



**HAL**  
open science

## Coordination et Communication Stratégique

Mael Le Treust

► **To cite this version:**

Mael Le Treust. Coordination et Communication Stratégique. Théorie de l'information [cs.IT]. CY Cergy Paris Université, 2022. tel-03896253

**HAL Id: tel-03896253**

**<https://hal.science/tel-03896253>**

Submitted on 13 Dec 2022

**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.



Distributed under a Creative Commons Attribution 4.0 International License

CY Cergy Paris Université

## Habilitation à Diriger des Recherches

présentée par

**Maël LE TREUST**

### Coordination et Communication Stratégique

dont la soutenance s'est tenue le mardi 29 novembre 2022 à 14h devant le jury :

Iryna	Andriyanova	Professeure Univ. CY	Référente CY
Pierre	Duhamel	Directeur de Recherche CNRS L2S	Examinateur
Inbar	Fijalkow	Professeure ENSEA	Garante HDR
Françoise	Forges	Professeure Univ. Paris-Dauphine	Rapporteure
Jean-Marie	Gorce	Professeur INSA Lyon	Examinateur
Olivier	Gossner	Directeur de Recherche CNRS CREST	Examinateur
Mikael	Skoglund	Professeur KTH Stockholm	Rapporteur
Bruno	Tuffin	Directeur de Recherche Inria Rennes	Examinateur
Michèle	Wigger	Professeure Télécom Paris	Rapporteure
Aylin	Yener	Professeure The Ohio State University	Examinatrice

ETIS UMR 8051, CY Cergy Paris Université, ENSEA, CNRS

[mael.le-treust@cnrs.fr](mailto:mael.le-treust@cnrs.fr)

<https://sites.google.com/site/maelletreust/>



## Résumé

Ce manuscrit d'HDR présente des résultats récents de Théorie de l'Information qui apportent un éclairage nouveau sur des problèmes de Théorie des Jeux, tels que la coordination des actions des joueurs ou le traitement de l'information stratégique. Bien que la Théorie de l'Information et la Théorie des Jeux aient des champs d'application distincts, ces deux théories s'appuient sur des outils mathématiques similaires. Les problèmes de coordination et d'information stratégique se posent pour la conception de réseaux d'appareils autonomes et pour l'étude du comportement d'agents rationnels. Les nouvelles générations de réseaux de communication intègrent des appareils qui prennent des décisions avec une plus grande autonomie. Ces machines observent leur environnement, interagissent, coopèrent avec d'autres machines ou avec des humains, s'adaptent dynamiquement à leur environnement et à la topologie du réseau.

Le théorème de codage de source avec perte de [Shannon \(1959\)](#) est la pierre angulaire de notre étude. Etant donnée une contrainte de capacité de communication, ce théorème caractérise le schéma de codage optimal pour compresser une source d'information. Les performances du codage sont évaluées via un critère de fidélité que nous interprétons comme une fonction de coût à minimiser. Nous revisitons ce problème de codage en adoptant le point de vue de la Théorie des Jeux, dans lequel chaque joueur optimise sa propre fonction de coût qui dépend des actions des autres joueurs.

Tout d'abord, nous étudions la capacité d'un encodeur et d'un décodeur à coordonner leurs actions avec les symboles générés par une source d'information. A cet effet, la suite d'actions de l'encodeur encapsule une version dégradée des symboles de source, qui sert à coordonner les futures actions du décodeur. Nous caractérisons les solutions de plusieurs instances du problème de coordination au Chap. [IV](#), et au Chap. [V](#), nous étendons ces résultats dans plusieurs directions, pour des problèmes de fuite d'état de canal, pour des problèmes de contrôle décentralisé et pour des problèmes de coordination forte. Le caractère stratégique de la transmission de l'information est étudié au Chap. [VI](#) lorsque les joueurs communiquent à travers un canal bruité. Nous considérons un encodeur s'engage à mettre en œuvre une stratégie, avant que les symboles de sources soient tirés, comme dans le jeu de persuasion Bayésienne.

En guise de perspectives, nous étudierons l'impact des contraintes de communication sur la transmission d'informations stratégiques dans les problèmes de "mechanism design" de [Jackson and Sonnenschein \(2007\)](#), et dans les jeux de "cheap talk" de [Crawford and Sobel \(1982\)](#). Une autre direction de recherche consiste à approfondir les problèmes d'optimisation qui apparaissent pour le codage stratégique, en utilisant des outils issus de la Théorie des Graphes et de l'Optimisation Convexe. Nous pensons que ces résultats apporteront un point de vue nouveau sur les problèmes ouverts en jeux répétés avec observation imparfaite et information incomplète. L'objectif est de construire une théorie qui englobe les résultats de codage standard en théorie de l'information et les résultats pour les jeux "sender-receiver".



## Abstract

This HDR manuscript presents recent Information-Theoretic tools that bring new insights into Game Theoretical problems, such as the coordination of players' actions and the processing of strategic information. Although Information Theory and Game Theory have distinct fields of application, these two theories rely on similar mathematical tools. Such problems arise for the design of autonomous devices and for the study of the behavior of rational agents. New generations of communication networks involve devices that take decisions with greater autonomy. These machines observe their environment, interact, cooperate with other machines or with humans, adapt dynamically to their environment and to the topology of the network.

The lossy source coding theorem of [Shannon \(1959\)](#) is the cornerstone of our study. Given a communication capacity constraint, this theorem characterizes the optimal coding scheme to compress an information source. The coding performances are evaluated via a fidelity criterion which we interpret as a cost function to be minimized. We revisit this coding problem by adopting a Game-Theoretical point of view, in which each player optimizes his own cost function which depends on the actions of other players.

First, we investigate the capacity of an encoder and a decoder to coordinate their actions with the symbols of an information source. For this purpose, the action sequence of the encoder encapsulates a degraded version of the source symbols, which serves to coordinate the future actions of the decoder. We characterize the solutions of several instances of the coordination problem in [Chap. IV](#), and in [Chap. V](#), we extend these results in several directions, i.e. for channel state leakage problems, for decentralized control problems and for strong coordination problems. The strategic nature of the information transmission is studied in [Chap. VI](#) when the communication is passed through a noisy channel. We consider that the encoder commits to implementing a signalling strategy before it observes the source symbols, as in the Bayesian persuasion game.

As perspectives, we will study the impact of communication constraints on the transmission of strategic information in mechanism design problems, as in [Jackson and Sonnenschein \(2007\)](#), and in cheap talk games, see [Crawford and Sobel \(1982\)](#). Another direction of research consists in deepening optimization problems that arise in strategic coding, by using tools from Graph Theory and Convex Optimization. We believe that these results will also bring a new point of view on open problems in repeated games with imperfect observation and incomplete information. The goal is to build a theory that encompasses the standard coding results in Information Theory, and the results for the sender-receiver games.



