A_\delta\bigg)\bigg] \leq \varepsilon_2, \tag{107}$$

$$\mathbb{E}_\Sigma\bigg[\mathcal{P}\bigg(\exists l' \neq l, \ \text{s.t.} \ \big(Z^n, W^n(m,l')\big) \in A_\delta\bigg)\bigg] \leq \varepsilon_2, \tag{108}$$

$$\mathbb{E}_\Sigma\bigg[\mathcal{P}\bigg(\exists m' \neq m, \ \text{s.t.} \ \big(Y^n, X^n(m')\big) \in A_\delta\bigg)\bigg] \leq \varepsilon_2, \tag{109}$$

Equation (107) comes from (103) and the covering lemma [69, pp. 208].
Equation (108) comes from (104) and the packing lemma [69, pp. 46].
Equation (109) comes from (105) and the packing lemma [69, pp. 46].

There exists a coding strategy σ with small error probability:

$$\forall \varepsilon_2 > 0, \ \forall \eta > 0, \ \exists \bar{\delta} > 0, \ \forall \delta \leq \bar{\delta}, \ \exists \bar{n} > 0, \ \forall n \geq \bar{n}, \quad \exists \sigma, \qquad \mathcal{P}_\sigma\big(E_\delta = 1\big) \leq \varepsilon_2. \tag{110}$$

Control of the posterior beliefs. We assume that the event $E_\delta = 0$ is realized. We denote by $\mathcal{P}_{\sigma,U_t}(\cdot|y^n, z^n, E_\delta = 0)$ the conditional probability distribution of U_t given $(y^n, z^n, E = 0)$, induced by the Wyner-Ziv's encoding strategy σ.

$$\mathbb{E}_\sigma\left[\frac{1}{n}\sum_{t=1}^{n} D\bigg(\mathcal{P}_{\sigma,U_t}(\cdot|Y^n,Z^n,E_\delta=0)\bigg|\bigg|\mathcal{Q}_{U_t}(\cdot|W_t,Z_t)\bigg)\right]$$

$$=\sum_{(w^n,z^n,y^n)\in A_\delta}\mathcal{P}_\sigma(w^n,z^n,y^n|E_\delta=0)\times\frac{1}{n}\sum_{t=1}^{n} D\bigg(\mathcal{P}_{\sigma,U_t}(\cdot|y^n,z^n,E_\delta=0)\bigg|\bigg|\mathcal{Q}_{U_t}(\cdot|w_t,z_t)\bigg) \tag{111}$$

$$=\frac{1}{n}\sum_{(u^n,z^n,w^n,y^n)\in A_\delta}\mathcal{P}_\sigma(u^n,z^n,w^n,y^n|E_\delta=0)\times\log_2\frac{1}{\prod_{t=1}^{n}\mathcal{Q}(u_t|w_t,z_t)}-\frac{1}{n}\sum_{t=1}^{n}H(U_t|Y^n,Z^n,E_\delta=0) \tag{112}$$

$$\leq H(U|W,Z)-\frac{1}{n}H(U^n|W^n,Y^n,Z^n,E_\delta=0)+\delta \tag{113}$$

$$\leq H(U|W,Z)-\frac{1}{n}H(U^n|W^n,Z^n,E_\delta=0)+\delta \tag{114}$$

$$=H(U|W,Z)-\frac{1}{n}H(U^n|E_\delta=0)+\frac{1}{n}I(U^n;W^n|E_\delta=0)$$
$$+\frac{1}{n}H(Z^n|W^n,E_\delta=0)-\frac{1}{n}H(Z^n|U^n,W^n,E_\delta=0)+\delta. \tag{115}$$

Equation (111)-(112) come from the hypothesis $E_\delta=0$ of typical sequences $(u^n,z^n,w^n,y^n)\in A_\delta$ and the definition of the conditional K-L divergence [72, pp. 24].

Equation (113) comes from property of typical sequences [69, pp. 26] and the conditioning that reduces entropy.

Equation (114) comes from the Markov chain $Z^n -\!\!\circ\!\!- U^n -\!\!\circ\!\!- W^n -\!\!\circ\!\!- Y^n$ induced by the channel and the strategy σ, that implies $H(U^n|W^n,Z^n,E_\delta=0)=H(U^n|W^n,Y^n,Z^n,E_\delta=0)$.

Equation (115) is a reformulation of (114).