# Contents

<b>Table of content</b>	<b>2</b>
<b>I Career achievements</b>	<b>5</b>
I.1 List of publications . . . . .	6
<b>II Introduction</b>	<b>12</b>
II.1 Research activity review . . . . .	13
II.1.a Point-to-point scenario . . . . .	13
II.1.b Coordination problem . . . . .	14
II.1.c Coordination and channel state masking with Matthieu Bloch . . . . .	15
II.1.d Coordination and decentralized control with Tobias Oechtering . . . . .	16
II.1.e Persuasion game with restricted communication with Tristan Tomala . . . . .	16
II.1.f HARQ protocols with Leszek Szczecinski . . . . .	17
II.2 Supervision of PhD students . . . . .	18
II.2.a Strong coordination with Giulia Cervia, Laura Luzzi and Matthieu Bloch . . . . .	18
II.2.b Multi-decoders Bayesian persuasion with Rony Bou Rouphael . . . . .	18
II.2.c Zero-error source coding problems with Nicolas Charpenay and Aline Roumy . . . . .	19
II.3 Content of this manuscript . . . . .	20
<b>III Information Theory review</b>	<b>21</b>
III.1 Shannon’s separation results . . . . .	22
III.1.a Perfect channel of restricted cardinality . . . . .	22
III.1.b General noisy channel . . . . .	24
III.1.c Entropy formulas . . . . .	25
III.2 Perfect channel of fixed capacity . . . . .	26
III.2.a Lossy source coding . . . . .	26
III.2.b Rational inattention . . . . .	27
III.2.c Multi-objective coding problem . . . . .	28
<b>IV Empirical coordination with a strictly causal or a causal decoder</b>	<b>29</b>
IV.1 Introduction . . . . .	30
IV.2 Empirical coordination with strictly causal decoder . . . . .	33
IV.2.a Problem statement . . . . .	33
IV.2.b Characterization of the set of achievable distributions . . . . .	34
IV.2.c Evaluation of an arbitrary expected cost function . . . . .	35
IV.2.d Example #1: coordination game of Gossner et al. (2006) . . . . .	36
IV.2.e Example #2: Hamming distortion and channel cost . . . . .	38
IV.3 General decoding functions . . . . .	40
IV.3.a Causal decoding with source feedforward . . . . .	40
IV.3.b Strictly causal decoding with source feedforward . . . . .	42
IV.4 State-dependent source and state-dependent channel . . . . .	43
IV.4.a Non-causal encoder and non-causal decoder . . . . .	43
IV.4.b Causal decoder and non-causal encoder . . . . .	47



---

IV.4.c Causal encoder and non-causal decoder . . . . .	49
IV.5 Chapter summary . . . . .	51
<b>V Impact of empirical coordination results</b>	<b>52</b>
V.1 State leakage and coordination with causal state knowledge at the encoder . . . . .	53
V.1.a Problem statement . . . . .	54
V.1.b Characterization of achievable triples of rate, leakage and distribution . . . . .	55
V.1.c Channel state estimation via cost function . . . . .	57
V.1.d Extensions to more general scenarios . . . . .	59
V.1.e Strictly causal encoding . . . . .	61
V.2 Decentralized control: Witsenhausen’s counterexample . . . . .	62
V.2.a System model . . . . .	63
V.2.b Coding Result . . . . .	64
V.2.c Control schemes of particular interest . . . . .	65
V.2.d Numerical Results . . . . .	68
V.2.e Discussion . . . . .	69
V.3 Strong coordination of signals and actions . . . . .	70
V.3.a Inner and outer bounds for the strong coordination region . . . . .	72
V.3.b Strong coordination region with state and side information . . . . .	73
V.3.c Characterization of the strong coordination region for special cases . . . . .	75
V.3.d Is separation optimal? . . . . .	78
V.4 Chapter summary . . . . .	81
<b>VI Bayesian persuasion with restricted communication</b>	<b>82</b>
VI.1 Persuasion with limited communication capacity . . . . .	83
VI.1.a Bayesian persuasion model . . . . .	86
VI.1.b Persuasion model with restricted communication . . . . .	87
VI.1.c Results . . . . .	90
VI.1.d Convexification with information constraint . . . . .	93
VI.1.e Illustrating example . . . . .	94
VI.1.f Beyond identical problems . . . . .	100
VI.2 Persuasion game with side information at the decoder . . . . .	101
VI.2.a System model . . . . .	102
VI.2.b Convex closure formulation . . . . .	104
VI.2.c Doubly symmetric binary source . . . . .	106
VI.3 Bayesian persuasion with several decoders . . . . .	111
VI.3.a System model . . . . .	111
VI.3.b Main result . . . . .	112
VI.4 Chapter summary . . . . .	113
<b>VII Conclusion and perspectives</b>	<b>114</b>
VII.1 Zero-error coding for computing problem . . . . .	115
VII.1.a Problem statement . . . . .	115
VII.1.b General setting . . . . .	116
VII.1.c Pairwise shared side information . . . . .	117
VII.2 Convex optimization algorithm for coordination problems . . . . .	118
VII.2.a Problem formulation . . . . .	118
VII.2.b Connection with Blahut-Arimoto’s algorithm . . . . .	119
VII.2.c Original optimization algorithm . . . . .	120
VII.3 Mechanism design and cheap talk problems . . . . .	120
VII.3.a System model . . . . .	121
VII.3.b Upper bound . . . . .	121
VII.4 Repeated games with imperfect monitoring and incomplete information . . . . .	122
<b>Bibliography</b>	<b>134</b>

---

# Chapter I

## Career achievements



Since October 2013, I work as a CNRS researcher. From Oct. 2013 to Aug. 2022, I was with ETIS UMR 8051 in Cergy, France. Since Sept. 2022, I am with IRISA UMR 6074 in Rennes, France. I received the B.Sc. in Mathematics from the Université de Rennes 1, France in 2005, the M.Sc. degree in Optimization, Game Theory & Economics (OJME) from the Université de Paris VI (UPMC), France in 2008 and the Ph.D. degree from the Université de Paris Sud XI in 2011, at the L2S UMR 8506, in Gif-sur-Yvette, France. In 2012-2013, I was a post-doctoral researcher at the LIGM UMR 8049 (Université Paris-Est) in Marne-la-Vallée, France, and at the Centre Énergie, Matériaux et Télécommunication (Université INRS) in Montréal, Canada.

### Contents

---

I.1 List of publications . . . . .	6
------------------------------------	---

---

## I.1 List of publications

### INTERNATIONAL JOURNALS (10)

- [J10] Le Treust, M. and Bloch, M. *State Leakage and Coordination of Actions*, IEEE Trans. on Information Theory, Volume: 67, Issue: 2, Pages 805-823, Feb. 2021.
- [J9] Cervia, G. and Luzzi, L. and Le Treust, M. and Bloch, M. *Strong coordination of signals and actions over noisy channels with two-sided state information*, IEEE Trans. on Information Theory, Volume: 66, Issue: 8, Pages: 4681-4708, Aug. 2020.
- [J8] Le Treust, M. and Tomala, T. *Persuasion with limited communication resources*, Journal of Economic Theory, vol. 184, page 104940, Nov. 2019.
- [J7] Le Treust, M. and Szczecinski, L. and F. Labeau *Rate Adaptation for Secure HARQ Protocols*, IEEE Trans. on Information Forensics and Security, Volume 13, Issue 12, Pages 2981 - 2994, May 2018.
- [J6] Le Treust, M. *Empirical Coordination for Joint Source-Channel Coding*, IEEE Trans. on Information Theory, Volume 63, Issue 8, Pages 5087 - 5114, Aug. 2017.
- [J5] Jabi, M. and Benyouss, A. and Le Treust, M. and Pierre-Doray, M. and Szczecinski, L. *Adaptive Cross-Packet HARQ*, IEEE Trans. on Communication, Volume 65, Issue 5, Pages: 2022-2035, May 2017.
- [J4] Mheich, Z. and Le Treust, M. and Alberge, F. and Duhamel, P. *Rate Adaptation for Incremental Redundancy Secure HARQ*, IEEE Trans. on Communications, Volume 64, Issue 2, pp 765-777, Feb. 2016.
- [J3] Le Treust, M. and Lasaulce, S. and Hayel, Y. and He, G., *Green power control in cognitive wireless networks*, IEEE Trans. on Vehicular Technology, Volume 62, Issue 4, pp 1741-1754, May 2013.
- [J2] Le Treust, M. and Lasaulce, S., *Transforming Monitoring Structures with Resilient Encoders. Application to Repeated Games*, Springer journal of Dynamic Games and Applications, Volume 3, Issue 1, pp 38-67, March 2013.
- [J1] Le Treust, M. and Lasaulce, S., *A repeated game formulation of energy-efficient decentralized power control*, IEEE Trans. on Wireless Communications, Volume 9, Issue 9, pp 2860-2869, Sept. 2010.

### INTERNATIONAL CONFERENCES (31)

- [IC31] Bou Roupheal, R., Le Treust, M., *Strategic Communication with Cost-Dependent Decoders via the Gray-Wyner Network*, IEEE Information Theory Workshop (ITW), Mumbai, 2022.
- [IC30] Charpenay, N., Le Treust, M. and Roumy, A., *Zero-error source coding when side information may be present*, International Zürich Seminar (IZS), March, 2022.
- [IC29] Bou Roupheal, R. and Le Treust, M., *Strategic Successive Refinement Coding for Bayesian Persuasion with Two Decoders*, IEEE Information Theory Workshop (ITW) 2021, Kanazawa, Japan.

- [IC28] Le Treust, M. and Tomala, T., *Strategic Communication with Decoder Side Information*, IEEE Information Symposium on Information Theory (ISIT), Melbourne, Australia, July 2021.
- [IC27] Le Treust, M. and Oechtering, T., *Continuous Random Variable Estimation is not Optimal for the Witsenhausen Counterexample*, IEEE Information Symposium on Information Theory (ISIT), Melbourne, Australia, July 2021.
- [IC26] Le Treust, M. and Tomala, T., *Point-to-Point Strategic Communication*, IEEE Information Theory Workshop (ITW), Riva del Garda, Italy, 2020.
- [IC25] Charpenay, N. and Le Treust, M., *Zero-Error Coding with a Generator Set of Variable-Length Words*, IEEE International Symposium on Information Theory (ISIT), Los Angeles, June 2020.
- [IC24] Bou Rouchael, R. and Le Treust, M., *Impact of Private Observation in Bayesian Persuasion*, International Conference on NETWORK Games, Control and Optimisation (NETGCOOP), Cargèse, Corsica, France on March 18-20, 2020.
- [IC23] Oechtering, T. and Le Treust, M., *Coordination Coding with Causal Decoder for Vector-valued Witsenhausen Counterexample Setups*, IEEE Information Theory Workshop (ITW), Visby, Aug. 2019.
- [IC22] Le Treust, M. and Oechtering, T. *Optimal Control Designs for Vector-valued Witsenhausen Counterexample Setups*, IEEE 56th Allerton Conference on Control Communication and Computing, Monticello, Illinois, 2018.
- [IC21] Cervia, G. and Luzzi, L. and Le Treust, M. and Bloch, M. *Strong Coordination Over Noisy Channels with Strictly Causal Encoding*, IEEE 56th Allerton Conference on Control Communication and Computing, Monticello, Illinois, 2018.
- [IC20] Le Treust, M. and Tomala, T. *Strategic Coordination with State Information at the Decoder*, International Zurich Seminar on Information and Communication, Zurich, Switzerland, 2018.
- [IC19] Cervia, G. and Luzzi, L. and Le Treust, M. and Bloch, M. *Strong Coordination of Signals and Actions over Noisy Channels*, IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 2017.
- [IC18] Le Treust, M. and Tomala, T., *Information Design for Strategic Coordination of Autonomous Devices with Non-Aligned Utilities*, IEEE Proc. of the 54th Allerton conference, Monticello, Illinois, 2016.
- [IC17] Cervia, G. and Luzzi, L. and Bloch, M. and Le Treust, M. *Polar coding for empirical coordination of signals and actions over noisy channels*, IEEE Proc. of the Information Theory Workshop (ITW), Cambridge, UK, 2016.

- [IC16] Le Treust, M. and Bloch, M., *Empirical Coordination, State Masking and State Amplification: Core of the Decoder's Knowledge*, IEEE Proc. of the International Symposium on Information Theory (ISIT), Barcelona, Spain, 2016.
- [IC15] Benyoussef, A. and Jabri, M. and Le Treust, M. and Szczecinski, L., *Joint Coding/Decoding for Multi-message HARQ*, IEEE Proc. of the Wireless Communications and Networking Conference (WCNC), Doha, Qatar, 2016.
- [IC14] Le Treust, M., *Empirical Coordination with Channel Feedback and Strictly Causal or Causal Encoding*, IEEE Proc. of the International Symposium on Information Theory (ISIT), Hong-Kong, China, 2015.
- [IC13] Le Treust, M., *Empirical Coordination with Two-Sided State Information and Correlated Source and State*, IEEE Proc. of the International Symposium on Information Theory (ISIT), Hong-Kong, China, 2015.
- [IC12] Le Treust, M., *Correlation between channel state and information source with empirical coordination constraint*, IEEE Proc. of the Information Theory Workshop (ITW2014), Hobart, Australia, 2014.
- [IC11] Mheich, Z. and Le Treust, M. and Alberge, F. and Duhamel, P. and Szczecinski, L. *Rate-adaptive secure HARQ protocol for block-fading channels*, IEEE Proc. of the 22nd European Signal Processing Conference (EUSIPCO), 2014.
- [IC10] Varma, V. S. and Mhiri, M. and Le Treust, M. and Lasaulce, S. and Samet, A. *On the benefits of repeated game models for green cross-layer power control in small cells*, IEEE Proc. of the First International Black Sea Conference on Communications and Networking (BlackSeaCom), 2013.
- [IC9] Le Treust, M. and Szczecinski L. and Labeau F., *Secrecy & Rate Adaptation for Secure HARQ Protocols*, IEEE Proc. of the Information Theory Workshop (ITW2013), Sevilla, Spain, 2013.
- [IC8] Le Treust, M. and Lasaulce, S., *Resilient Source Coding*, IEEE Proc. of the international conference on network games, control and optimization (NETGCOOP11), (Paris) 2011.
- [IC7] Le Treust, M. and Zaidi, A. and Lasaulce, S., *An Achievable Rate Region for the Broadcast Wiretap Channel with Asymmetric Side Information*, IEEE Proc. of the 49th Allerton conference, Monticello, Illinois, 2011.
- [IC6] Meriaux, F. and Le Treust, M. and Lasaulce, S. and Kieffer, M., *Energy-efficient power control strategies for stochastic games*, IEEE Proc. of the 17th International Conference on Digital Signal Processing (DSP), 2011.

- [IC5] Le Treust, M. and Lasaulce, S., *The price of re-establishing almost perfect monitoring in games with arbitrary monitoring structures*, ACM Proc. of the 4th International Workshop on Game Theory in Communication Networks (GAMECOMM11), Cachan (Paris), France, 2011.
- [IC4] Le Treust, M. and Hayel, Y. and Lasaulce, S. and Debbah, M., *To Sense or not to Sense in Energy-Efficient Power Control Games*, IEEE Proc. of the 2th International Conference on Game Theory for Networks (GAMENETS11), Shanghai, China, 2011.
- [IC3] Le Treust, M. and Tembine, H. and Lasaulce, S. and Debbah, M., *Coverage games in small cells networks*, IEEE Proc. of the Future Network and Mobile Summit (FUNEMS10), Florence, Italy, 2010.
- [IC2] Le Treust, M. and Lasaulce, S. and Debbah, M., *Implicit cooperation in distributed energy-efficient networks*, IEEE Proc. of the 4th International Symposium on Communications, Control and Signal Processing (ISCCSP10), Limassol, Cyprus, 2010.
- [IC1] Bennis, M. and Le Treust, M. and Lasaulce, S. and Debbah, M., *Spectrum sharing games in the interference channel*, IEEE Proc. of the 1st International Conference on Game Theory for Networks (GAMENETS09), Istanbul, Turkey, 2009.