We denote by $A_\delta(z^n|w^n)$ the set of sequences $z^n\in\mathcal{Z}^n$ that are jointly typical with w^n, i.e. $||Q^n_{WZ}-\mathcal{Q}_{WZ}||_1<\delta$.

$$\frac{1}{n}H(U^n|E_\delta=0)\geq H(U)-\frac{1}{n}-\log_2|\mathcal{U}|\cdot\mathcal{P}_\sigma\big(E_\delta=1\big), \tag{116}$$

$$\frac{1}{n}I(U^n;W^n|E_\delta=0)\leq \mathsf{R}+\mathsf{R_L}=I(U;W)+\eta, \tag{117}$$

$$\frac{1}{n}H(Z^n|W^n,E_\delta=0)\leq\frac{1}{n}\log_2|A_\delta(z^n|w^n)|\leq H(Z|W)+\delta, \tag{118}$$

$$\frac{1}{n}H(Z^n|U^n,W^n,E_\delta=0)\geq\frac{1}{n}H(Z^n|U^n,W^n)-\frac{1}{n}-\log_2|\mathcal{U}|\cdot\mathcal{P}_\sigma\big(E_\delta=1\big) \tag{119}$$

$$=H(Z|U,W)-\frac{1}{n}-\log_2|\mathcal{U}|\cdot\mathcal{P}_\sigma\big(E_\delta=1\big). \tag{120}$$

Equation (116) comes from the i.i.d. source and Fano's inequality.

Equation (117) comes from the cardinality of codebook given by (103). This argument is also used in [73, Eq. (23)].

Equation (118) comes from the cardinality of $A_\delta(z^n|w^n)$, see also [69, pp. 27].

Equation (119) comes from Fano's inequality.

Equation (119) comes from $H(Z^n|U^n, W^n) = H(Z^n|U^n) = H(Z|U) = H(Z|U, W)$ due to the Markov chain $Z^n -\!\!\!\ominus\!\!\!- U^n -\!\!\!\ominus\!\!\!- W^n$ of the encoding σ, the i.i.d. property of the source (U, Z), the Markov chain $Z -\!\!\!\ominus\!\!\!- U -\!\!\!\ominus\!\!\!- W$ of the single-letter characterization $\mathcal{Q}_{UZW} \in \mathbb{Q}_0$.

Equations (115)-(119) shows that on average, the posterior beliefs $\mathcal{P}_{\sigma,U_t}(\cdot|y^n, z^n, E_\delta = 0)$ induced by strategy σ is close to the target probability distribution $\mathcal{Q}_U(\cdot|w, z)$.

$$\begin{aligned}&\mathbb{E}_\sigma\left[\frac{1}{n}\sum_{t=1}^{n} D\bigg(\mathcal{P}_{\sigma,U_t}(\cdot|Y^n, Z^n, E_\delta = 0)\bigg|\bigg|\mathcal{Q}_{U_t}(\cdot|W_t, Z_t)\bigg)\right]\\ \leq& 2\delta + \eta + \frac{2}{n} + 2\log_2|\mathcal{U}|\cdot\mathcal{P}_\sigma\big(E_\delta = 1\big) := \epsilon. \end{aligned}\tag{121}$$

Then we have:

$$\begin{aligned}\mathcal{P}_\sigma(B^c_{\alpha,\gamma,\delta}) =& 1 - \mathcal{P}_\sigma(B_{\alpha,\gamma,\delta})\\ =& \mathcal{P}_\sigma(E_\delta = 1)\mathcal{P}_\sigma(B^c_{\alpha,\gamma,\delta}|E_\delta = 1) + \mathcal{P}_\sigma(E_\delta = 0)\mathcal{P}_\sigma(B^c_{\alpha,\gamma,\delta}|E_\delta = 0)\\ \leq& \mathcal{P}_\sigma(E_\delta = 1) + \mathcal{P}_\sigma(B^c_{\alpha,\gamma,\delta}|E_\delta = 0)\\ \leq& \varepsilon_2 + \mathcal{P}_\sigma(B^c_{\alpha,\gamma,\delta}|E_\delta = 0). \end{aligned}\tag{122}$$

Moreover:

$$\mathcal{P}_\sigma(B^c_{\alpha,\gamma,\delta}|E_\delta = 0)$$

$$= \sum_{w^n,y^n,z^n} \mathcal{P}_\sigma\Big((w^n, y^n, z^n) \in B^c_{\alpha,\gamma,\delta}\Big|E_\delta = 0\Big) \tag{123}$$

$$= \sum_{w^n,y^n,z^n} \mathcal{P}_\sigma\left((w^n, y^n, z^n) \quad \text{s.t.} \quad \frac{|T_\alpha(w^n, y^n, z^n)|}{n} < 1 - \gamma\middle|E_\delta = 0\right) \tag{124}$$

$$= \mathcal{P}_\sigma\left(\frac{1}{n}\cdot\left|\left\{t, \text{ s.t. } D\Big(\mathcal{P}_{\sigma,U_t}(\cdot|y^n, z^n)\Big|\Big|\mathcal{Q}_{U_t}(\cdot|w_t, z_t)\Big) \leq \frac{\alpha^2}{2\ln 2}\right\}\right| < 1 - \gamma\middle|E_\delta = 0\right) \tag{125}$$

$$= \mathcal{P}_\sigma\left(\frac{1}{n}\cdot\left|\left\{t, \text{ s.t. } D\Big(\mathcal{P}_{\sigma,U_t}(\cdot|y^n, z^n)\Big|\Big|\mathcal{Q}_{U_t}(\cdot|w_t, z_t)\Big) > \frac{\alpha^2}{2\ln 2}\right\}\right| \geq \gamma\middle|E_\delta = 0\right) \tag{126}$$

$$\leq \frac{2\ln 2}{\alpha^2\gamma}\cdot\mathbb{E}_\sigma\left[\frac{1}{n}\sum_{t=1}^{n} D\Big(\mathcal{P}_{\sigma,U_t}(\cdot|y^n, z^n)\Big|\Big|\mathcal{Q}_{U_t}(\cdot|w_t, z_t)\Big)\right] \tag{127}$$

$$\leq \frac{2\ln 2}{\alpha^2\gamma}\cdot\bigg(\eta + \delta + \frac{2}{n} + 2\log_2|\mathcal{U}|\cdot\mathcal{P}_\sigma\big(E_\delta = 1\big)\bigg). \tag{128}$$