#### NATIONAL CONFERENCES (11)

- [NC11] Charpenay, N. and Le Treust, M. and Roumy, A., *Codage de source zéro-erreur pour un réseau de Gray-Wyner lorsque l'information adjacente peut être présente*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Nancy, France, 2022.
- [NC10] Bou Roupheal, R. and Le Treust, M., *Communication stratégique via le réseau Gray-Wyner*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Nancy, France, 2022.
- [NC9] Le Treust, M. and Bloch, M., *Jeu d'estimation de l'état de canal et coordination*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Lille, France, 2019.
- [NC8] Le Treust, M. and Tomala, T., *Persuasion Bayésienne pour la Coordination Stratégique d'Appareils Autonomes ayant des Objectifs Non-Alignés.*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Juan-les-Pins, France, 2017.
- [NC7] Cervia, Giulia and Luzzi, Laura and Le Treust, Mael and Bloch, Matthieu R., *Polar codes for empirical coordination over noisy channels with strictly causal encoding*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Juan-les-Pins, France, 2017.
- [NC6] Le Treust, M., *Coordination d'appareils autonomes dans un réseau décentralisé.*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Lyon, France, 2015.

- [NC5] Meriaux, F. and LeTreust, M. and Lasaulce, S. and Kieffer, M., *Jeux stochastiques et contrôle de puissance distribué*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI11), Bordeaux, France, 2011.
- [NC4] Le Treust, M. and Lasaulce, S. , *Codage de Source Stratégique avec Information Adjacente au Décodeur*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI11), Bordeaux, France, 2011.
- [NC3] Debbah, M. and Lasaulce, S. and LeTreust, M. and Tembine, H. , *Contrôle de puissance dynamique pour les communications sans fils*, Actes de la Société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF10), Toulouse, 2010.
- [NC2] Le Treust, M. and Lasaulce, S., *Approche bayésienne pour le contrôle de puissance décentralisé efficace énergétiquement*, Actes de la Société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF10), Toulouse, 2010.
- [NC1] Le Treust, M and Medina Perlaza, S. and Lasaulce, S., *Contrôle de puissance distribué efficace énergétiquement et jeux répétés*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI09), Dijon, France, 2009.

#### POSTER (2)

- [P2] Bou Rouphael, R. and Le Treust, M., *Impact of Private Observation in Persuasion Game*, GAMENET Conference 2019, Prague, Czech Republic, Nov. 21-22, 2019
- [P1] Cervia, G., Luzzi, L., Bloch, M., and Le Treust, M., *Polar coding for empirical coordination of signals and actions over noisy channels*, European School on Information Theory, Gothenburg (ESIT), Sweden, April 4-8, 2016

#### WORKING PAPERS (9)

- [W9] Le Treust, M. and Oechtering, T., *Power-Estimation Trade-off of Vector-valued Witsenhausen Counterexample with Causal Decoder*, submitted to IEEE Transactions on Information Theory, [on-line] available: <http://arxiv.org/abs/2211.09200>, Nov. 2022.
- [W8] Charpenay, N., Le Treust, M. and Roumy, A., *Zero-Error Coding for Computing with Encoder Side-Information*, submitted to IEEE Information Theory Workshop (ITW), Saint-Malo, France, [on-line] available: <https://arxiv.org/abs/2211.03649>, 2023.
- [W7] Bou Rouphael, R. and Le Treust, M., *Strategic Communication via Cascade Multiple Description Network*, [on-line] available: <https://arxiv.org/abs/2210.00533>, 2022.
- [W6] Charpenay, N., Le Treust, M. and Roumy, A., *Zero-error source coding with two-sided side-information*, working paper, Apr. 2022.
- [W5] Bou Rouphael, R. and Le Treust, M., *Strategic Successive Refinement for Bayesian Persuasion with Cost-Dependent Decoders*, second round of review for IEEE JSAIT, Apr. 2022.
- [W4] Charpenay, N. and Le Treust, M. *Variable-Length Coding for Zero-Error Channel Capacity*, preliminary draft, [on-line] available: <https://arxiv.org/abs/2001.03523>, Jan. 2020.
- [W3] Le Treust, M. and Tomala, T. *Strategic Communication with Side Information at the Decoder*, preliminary draft, [on-line] available: <https://arxiv.org/abs/1911.04950>, Nov. 2019.

[W2] Le Treust, M. and Tomala, T. *Information-Theoretic Limits of Strategic Communication*, preliminary draft, [on-line] available: <https://arxiv.org/abs/1807.05147>, Jul. 2018.

[W1] Le Treust, M. *Coding Theorems for Empirical Coordination*, preliminary draft, [on-line] available: <https://hal.archives-ouvertes.fr/hal-01865569>, Apr. 2015.

MANUSCRIPTS (3)

[M3] Le Treust, M., *Coordination and Strategic Communication*, Habilitation à Diriger les Recherches (HDR), CY Cergy Paris Université, Dec. 2022

[M2] Le Treust, M., *Théorie de l'information, jeux répétés avec observation imparfaite et réseaux de communication décentralisés*, PhD Thesis, Université Paris Sud XI, Dec. 2011

[M1] Le Treust, M., *Folk theorem and Imperfect Monitoring*, Master Thesis, Université Pierre et Marie Curie Paris VI, Sept. 2008



# Chapter II

## Introduction



(The Cardsharps, Caravaggio)

This manuscript presents an overview of the results obtained at the interplay between Information Theory and Game Theory. The goal is to build a theory that encompasses the standard coding results in Information Theory, and the results for the sender-receiver games. Our findings will be applied to decentralized decision-making problems and digital communication networks. Specific topics of interest include Convex Optimization, Graph Theory and zero-error coding, repeated game with incomplete information and imperfect monitoring.

### Contents

---

II.1	Research activity review . . . . .	<b>13</b>
II.1.a	Point-to-point scenario . . . . .	13
II.1.b	Coordination problem . . . . .	14
II.1.c	Coordination and channel state masking with Matthieu Bloch . . . . .	15
II.1.d	Coordination and decentralized control with Tobias Oechtering . . . . .	16
II.1.e	Persuasion game with restricted communication with Tristan Tomala . . . . .	16
II.1.f	HARQ protocols with Leszek Szczecinski . . . . .	17
II.2	Supervision of PhD students . . . . .	<b>18</b>
II.2.a	Strong coordination with Giulia Cervia, Laura Luzzi and Matthieu Bloch . . . . .	18
II.2.b	Multi-decoders Bayesian persuasion with Rony Bou Rouphael . . . . .	18
II.2.c	Zero-error source coding problems with Nicolas Charpenay and Aline Roumy . . . . .	19
II.3	Content of this manuscript . . . . .	<b>20</b>

---

## II.1 Research activity review

Although Information Theory and Game Theory have distinct fields of application, they rely on similar mathematical tools from Probability, Statistics, Convex Optimization, Graph Theory, Combinatorial Optimization. Information Theory has applications in Electrical Engineering and Computer Science for compression, storage, image processing and digital communication problems. Game Theory has applications in Economics, in Social Sciences, in Biology and in Computer Science for decentralized decision-making problems.

This manuscript presents recent Information Theoretic tools in order to study the coordination of players' actions and the processing of strategic information in Game Theory. This question is at the intersection of many current research directions regarding the design of autonomous devices and the study of the behavior of rational agents. New generations of communication networks involve devices or agents that take into account the communication constraints and that make decisions with greater autonomy. They observe their environment, interact, cooperate with other machines or humans, adapt dynamically to their environment and to the topology of the network.