Equation (123) to (126) are simple reformulations.

Equation (127) comes from the double use of Markov's inequality as in [4, Lemma A.21, pp. 42].

Equation (128) comes from (121).

Combining equations (110), (122), (128) and choosing $\eta > 0$ small, we obtain the following statement:

$$\forall \varepsilon > 0,\ \forall \alpha > 0,\ \forall \gamma > 0,\ \exists \bar{\delta} > 0,\ \forall \delta < \bar{\delta},\ \exists \bar{n} \in \mathbb{N}^{\star},\ \forall n \geq \bar{n}, \exists \sigma,\ \text{s.t. } \mathcal{P}_{\sigma}(B^{c}_{\alpha,\gamma,\delta}) \leq \varepsilon. \tag{129}$$

This concludes the proof of Proposition B.1.

Appendix C
Converse proof of Theorem III.3

We consider an encoding strategy σ of length $n \in \mathbb{N}^{\star}$. We denote by T the uniform random variable over $\{1,\ldots,n\}$ and let Z^{-T} stand for $(Z_1,\ldots,Z_{t-1},Z_{t+1},\ldots Z_n)$, where Z_T has been removed. We introduce the auxiliary random variable $W = (Y^n, Z^{-T}, T)$ whose joint probability distribution $\mathcal{P}_{UZW}$ is defined by

$$\begin{aligned}\mathcal{P}(u,z,w) =& \mathcal{P}_{\sigma}\big(u_T, z_T, y^n, z^{-T}, T\big)\\ =& \mathcal{P}(T=t)\cdot \mathcal{P}_{\sigma}\big(u_T, z_T, y^n, z^{-T}\big| T=t\big)\\ =& \frac{1}{n}\cdot \mathcal{P}_{\sigma}\big(u_t, z_t, y^n, z^{-t}\big), \quad \forall (u,w,z,u^n,z^n,y^n).\end{aligned} \tag{130}$$

This identification ensures that the Markov chain $W -\!\!\!\ominus\!\!\!- U_T -\!\!\!\ominus\!\!\!- Z_T$ is satisfied. Let us fix a decoding strategy $\tau_{V^n|Y^nZ^n}$ and define $\tilde{\tau}_{V|WZ} = \tilde{\tau}_{V|Y^nZ^{-T}TZ} = \tau_{V_T|Y^nZ^n}$. The encoder's long-run distortion writes:

$$d^n_{\mathsf{e}}(\sigma,\tau) = \sum_{u^n,z^n,y^n} \mathcal{P}_{\sigma}(u^n,z^n,y^n) \sum_{v^n} \tau(v^n|y^n,z^n)\cdot \left[\frac{1}{n}\sum_{t=1}^{n} d_{\mathsf{e}}(u_t,z_t,v_t)\right] \tag{131}$$

$$= \sum_{t=1}^{n} \sum_{\substack{u_t,z_t,\\ z^{-t},y^n}} \frac{1}{n}\cdot \mathcal{P}_{\sigma}(u_t,z^n,y^n) \sum_{v_t} \tau(v_t|y^n,z^n)\cdot d_{\mathsf{e}}(u_t,z_t,v_t) \tag{132}$$

$$= \sum_{\substack{u_t,z_t,y^n,\\ z^{-t},t}} \mathcal{P}_{\sigma}(u_t,z_t,y^n,z^{-t},t) \sum_{v_t} \tau(v_t|z_t,y^n,z^{-t},t)\cdot d_{\mathsf{e}}(u_t,z_t,v_t) \tag{133}$$

$$= \sum_{u,z,w} \mathcal{P}(u,z,w) \sum_{v} \tilde{\tau}(v|w,z)\cdot d_{\mathsf{e}}(u,z,v). \tag{134}$$

Equations (131) - (133) are reformulations and re-orderings.

Equation (134) comes from replacing the random variables (Y^n, Z^{-T}, T) by W whose distribution is defined in (130).