### II.1.a Point-to-point scenario

To address this problem, we revisit seminal point-to-point results such as the channel coding theorem of Shannon (1948) and the lossy source coding theorem of Shannon (1959). This last problem depends of a fidelity criterion which we interpret as a decision maker's cost function to be minimized. In this direction, the scenario depicted in Fig. III.3 is of particular interest.

In this manuscript, the sender and the receiver are referred to as the encoder and the decoder. We denote by  $\Delta(\mathcal{X})$  the set of probability distributions  $\mathcal{P}_X$  over the finite set  $\mathcal{X}$ , and by  $\Delta(\mathcal{Y})^{|\mathcal{X}|}$  the set of conditional probability distributions  $\mathcal{P}_{Y|X}$ . The information source  $(\mathcal{U}, \mathcal{P}_U)$  is defined by a finite set  $\mathcal{U}$  and a probability distribution  $\mathcal{P}_U \in \Delta(\mathcal{U})$  such that the symbol  $u \in \mathcal{U}$  is selected with probability  $\mathcal{P}_U(u) \in [0, 1]$ . The cost function  $c : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}$  is evaluated over  $\mathcal{U}$  and the finite set of decoder actions  $\mathcal{V}$ . The channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$  is defined by two finite sets of channel inputs  $\mathcal{X}$  and channel outputs  $\mathcal{Y}$ , and a conditional probability distribution  $\mathcal{T}_{Y|X} \in \Delta(\mathcal{Y})^{|\mathcal{X}|}$  such that when the encoder chooses input symbol  $x$ , the decoder receives output symbol  $y$  with probability  $\mathcal{T}_{Y|X}(y|x) \in [0, 1]$ . Throughout this manuscript, calligraphic fonts, e.g.  $\mathcal{X}$ , stand for finite sets, capital letters, e.g.  $X$ , denote random variables while lowercase letters, e.g.  $x \in \mathcal{X}$  denote the realizations. Sequences of length  $n \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$  of random variables and realizations are denoted respectively by  $X^n = (X_1, \dots, X_t, \dots, X_n)$  and  $x^n = (x_1, \dots, x_t, \dots, x_n)$ , for  $t \in \{1, \dots, n\}$ .

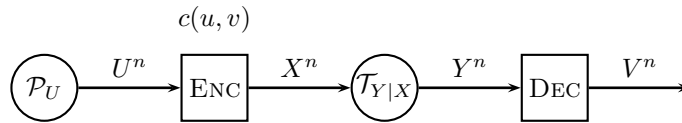


Figure II.1 – Point-to-point scenario with an information source  $(\mathcal{U}, \mathcal{P}_U)$ , a channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$  and a cost function  $c : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}$ .

The information source generates an i.i.d. sequence  $U^n$  of length  $n \in \mathbb{N}^*$ . The encoder observes the realization  $u^n \in \mathcal{U}^n$  and selects a sequence of channel inputs  $x^n \in \mathcal{X}^n$ . The channel generates an output sequence  $Y^n$  according to the i.i.d. conditional distribution  $\mathcal{T}_{Y|X}$  evaluated with respect to  $x^n \in \mathcal{X}^n$ . The decoder observes the realization  $y^n \in \mathcal{Y}^n$  and selects a sequence of actions  $v^n \in \mathcal{V}^n$ . The encoding function  $\sigma : \mathcal{U}^n \rightarrow \mathcal{X}^n$  and the decoding functions  $\tau : \mathcal{Y}^n \rightarrow \mathcal{V}^n$  are selected in order to minimize the expected long-run cost. The coding problem is defined by

$$C^n = \min_{\sigma, \tau} \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n c(U_t, V_t) \right]. \quad (\text{II.1})$$

In [Shannon \(1948\)](#) and [Shannon \(1959\)](#), the asymptotic value of  $C^n$  is characterized by

$$C^* = \min_{\substack{\mathcal{Q}_{V|U} \in \Delta(\mathcal{V})^{|\mathcal{U}|}, \\ I(U;V) \leq \max_{\mathcal{P}_X} I(X;Y)}} \mathbb{E}[c(U, V)], \quad (\text{II.2})$$

where the mutual information

$$I(U; V) = \sum_{(u,v) \in \text{supp } \mathcal{P}_U \mathcal{Q}_{V|U}} \mathcal{P}_U(u) \mathcal{Q}_{V|U}(v|u) \log_2 \frac{\mathcal{Q}_{V|U}(v|u)}{\sum_{u'} \mathcal{P}_U(u') \mathcal{Q}_{V|U}(v|u')}, \quad (\text{II.3})$$

measures the correlation of  $U$  and  $V$ . An important feature of [\(II.2\)](#) is the restriction imposed on the set of achievable conditional distributions  $\mathcal{Q}_{V|U} \in \Delta(\mathcal{V})^{|\mathcal{U}|}$ , by the information constraint  $I(U;V) \leq \max_{\mathcal{P}_X} I(X;Y)$ . This captures the restrictions imposed by the repeated use of the channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ , on the empirical distribution of the random variables  $U$  and  $V$ .

In the Game Theory literature, the optimization problem of [\(II.2\)](#) gave rise to as the rational inattention problem, formulated in [Sims \(2003\)](#). The goal is to determine the optimal trade-off between the cost of information extraction  $I(U;V)$  and the benefit of information exploitation for the minimization of  $\mathbb{E}[c(U, V)]$ . Note that the setting of [Fig. III.3](#) is also related to the model of repeated game of incomplete information [Aumann and Maschler \(1995\)](#) and of the sender-receiver game [Crawford and Sobel \(1982\)](#).

We present a brief summary of the different problems under study in this manuscript and the main contributions.

### II.1.b Coordination problem

An intriguing information constraint  $I(U;V) \leq H(X|U, V)$  was stated in [Gossner et al. \(2006\)](#) for a repeated game problem with incomplete information. The encoder and the decoder are the players of the game, the decoder actions only depends on past channel outputs, and the channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$  is perfect, i.e.  $\mathcal{X} = \mathcal{Y}$  and  $\mathcal{T}_{Y|X}(y|x) = \mathbb{1}\{x = y\}$ , where the indicator function is equal to 1 if  $x = y$ , and 0 otherwise.

The main difference with the model of [Sec. II.1.a](#) is the cost function

$$c : \mathcal{U} \times \mathcal{X} \times \mathcal{V} \rightarrow \mathbb{R}, \quad (\text{II.4})$$

which also depends on the encoder action  $X$ , i.e. the channel input symbol.

When selecting the actions, the encoder faces a trade-off between minimizing of the cost function with  $X$  and sending information to the decoder so that it can also minimize the cost function with  $V$ . The goal of the players is to implement sequences of actions that are empirically correlated with the source of information. The authors show that the optimal solution is characterized via the information constraint  $I(U;V) \leq H(X|U, V)$ . This problem has given rise to a literature on empirical coordination problems in Information Theory, see [Cuff et al. \(2010\)](#). The key is to control the empirical distribution of symbols of source  $U$  and actions  $(X, V)$ , induced by the coding scheme.