Equations (131) - (134) are also valid for the decoder's distortion $d_{\mathsf{d}}^n(\sigma,\tau) = \sum_{\substack{u,z,\\ w,v}} \mathcal{P}(u,z,w)\tilde{\tau}(v|w,z)\cdot d_{\mathsf{d}}(u,z,v)$. A best-reply strategy $\tau \in \mathsf{BR}_{\mathsf{d}}(\sigma)$ reformulates as:

$$\tau \in \mathrm{argmin}_{\tau'_{V^n|Y^nZ^n}} \sum_{\substack{u^n,z^n,\\ x^n,y^n,v^n}} \mathcal{P}_\sigma(u^n,z^n,x^n,y^n)\cdot\tau'(v^n|y^n,z^n)\cdot\left[\frac{1}{n}\sum_{t=1}^n \phi_{\mathsf{d}}(u_t,z_t,v_t)\right] \tag{135}$$

$$\Longleftrightarrow \tilde{\tau}_{V|WZ} \in \mathrm{argmin}_{\tilde{\tau}'_{V|WZ}} \sum_{u,z,w} \mathcal{P}(u,z,w)\cdot\tilde{\tau}'(v|w,z)\cdot\phi_{\mathsf{d}}(u,z,v) \tag{136}$$

$$\Longleftrightarrow \tilde{\tau}_{V|WZ} \in \mathbb{Q}_2\big(\mathcal{P}_{UZW}\big). \tag{137}$$

We now prove that the distribution $\mathcal{P}_{UZW}$ defined in (130), satisfies the information constraint of the set $\mathbb{Q}_0$.

$$0 \leq I(X^n;Y^n) - I(U^n,Z^n;Y^n) \tag{138}$$

$$\leq \sum_{t=1}^n H(Y_t) - \sum_{t=1}^n H(Y_t|X_t) - I(U^n;Y^n|Z^n) \tag{139}$$

$$\leq n\cdot\max_{\mathcal{P}_X} I(X;Y) - \sum_{t=1}^n I(U_t;Y^n|Z^n,U^{t-1}) \tag{140}$$

$$= n\cdot\max_{\mathcal{P}_X} I(X;Y) - \sum_{t=1}^n I(U_t;Y^n,Z^{-t},U^{t-1}|Z_t) \tag{141}$$

$$\leq n\cdot\max_{\mathcal{P}_X} I(X;Y) - \sum_{t=1}^n I(U_t;Y^n,Z^{-t}|Z_t) \tag{142}$$

$$= n\cdot\max_{\mathcal{P}_X} I(X;Y) - n\cdot I(U_T;Y^n,Z^{-T}|Z_T,T) \tag{143}$$

$$= n\cdot\max_{\mathcal{P}_X} I(X;Y) - n\cdot I(U_T;Y^n,Z^{-T},T|Z_T) \tag{144}$$

$$= n\cdot\max_{\mathcal{P}_X} I(X;Y) - n\cdot I(U;W|Z) \tag{145}$$

$$= n\cdot\left(\max_{\mathcal{P}_X} I(X;Y) - I(U;W) + I(Z;W)\right). \tag{146}$$

Equation (138) comes from the Markov chain $Y^n -\!\!\circ\!\!- X^n -\!\!\circ\!\!- (U^n,Z^n)$.

Equation (139) comes from the memoryless property of the channel and from removing the positive term $I(U^n;Z^n)\geq 0$.

Equation (140) comes from taking the maximum over $\mathcal{P}_X$ and the chain rule.

Equation (141) comes from the i.i.d. property of the source (U,Z) which implies $I(U_t,Z_t;Z^{-t},U^{t-1}) = I(U_t;Z^{-t},U^{t-1}|Z_t) = 0$.

Equation (142) comes from removing $I(U_t;U^{t-1}|Y^n,Z^{-t},Z_t)\geq 0$.

Equation (143) comes from the uniform random variable $T\in\{1,\dots,n\}$.

Equation (144) comes from the independence between T and the source (U,Z), which implies

$I(U_T, Z_T; T) = I(U_T; T|Z_T) = 0$.

Equation (145) comes from the identification $W = (Y^n, Z^{-T}, T)$.

Equation (146) comes from the Markov chain $W -\!\!\ominus\!\!- U_T -\!\!\ominus\!\!- Z_T$. This proves that the distribution $\mathcal{P}_{\sigma,UZW}$ belongs to the set $\mathbb{Q}_0$.

Therefore, for any encoding strategy σ and all n, we have:

$$\max_{\tau \in \mathsf{BR}_{\mathsf{d}}(\sigma)} d_{\mathsf{e}}^n(\sigma, \tau) \tag{147}$$

$$= \max_{\substack{\tilde{\tau}_{V|WZ} \in \\ \mathbb{Q}_2(\mathcal{P}_{UZW})}} \sum_{u,z,w} \mathcal{P}(u, z, w) \sum_{v} \tilde{\tau}(v|w, z) \cdot d_{\mathsf{e}}(u, z, v) \tag{148}$$

$$= \max_{\substack{\tilde{\tau}_{V|WZ} \in \\ \mathbb{Q}_2(\mathcal{P}_{UZW})}} \mathbb{E}_{\substack{\mathcal{P}(u,z,w) \\ \tilde{\tau}_{V|WZ}}} \Big[d_{\mathsf{e}}(U, Z, V) \Big] \tag{149}$$

$$\geq \inf_{\mathcal{Q}_{UZW} \in \mathbb{Q}_0} \max_{\substack{\mathcal{Q}_{V|WZ} \in \\ \mathbb{Q}_2(\mathcal{Q}_{UZW})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW} \\ \mathcal{Q}_{V|WZ}}} \Big[d_{\mathsf{e}}(U, Z, V) \Big] = D_{\mathsf{e}}^{\star}. \tag{150}$$

The last inequality comes from the probability distribution $\mathcal{P}_{UZW}$ satisfying the information constraint of the set $\mathbb{Q}_0$.

The proof for the cardinality bound $|\mathcal{W}| = \min(|\mathcal{U}| + 1, |\mathcal{V}|^{|\mathcal{Z}|})$ follows two arguments. The bound $|\mathcal{W}| = |\mathcal{U}| + 1$ comes [71, Corollary 17.1.5, pp. 157], also in [4, Corollary A.2, pp. 26]. The bound $|\mathcal{W}| = |\mathcal{V}|^{|\mathcal{Z}|}$ comes from assuming that the encoder tells to the decoder to select a function from the side information $\mathcal{Z}$ to the symbols $\mathcal{V}$, as discussed in Sec. III-D.

This concludes the proof of (19) in Theorem III.3.

Appendix D
Proof of Proposition IV.2

The average distortion $\Psi_{\mathsf{e}}(q)$ defined in (52) is piece-wise linear, hence the optimal triple of posteriors may belong to distinct intervals $q_1 \in [0, \nu_1)$, $q_2 \in [\nu_1, \nu_2)$, $q_3 \in [\nu_2, 1]$. Since $\delta_0 = \delta_1$ and $\kappa = 0$, the function $\Psi_{\mathsf{e}}(q)$ is symmetric and constant over the interval $[\nu_1, \nu_2]$, as depicted in Fig 7. The optimal splitting must satisfy $(q_1, q_2, q_3) = (q_1, \frac{1}{2}, 1 - q_1)$, since $q_2 = \frac{1}{2}$ provides the highest entropy $h(q_2) = H(U|Z)$. Equations (57)-(59) reformulate as

$$\lambda_1 = \frac{1}{2} \cdot \frac{C}{H(U|Z) - h(q_1)}, \tag{151}$$

$$\lambda_2 = 1 - \frac{C}{H(U|Z) - h(q_1)}, \tag{152}$$

$$\lambda_3 = \frac{1}{2} \cdot \frac{C}{H(U|Z) - h(q_1)}. \tag{153}$$

We examine the feasibility conditions, i.e. $(\lambda_1,\lambda_2,\lambda_3) \in [0,1]^3$. We recall that the notation $h^{-1}\big(H(U|Z)-C\big)$ stands for the unique solution $q\in[0,\frac{1}{2}]$ of the equation $h(q) = H(U|Z)-C$. Since $H(U|Z)\geq h(q_1)$ for all q_1, we have $\lambda_1\geq 0$, $\lambda_2\leq 1$ and $\lambda_3\geq 0$, moreover

$$\lambda_1 \leq 1 \Longleftrightarrow h(q_1) \leq H(U|Z) - \frac{1}{2}\cdot C \Longleftrightarrow q_1 \leq h^{-1}\Big(H(U|Z) - \frac{1}{2}\cdot C\Big) \tag{154}$$

$$\lambda_2 \geq 0 \Longleftrightarrow h(q_1) \leq H(U|Z) - C \Longleftrightarrow q_1 \leq h^{-1}\Big(H(U|Z) - C\Big). \tag{155}$$

Since the function h^{-1} is increasing over the interval $[0,1]$, we have $h^{-1}\Big(H(U|Z)-\frac{1}{2}\cdot C\Big) \leq h^{-1}\Big(H(U|Z)-C\Big)$. This proves the following Lemma.

Lemma 5 *The splitting $(q_1,q_2,q_3) = (q_1,\frac{1}{2},1-q_1)$ is feasible if and only if*

$$q_1 \leq h^{-1}\Big(H(U|Z)-C\Big). \tag{156}$$

We assume that $(q_1,q_2,q_3) = (q_1,\frac{1}{2},1-q_1)$, we define the *encoder's distortion function* by

$$\begin{aligned}
\Phi_{\mathsf{e}}(q_1) =& \lambda_1\Psi_{\mathsf{e}}(q_1) + \lambda_2\Psi_{\mathsf{e}}\Big(\frac{1}{2}\Big) + \lambda_3\Psi_{\mathsf{e}}(1-q_1) & (157)\\
=&\frac{1}{2}\cdot\frac{C}{H(U|Z)-h(q_1)}\cdot q_1 + \bigg(1-\frac{C}{H(U|Z)-h(q_1)}\bigg)\cdot\delta + \frac{1}{2}\cdot\frac{C}{H(U|Z)-h(q_1)}\cdot q_1 & (158)\\
=&\delta - \frac{(\delta-q_1)\cdot C}{H(U|Z)-h(q_1)}, & (159)
\end{aligned}$$