A series of contributions consists in characterizing the solutions of several instances of the coordination problem, when the observation of actions is imperfect, when the decoder also observes past source symbols, when the decoder or the encoder operates in a strictly causal, causal or non-causal manner, when the channel and the sources depend on state parameters partially observed by the encoder and by the decoder. We show that the coordination problem is related to the joint source-channel coding problem with two-sided state information, which is a famous object of study in Information Theory, see [Merhav and Shamai \(2003\)](#). We notice that the source symbol observed by the encoder must be considered as the state of a state-dependent channel, as in [Gel'fand and Pinsker \(1980\)](#), whereas the channel output must be considered as side information of the decoder, as in [Wyner and Ziv \(1976\)](#). This research work resulted in the following publications:

- [J6] [Le Treust, M. Empirical Coordination for Joint Source-Channel Coding, IEEE Transactions on Information Theory, Volume 63, Issue 8, Pages 5087 - 5114, Aug. 2017.](#)

- [IC14] Le Treust, M., *Empirical Coordination with Channel Feedback and Strictly Causal or Causal Encoding*, IEEE Proc. of the International Symposium on Information Theory (ISIT), Hong-Kong, China, 2015.
- [IC13] Le Treust, M., *Empirical Coordination with Two-Sided State Information and Correlated Source and State*, IEEE Proc. of the International Symposium on Information Theory (ISIT), Hong-Kong, China, 2015.
- [IC12] Le Treust, M., *Correlation between channel state and information source with empirical coordination constraint*, IEEE Proc. of the Information Theory Workshop (ITW2014), Hobart, Australia, 2014.
- [NC6] Le Treust, M., *Coordination d'appareils autonomes dans un réseau décentralisé.*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Lyon, France, 2015.
- [W1] Le Treust, M. *Coding Theorems for Empirical Coordination: Technical Report*, [on-line] available: <https://hal.archives-ouvertes.fr/hal-01865569>, Apr. 2015.

### II.1.c Coordination and channel state masking with Matthieu Bloch

The study of state-dependent channels can be traced back to the early works of Shannon (1958) and Gel'fand and Pinsker (1980), which identified optimal coding strategies to transmit reliably in a noisy environment when the encoder has some observation of the noise random variable, referred to as the state of the channel. In Kim et al. (2008), the authors design a coding scheme that conveys the channel state to the decoder, on top of the message transmission. They characterize the optimal trade-off between the rate capacity of reliable communication and the reduction of uncertainty about the channel state. In Merhav and Shamai (2007), the goal of the coding scheme is to hide the channel state variable from the decoder by minimizing the channel state leakage, while guaranteeing a certain transmission rate. The rate-leakage capacity region of state masking has been successfully characterized for both causal and non-causal state knowledge.

With Matthieu Bloch, Professor at Georgia Tech Atlanta and member of GeorgiaTech-CNRS IRL 2958, we revisit the problems of state masking and state amplification through the lens of empirical coordination. We introduce a channel state estimation zero-sum game in which the decoder minimizes the channel state estimation cost function. The coordination approach allows us to characterize the optimal solution of this zero-sum game, which was considered to be difficult, see (Merhav and Shamai, 2007, <sup>2</sup> pp. 2255). We characterize the optimal trade-offs between the coordination capacity, the channel state information leakage and the information rate transmitted to the decoder. We extend the results of (Le Treust, 2015b, [IC14]) when the encoder has noisy channel output feedback. This research work resulted in the following publications:

- [J10] Le Treust, M. and Bloch, M. *State Leakage and Coordination of Actions*, IEEE Transactions on Information Theory, Volume: 67, Issue: 2, Pages 805-823, Feb. 2021.
- [IC16] Le Treust, M. and Bloch, M., *Empirical Coordination, State Masking and State Amplification: Core of the Decoder's Knowledge*, IEEE Proc. of the International Symposium on Information Theory (ISIT), Barcelona, Spain, 2016.
- [NC9] Le Treust, M. and Bloch, M., *Jeu d'estimation de l'état de canal et coordination*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Lille, France, 2019.

### II.1.d Coordination and decentralized control with Tobias Oechtering

Distributed decision-making systems arise in many engineering problems where decentralized agents choose actions based on locally available information, so as to minimize a common cost function. The design of optimal strategies for such problems is considered to be notoriously difficult. In [Witsenhausen \(1968\)](#), the author introduces an outstanding toy example that help to grasp the fundamental difficulty when the actions serve two purposes, a control purpose affecting the system state and a communication purpose where information is sent to other agents, see [Yüksel and Başar \(2013\)](#). The coordination approach of [Sec. II.1.b](#) offers a new perspective on decentralized control problems.

With Tobias Oechtering, Professor at KTH Stockholm, we consider the vector version of the Witsenhausen’s counterexample introduced in [Grover and Sahai \(2010\)](#). We extend the finite alphabet coding scheme of ([Le Treust, 2017, \[J6\]](#)) to continuous alphabet by adapting the notion of weak typicality. We determine new coding strategies and we evaluate their performances. Interestingly, we show that a pair of discrete and continuous auxiliary random variables, outperforms both Witsenhausen’s two point strategy and the best affine policies. This research work resulted in the following publications:

- [IC27] Le Treust, M. and Oechtering, T., *Continuous Random Variable Estimation is not Optimal for the Witsenhausen Counterexample*, IEEE Information Symposium on Information Theory (ISIT), Melbourne, Australia, July 2021.
- [IC23] Oechtering, T. and Le Treust, M., *Coordination Coding with Causal Decoder for Vector-valued Witsenhausen Counterexample Setups*, IEEE Information Theory Workshop (ITW), Visby, Aug. 2019.
- [IC22] Le Treust, M. and Oechtering, T. *Optimal Control Designs for Vector-valued Witsenhausen Counterexample Setups*, IEEE 56th Allerton Conference on Control Communication and Computing, Monticello, Illinois, 2018.

### II.1.e Persuasion game with restricted communication with Tristan Tomala

The nature of the problem changes when the encoder and the decoder, see [Fig. III.3](#), selfishly optimize distinct cost functions

$$c_e : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R} \quad \neq \quad c_d : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}. \quad (\text{II.5})$$

The transmission of information becomes strategic. The encoder and the decoder are the players of a game with incomplete information. This framework has been well explored in the Game Theory literature, when the encoder can send a large number of messages that are perfectly received by the decoder.

In the cheap talk game of [Crawford and Sobel \(1982\)](#), the encoder observes a Gaussian random variable and selects a real number, which is transmitted and perfectly received by the decoder. The communication is unrestricted and the signals do not affect the cost functions. In this model, neither the encoder nor the decoder commit to a strategy. The natural solution concept is the Nash equilibrium, see [Nash \(1951\)](#).

In the Bayesian persuasion game of [Kamenica and Gentzkow \(2011\)](#), the encoder commits to implementing a coding strategy before the game plays. By knowing the encoding scheme, the decoder selects a best response action that depends on its posterior beliefs. The natural solution is the Stackelberg equilibrium, see [von Stackelberg \(1934\)](#), in which the encoder is the Stackelberg leader and the decoder is the Stackelberg follower.

With Tristan Tomala, Professor at HEC Paris and member of GREGHEC UMR 2959, we investigate the Bayesian persuasion problem when the encoder and the decoder communicate through a noisy channel, when the number of messages is restricted, and when the decoder observes a degraded version of the information source. This research work resulted in the following publications:

- [J8] Le Treust, M. and Tomala, T. *Persuasion with limited communication resources*, Journal of Economic Theory, vol. 184, page 104940, Nov. 2019.

- [IC28] Le Treust, M. and Tomala, T., *Strategic Communication with Decoder Side Information*, IEEE Information Symposium on Information Theory (ISIT), Melbourne, Australia, July 2021.
- [IC26] Le Treust, M. and Tomala, T., *Point-to-Point Strategic Communication*, IEEE Information Theory Workshop (ITW), Riva del Garda, Italy, 2020.
- [IC20] Le Treust, M. and Tomala, T. *Strategic Coordination with State Information at the Decoder*, International Zurich Seminar on Information and Communication, Zurich, Switzerland, 2018.
- [IC18] Le Treust, M. and Tomala, T., *Information Design for Strategic Coordination of Autonomous Devices with Non-Aligned Utilities*, IEEE Proc. of the 54th Allerton conference, Monticello, Illinois, 2016.
- [NC8] Le Treust, M. and Tomala, T., *Persuasion Bayésienne pour la Coordination Stratégique d'Appareils Autonomes ayant des Objectifs Non-Alignés.*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Juan-les-Pins, France, 2017.
- [W3] Le Treust, M. and Tomala, T. *Strategic Communication with Side Information at the Decoder*, preliminary draft, [on-line] available: <https://arxiv.org/abs/1911.04950>, Nov. 2019.
- [W2] Le Treust, M. and Tomala, T. *Information-Theoretic Limits of Strategic Communication*, preliminary draft, [on-line] available: <https://arxiv.org/abs/1807.05147>, Jul. 2018.

### II.1.f HARQ protocols with Leszek Szczecinski

During my post-doc at INRS, in Montreal in 2012 to 2013 under the supervision of Leszek Szczecinski, I worked on Hybrid Automatic Retransmission reQuest (HARQ) protocols for digital communication. In a fading environment, the channel capacity is a random variable which might be observed by the decoder only. When the rate parameter of the coding scheme exceeds the channel capacity, the decoder sends a NACK feedback to the encoder, that retransmits the codeword, as in [Caire and Tuninetti \(2001\)](#). We investigate several HARQ protocols that are able to adapt dynamically to the changes in the channel capacity. Following the work of [Tang et al. \(2009\)](#), we design a HARQ protocol tailored to ensures the secrecy of the transmission. This research work resulted in the following publications:

- [J7] Le Treust, M. and Szczecinski, L. and F. Labeau *Rate Adaptation for Secure HARQ Protocols*, IEEE Transactions on Information Forensics and Security, Volume 13, Issue 12, Pages 2981 - 2994, May 2018.
- [J5] Jabi, M. and Benyouss, A. and Le Treust, M. and Pierre-Doray, M. and Szczecinski, L. *Adaptive Cross-Packet HARQ*, IEEE Transactions on Communication, Volume 65, Issue 5, Pages: 2022-2035, May 2017.
- [J4] Mheich, Z. and Le Treust, M. and Alberge, F. and Duhamel, P. *Rate Adaptation for Incremental Redundancy Secure HARQ*, IEEE Transactions on Communications, Volume 64, Issue 2, pp 765-777, Feb. 2016.
- [IC15] Benyouss, A. and Jabi, M. and Le Treust, M. and Szczecinski, L., *Joint Coding/Decoding for Multi-message HARQ*, IEEE Proc. of the Wireless Communications and Networking Conference (WCNC), Doha, Qatar, 2016.
- [IC11] Mheich, Z. and Le Treust, M. and Alberge, F. and Duhamel, P and Szczecinski, L. *Rate-adaptive secure HARQ protocol for block-fading channels*, IEEE Proc. of the 22nd European Signal Processing Conference (EUSIPCO), 2014.

- [IC9] Le Treust, M. and Szczecinski L. and Labeau F., *Secrecy & Rate Adaptation for Secure HARQ Protocols*, IEEE Proc. of the Information Theory Workshop (ITW2013), Sevilla, Spain, 2013.

## II.2 Supervision of PhD students

### II.2.a Strong coordination with Giulia Cervia, Laura Luzzi and Matthieu Bloch

From 2015 to 2018, Giulia Cervia was PhD student at ETIS UMR 8051 in Cergy, under the direction of Inbar Fijalkow and the supervision of Laura Luzzi and myself. We worked in collaboration with Matthieu Bloch on the problem of strong coordination introduced in [Cuff et al. \(2010\)](#). Contrary to the notion of empirical coordination presented in [Sec. II.1.b](#) and [II.1.c](#), strong coordination requires that the probability distribution of the sequences of symbols converges to a target i.i.d. probability distribution.

Giulia Cervia investigates a joint source-channel coding problem in which both encoder and decoder have state information, as in ([Le Treust, 2015c](#), [\[IC13\]](#)). She proposes a practical coding scheme based on polar codes that ensures the strong coordination of the symbols of source and channel. This research work resulted in the following publications:

- [J9] Cervia, G. and Luzzi, L. and Le Treust, M. and Bloch, M. *Strong coordination of signals and actions over noisy channels with two-sided state information*, IEEE Transactions on Information Theory, Volume: 66, Issue: 8, Pages: 4681-4708, Aug. 2020.
- [IC21] Cervia, G. and Luzzi, L. and Le Treust, M. and Bloch, M. *Strong Coordination Over Noisy Channels with Strictly Causal Encoding*, IEEE 56th Allerton Conference on Control Communication and Computing, Monticello, Illinois, 2018.
- [IC19] Cervia, G. and Luzzi, L. and Le Treust, M. and Bloch, M. *Strong Coordination of Signals and Actions over Noisy Channels*, IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 2017.
- [IC17] Cervia, G. and Luzzi, L. and Bloch, M. and Le Treust, M. *Polar coding for empirical coordination of signals and actions over noisy channels*, IEEE Proc. of the Information Theory Workshop (ITW), Cambridge, UK, 2016.
- [NC7] Cervia, Giulia and Luzzi, Laura and Le Treust, Mael and Bloch, Matthieu R., *Polar codes for empirical coordination over noisy channels with strictly causal encoding*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Juan-les-Pins, France, 2017.
- [P1] Cervia, G., Luzzi, L., Bloch, M., and Le Treust, M., *Polar coding for empirical coordination of signals and actions over noisy channels*, European School on Information Theory, Gothenburg (ESIT), Sweden, April 4-8, 2016

### II.2.b Multi-decoders Bayesian persuasion with Rony Bou Rouphael

From 2019 to 2022, Rony Bou Rouphael is doing a PhD at ETIS UMR 8051 in Cergy under the direction of Inbar Fijalkow and my supervision. He works on three extensions of the Bayesian persuasion model of ([Le Treust and Tomala, 2019](#), [\[J8\]](#)), with several decoders and restricted communication.

In the model of successive refinement of [Equitz and Cover \(1991\)](#) and [Rimoldi \(1994\)](#), the encoder uses a common channel to both decoders and a private channel to the first decoder. Rony Bou Rouphael characterizes the single-letter solution of the Bayesian persuasion game where the encoder performs a successive refinement coding. In the model of [Gray and Wyner \(1974\)](#), an additional private channel is linked to the second decoder. Rony Bou Rouphael provides lower and upper bounds on the optimal cost of the encoder, by assuming that the decoder cost functions depend on the actions of the other

decoder. He also investigates the strategic version of the cascade multiple description network problem of Yamamoto (1981). This research work resulted in the following publications:

- [IC31] Bou Rouphael, R., Le Treust, M., *Strategic Communication with Cost-Dependent Decoders via the Gray-Wyner Network*, IEEE Information Theory Workshop (ITW), Mumbai, 2022.
- [IC29] Bou Rouphael, R. and Le Treust, M., *Strategic Successive Refinement Coding for Bayesian Persuasion with Two Decoders*, IEEE Information Theory Workshop (ITW) 2021, Kanazawa, Japan.
- [IC24] Bou Rouphael, R. and Le Treust, M., *Impact of Private Observation in Bayesian Persuasion*, International Conference on NETWORK Games, Control and Optimisation (NETGCOOP), Cargèse, Corsica, France on March 18-20, 2020.
- [NC10] Bou Rouphael, R. and Le Treust, M., *Communication stratégique via le réseau Gray-Wyner*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Nancy, France, 2022.
- [P2] Bou Rouphael, R. and Le Treust, M., *Impact of Private Observation in Persuasion Game*, GAMENET Conference 2019, Prague, Czech Republic, Nov. 21-22, 2019
- [W5] Bou Rouphael, R. and Le Treust, M., *Strategic Successive Refinement for Bayesian Persuasion with Cost-Dependent Decoders*, submitted to IEEE JSAIT, Apr. 2022.