Its derivative is

$$\Phi'_{\mathsf{e}}(q_1) = \frac{C}{(H(U|Z)-h(q_1))^2}\cdot\bigg(H(U|Z)-h(q_1)-h'(q_1)\cdot(\delta-q_1)\bigg), \tag{160}$$

where the derivative of the entropy $h'(q)$ writes

$$h'(q) = \log_2\frac{1-q}{q} - (1-2\cdot\delta)\cdot\log_2\frac{1-q\star\delta}{q\star\delta}. \tag{161}$$

We examine the term $k(q) = \Big(H(U|Z)-h(q)-h'(q)\cdot(\delta-q)\Big)$ of the function $\Phi'_{\mathsf{e}}(q)$. We have $\lim_{q\to 0} k(q) = -\infty$ since $\lim_{q\to 0} h'(q) = +\infty$ and $k(\delta) = H(U|Z)-h(\delta) = H_b(\delta\star\delta) - H_b(\delta) > 0$ since $\delta<\frac{1}{2}$. The derivative $k'(q) = -h''(q)\cdot(\delta-q) > 0$ is strictly positive because $\delta > q$ and the entropy is strictly concave $h''(q)<0$. Hence the equation $k(q)=0$ has a unique solution $q^\star\in(0,\delta[$.

The derivative $\Phi'_{\mathsf{e}}(q)$ is non-positive on the interval $q\in(0,q^\star]$ and non-negative on the interval $q\in[q^\star,\delta)$, hence the distortion $\Phi_{\mathsf{e}}(q)$ reaches its minimum in $q^\star$.

1) If $q^\star \leq h^{-1}\Big(H(U|Z)-C\Big)$, then the optimal splitting is

$q_1 = q^\star$	$q_2 = \frac{1}{2}$	$q_3 = 1-q^\star$
$\lambda_1 = \frac{1}{2}\cdot\frac{C}{H(U\|Z)-h(q^\star)}$	$\lambda_2 = 1-\frac{C}{H(U\|Z)-h(q^\star)}$	$\lambda_3 = \frac{1}{2}\cdot\frac{C}{H(U\|Z)-h(q^\star)}$

corresponding to the strategies

$\alpha_1 = (1-q^\star)\cdot\frac{C}{H(U\|Z)-h(q^\star)}$	$\alpha_2 = 1-\frac{C}{H(U\|Z)-h(q^\star)}$	$\alpha_3 = q^\star\cdot\frac{C}{H(U\|Z)-h(q^\star)}$
$\beta_1 = q^\star\cdot\frac{C}{H(U\|Z)-h(q^\star)}$	$\beta_2 = 1-\frac{C}{H(U\|Z)-h(q^\star)}$	$\beta_3 = (1-q^\star)\cdot\frac{C}{H(U\|Z)-h(q^\star)}$

and the optimal distortion is

$$D_{\mathsf{e}}^\star = \delta - C\cdot\frac{\delta - q^\star}{H(U|Z)-h(q^\star)}. \tag{162}$$

2) If $q^\star > h^{-1}\Big(H(U|Z)-C\Big)$, then the optimal splitting is

$q_1 = h^{-1}\big(H(U\|Z)-C\big)$	$q_2 = \frac{1}{2}$	$q_3 = 1-h^{-1}\big(H(U\|Z)-C\big)$
$\lambda_1 = \frac{1}{2}$	$\lambda_2 = 0$	$\lambda_3 = \frac{1}{2}$

corresponding to the strategies

$\alpha_1 = 1-h^{-1}\big(H(U\|Z)-C\big)$	$\alpha_2 = 0$	$\alpha_3 = h^{-1}\big(H(U\|Z)-C\big)$
$\beta_1 = h^{-1}\big(H(U\|Z)-C\big)$	$\beta_2 = 0$	$\beta_3 = 1-h^{-1}\big(H(U\|Z)-C\big)$

and the optimal distortion is

$$D_{\mathsf{e}}^\star = h^{-1}\big(H(U|Z)-C\big). \tag{163}$$

3) If $C > H(U|Z)$, then the extreme splitting $(q_1, q_3) = (0, 1)$ is feasible and $D_{\mathsf{e}}^\star = 0$.

This concludes the proof of Proposition IV.2.

References

[1] M. Le Treust and T. Tomala, "Information design for strategic coordination of autonomous devices with non-aligned utilities," in *54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 233–242, Sept. 2016.

[2] ——, "Persuasion bayésienne pour la coordination stratégique d'appareils autonomes ayant des objectifs non-alignés," in *Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI17), Juan-les-Pins, France*, 2017.

[3] ——, "Strategic coordination with state information at the decoder," *Proc. of 2018 International Zurich Seminar on Information and Communication*, 2018.

[4] ——, "Persuasion with limited communication capacity," *Journal of Economic Theory*, vol. 184, p. 104940, 2019.

[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–11, Jan. 1976.

[6] N. Merhav and S. Shamai, "On joint source-channel coding for the Wyner-Ziv source and the Gel'fand-Pinsker channel," *IEEE Transactions on Information Theory*, vol. 49, no. 11, pp. 2844–2855, 2003.

[7] E. Kamenica and M. Gentzkow, “Bayesian persuasion,” *American Economic Review*, vol. 101, pp. 2590–2615, 2011.

[8] E. Akyol, C. Langbort, and T. Başar, “Strategic compression and transmission of information,” in *IEEE Information Theory Workshop - Fall (ITW)*, Oct 2015, pp. 219–223.

[9] ——, “On the role of side information in strategic communication,” in *IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1626–1630.

[10] ——, “Information-theoretic approach to strategic communication as a hierarchical game,” *Proceedings of the IEEE*, vol. 105, no. 2, pp. 205–218, 2017.

[11] V. P. Crawford and J. Sobel, “Strategic information transmission,” *Econometrica*, vol. 50, no. 6, pp. 1431–1451, 1982.

[12] V. S. S. Nadendla, C. Langbort, and T. Başar, “Effects of subjective biases on strategic information transmission,” *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6040–6049, Dec. 2018.

[13] S. Sarıtaş, S. Yüksel, and S. Gezici, “Dynamic signaling games with quadratic criteria under nash and stackelberg equilibria,” *preprint available [on-line] https://arxiv.org/abs/1704.03816*, 2017.

[14] ——, “Quadratic multi-dimensional signaling games and affine equilibria,” *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 605–619, Feb. 2017.

[15] S. Sarıtaş, S. Gezici, and S. Yüksel, “Hypothesis testing under subjective priors and costs as a signaling game,” *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 5169–5183, Oct. 2019.

[16] F. Farokhi, A. M. H. Teixeira, and C. Langbort, “Estimation with strategic sensors,” *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 724–739, Feb. 2017.

[17] J. Y. Y. Jakub Mareček, Robert Shorten, “Signaling and obfuscation for congestion control,” *ACM SIGecom Exchanges*, vol. 88, no. 10, pp. 2086–2096, 2015.

[18] H. Tavafoghi and D. Teneketzis, “Informational incentives for congestion games,” in *55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1285–1292, Oct. 2017.

[19] S. Dughmi and H. Xu, “Algorithmic bayesian persuasion,” *In Proceedings of the 47th ACM Symposium on Theory of Computing (STOC)*, 2016.

[20] S. Dughmi, “Algorithmic information structure design: A survey,” *ACM SIGecom Exchanges*, vol. 15, no. 2, pp. 2–24, Jan 2017.

[21] S. Dughmi, D. Kempe, and R. Qiang, “Persuasion with limited communication,” *In Proceedings of the 17th ACM conference on Economics and Computation (ACM EC’17), Maastricht, The Netherlands*, 2016.

[22] R. Berry and D. Tse, “Shannon meets Nash on the interference channel,” *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2821–2836, May 2011.

[23] S. M. Perlaza, R. Tandon, H. V. Poor, and Z. Han, “Perfect output feedback in the two-user decentralized interference channel,” *IEEE Transactions on Information Theory*, vol. 61, no. 10, pp. 5441–5462, Oct. 2015.

[24] A. E. Gamal and T. Cover, “Achievable rates for multiple descriptions,” *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.

[25] A. Lapidoth, A. Malär, and M. Wigger, “Constrained source-coding with side information,” *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3218–3237, June 2014.

[26] A. Dembo and T. Weissman, “The minimax distortion redundancy in noisy source coding,” *IEEE Transactions on Information Theory*, vol. 49, no. 11, pp. 3020–3030, Nov. 2003.

[27] R. Timo, A. Grant, and G. Kramer, “Lossy broadcasting with complementary side information,” *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 104–131, Jan. 2013.

[28] H. Yamamoto, "A rate-distortion problem for a communication system with a secondary decoder to be hindered," *IEEE Transactions on Information Theory*, vol. 34, no. 4, pp. 835–842, July 1988.

[29] H. Yamamoto, "Rate-distortion theory for the shannon cipher system," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 827–835, May 1997.

[30] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7584–7605, Dec 2014.

[31] ——, "The henchman problem: Measuring secrecy by the minimum distortion in a list," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3436–3450, June 2016.

[32] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Transactions on Information Theory*, vol. 43, no. 1, pp. 38–47, Jan. 1997.

[33] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai Shitz, "On information rates for mismatched decoders," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[34] R. Zamir, "The index entropy of a mismatched codebook," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 523–528, Feb. 2002.

[35] A. Somekh-Baruch, "A general formula for the mismatch capacity," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 4554–4568, Sep. 2015.

[36] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Multiuser random coding techniques for mismatched decoding," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 3950–3970, July 2016.

[37] L. Zhou, V. Y. F. Tan, and M. Motani, "The dispersion of mismatched joint source-channel coding for arbitrary sources and additive channels," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2234–2251, April 2019.

[38] F. Forges, "Non-zero-sum repeated games and information transmission," *in: N. Meggido, Essays in Game Theory in Honor of Michael Maschler, Springer-Verlag*, no. 6, pp. 65–95, 1994.

[39] H. von Stackelberg, *Marketform und Gleichgewicht*. Oxford University Press, 1934.

[40] J. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, pp. 286–295, 1951.

[41] D. Bergemann and S. Morris, "Information design, bayesian persuasion, and bayes correlated equilibrium," *American Economic Review Papers and Proceedings*, vol. 106, no. 5, pp. 586–591, May 2016.

[42] I. Taneva, "Information design," *Manuscript, School of Economics, The University of Edinburgh*, 2016.

[43] D. Bergemann and S. Morris, "Information design: a unified perspective," *Cowles Foundation Discussion Paper No 2075*, 2017.

[44] R. Alonso and O. Câmara, "Bayesian persuasion with heterogeneous priors," *Journal of Economic Theory*, vol. 165, pp. 672–706, 2016.

[45] M. Laclau and L. Renou, "Public persuasion," *working paper*, Feb. 2017.

[46] E. Tsakas and N. Tsakas, "Noisy persuasion," *working paper*, 2017.

[47] A. Blume, O. J. Board, and K. Kawamura, "Noisy talk," *Theoretical Economics*, vol. 2, pp. 395–440, July 2007.

[48] P. Hernández and B. von Stengel, "Nash codes for noisy channels," *Operations Research*, vol. 62, no. 6, pp. 1221–1235, Nov. 2014.

[49] C. Sims, "Implication of rational inattention," *Journal of Monetary Economics*, vol. 50, no. 3, pp. 665–690, April 2003.

[50] M. Gentzkow and E. Kamenica, "Costly persuasion," *American Economic Review*, vol. 104, pp. 457–462, 2014.

[51] A. Neyman and D. Okada, "Strategic entropy and complexity in repeated games," *Games and Economic Behavior*, vol. 29, no. 1–2, pp. 191–223, 1999.

[52] ——, "Repeated games with bounded entropy," *Games and Economic Behavior*, vol. 30, no. 2, pp. 228–247, 2000.

[53] ——, “Growth of strategy sets, entropy, and nonstationary bounded recall,” *Games and Economic Behavior*, vol. 66, no. 1, pp. 404–425, 2009.

[54] O. Gossner and N. Vieille, “How to play with a biased coin?” *Games and Economic Behavior*, vol. 41, no. 2, pp. 206–226, 2002.

[55] O. Gossner and T. Tomala, “Empirical distributions of beliefs under imperfect observation,” *Mathematics of Operation Research*, vol. 31, no. 1, pp. 13–30, 2006.

[56] ——, “Secret correlation in repeated games with imperfect monitoring,” *Mathematics of Operation Research*, vol. 32, no. 2, pp. 413–424, 2007.

[57] O. Gossner, R. Laraki, and T. Tomala, “Informationally optimal correlation,” *Mathematical Programming*, vol. 116, no. 1-2, pp. 147–172, 2009.

[58] O. Gossner, P. Hernández, and A. Neyman, “Optimal use of communication resources,” *Econometrica*, vol. 74, no. 6, pp. 1603–1636, 2006.

[59] P. Cuff and L. Zhao, “Coordination using implicit communication,” *IEEE Information Theory Workshop (ITW)*, pp. 467–471, 2011.

[60] P. Cuff, H. Permuter, and T. Cover, “Coordination capacity,” *Information Theory, IEEE Transactions on*, vol. 56, no. 9, pp. 4181–4206, Sept. 2010.

[61] P. Cuff and C. Schieler, “Hybrid codes needed for coordination over the point-to-point channel,” in *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 235–239, Sept. 2011.

[62] M. Le Treust, “Joint empirical coordination of source and channel,” *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5087–5114, Aug. 2017.

[63] G. Cervia, L. Luzzi, M. Le Treust, and M. R. Bloch, “Strong coordination of signals and actions over noisy channels with two-sided state information,” *preprint available [on-line] https://arxiv.org/abs/1801.10543*, 2018.

[64] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[65] M. Le Treust, “Empirical coordination with two-sided state information and correlated source and state,” in *IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 466–470.

[66] M. Le Treust and M. Bloch, “Empirical coordination, state masking and state amplification: Core of the decoder’s knowledge,” *IEEE International Symposium on Information Theory (ISIT)*, pp. 895–899, 2016.

[67] ——, “State leakage and coordination of actions: Core of decoder’s knowledge,” *preprint available [on-line] https://arxiv.org/abs/1812.07026*, Dec. 2018.

[68] M. Fekete, “Über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten,” *Mathematische Zeitschrift*, vol. 17, no. 1, pp. 228–249, Dec. 1923.

[69] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, Dec. 2011.

[70] R. Aumann and M. Maschler, *Repeated Games with Incomplete Information*. MIT Press, Cambrige, MA, 1995.

[71] R. Rockafellar, *Convex Analysis*, ser. Princeton landmarks in mathematics and physics.

[72] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: 2nd. Ed., Wiley-Interscience, 2006.

[73] N. Merhav and S. Shamai, “Information rates subject to state masking,” *Information Theory, IEEE Transactions on*, vol. 53, no. 6, pp. 2254–2261, June 2007.