### II.2.c Zero-error source coding problems with Nicolas Charpenay and Aline Roumy

From 2020 to 2023, Nicolas Charpenay is doing a PhD at IRISA UMR 6074 in Rennes under the direction of Aline Roumy and my supervision. He works on zero-error source coding problems with side information at the decoder, that are related to the video compression. The zero-error assumption requires that the probability of error in the coding, is exactly zero. The optimal compression rate of this difficult open problem corresponds to the minimal entropy of a coloring of the characteristic graph, see Witsenhausen (1976) and Alon and Orlitsky (1996).

Nicolas Charpenay characterizes the zero-error coding rate for the successive refinement coding problem, i.e. with several decoders, in which the encoder observes the decoder side information. This scenario arises in interactive compression, where the user can randomly access part of the data directly in the compressed domain. He also investigates the model with one decoder and two-sided state-information, i.e. when the encoder observes a degraded version of the decoder side information. Following up the work of Orlitsky and Roche (2001), he determines the appropriate graph to study the problem of zero-error coding for computing. During the Master internship, Nicolas Charpenay used to work on zero-error channel coding problem, introduced in Shannon (1956). This research work resulted in the following publications:

- [IC30] Charpenay, N., Le Treust, M. and Roumy, A., *Zero-error source coding when side information may be present*, International Zürich Seminar, March, 2022.
- [IC25] Charpenay, N. and Le Treust, M., *Zero-Error Coding with a Generator Set of Variable-Length Words*, IEEE International Symposium on Information Theory (ISIT), Los Angeles, June 2020.
- [NC11] Charpenay, N. and Le Treust, M. and Roumy, A., *Codage de source zéro-erreur pour un réseau de Gray-Wyner lorsque l'information adjacente peut être présente*, Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI), Nancy, France, 2022.
- [W6] Charpenay, N., Le Treust, M. and Roumy, A., *Zero-error source coding with two-sided side-information*, submitted to IEEE JSAIT, Apr. 2022.



[W4] Charpenay, N. and Le Treust, M. *Variable-Length Coding for Zero-Error Channel Capacity*, preliminary draft, [on-line] available: <https://arxiv.org/abs/2001.03523>, Jan. 2020.

## II.3 Content of this manuscript

This manuscript is organized as follows.

Chap. III presents some key results from the Information Theory literature such as the lossy source coding theorem of Shannon (1959), the channel coding theorem of Shannon (1948) and show the connection to the problem of rational inattention of Sims (2003) in Game Theory.

Chap. IV introduces the coordination problem of Gossner et al. (2006) and its reformulation in Cuff et al. (2010) and Cuff and Zhao (2011). We present the main contributions to the coordination problem described in Sec. II.1.b, when the channel between the encoder and the decoder is noisy.

Chap. V presents three different research directions that are extensions of the coordination results: 1) the channel state masking problem of Merhav and Shamai (2007) described in Sec. II.1.c, 2) the decentralized control problem of Witsenhausen (1968) described in Sec. II.1.d and 3) the strong coordination problem of Cuff et al. (2010) described in Sec. II.2.a.

Chap. VI details the contributions of Sec. II.1.e and II.2.b regarding the Bayesian persuasion game of Kamenica and Gentzkow (2011) with restricted communication, with decoder side information and with several decoders.

Chap. VII presents the on-going research work on zero-error source coding of II.2.c, on convex optimization algorithms for coordination problems, and on mechanism design and cheap talk problems. Our information theoretic approach will further be used to study repeated games with imperfect monitoring and incomplete information.

# Chapter III

## Information Theory review

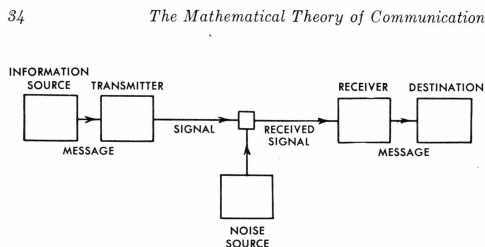


Fig. 1.— Schematic diagram of a general communication system.

(Fig. 1 in Shannon (1948))

In this chapter, we recall the lossy source coding result of Shannon (1959), the notion of channel capacity of Shannon (1948) and the “separation result” for the joint source-channel coding problem (El Gamal and Kim, 2011, Sec. 3.9). We discuss the connection with the problem of rational inattention of Sims (2003) and we present the optimality conditions of the distortion-rate function in (Steiner et al., 2017, Theorem 1). We extend the lossy source coding result to several cost functions.

### Contents

---

III.1 Shannon’s separation results . . . . .	22
III.1.a Perfect channel of restricted cardinality . . . . .	22
III.1.b General noisy channel . . . . .	24
III.1.c Entropy formulas . . . . .	25
III.2 Perfect channel of fixed capacity . . . . .	26
III.2.a Lossy source coding . . . . .	26
III.2.b Rational inattention . . . . .	27
III.2.c Multi-objective coding problem . . . . .	28

---

### III.1 Shannon's separation results

We recall some notations. We denote by  $\Delta(\mathcal{X})$  the set of probability distributions  $\mathcal{P}_X$  over the finite set  $\mathcal{X}$ , i.e. the symbol  $x \in \mathcal{X}$  is selected with probability  $\mathcal{P}_X(x) \in [0, 1]$ . We denote by  $\Delta(\mathcal{Y})^{|\mathcal{X}|}$  the set of conditional probability distributions  $\mathcal{P}_{Y|X}$ .

#### III.1.a Perfect channel of restricted cardinality

Consider a source of information  $(\mathcal{U}, \mathcal{P}_U)$  with  $|\mathcal{U}| = 3$  symbols and a perfect channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ , i.e.  $\mathcal{T}_{Y|X}(y|x) = \mathbb{1}\{x = y\}$ , with  $|\mathcal{X}| = |\mathcal{Y}| = 2$  symbols. Is it possible to convey a source of information through a perfect channel which has fewer symbols?

Information Theory provides an answer to this question by applying the law of large numbers and by considering a repeated version of this problem. Unlikely sequences having an incorrect empirical distribution  $\tilde{u}^n \in \mathcal{U}^n$  are assigned to a single sequence of channel symbols  $\tilde{x}^n \in \mathcal{X}^n$ , while the sequences with a correct empirical distribution  $u^n \in \mathcal{U}^n$  are perfectly recovered with high probability. The number of sequences with a correct empirical distribution scales exponentially with the Shannon's entropy of the source distribution  $\mathcal{P}_U \in \Delta(\mathcal{U})$ . We present seminal results of [Shannon \(1948, 1959\)](#) addressing this question.

In this section, we consider a source of information  $(\mathcal{U}, \mathcal{P}_U)$ , a perfect channel  $\mathcal{X}$  of fixed cardinality  $|\mathcal{X}|$ , a finite set of decoder actions  $\mathcal{V}$ , and a cost function

$$c : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}. \quad (\text{III.1})$$

The information source generates an i.i.d. sequence  $U^n$  of length  $n \in \mathbb{N}^*$ . The encoder observes the realization of the source sequence  $u^n \in \mathcal{U}^n$  and selects a channel inputs sequence  $x^n \in \mathcal{X}^n$ . The decoder

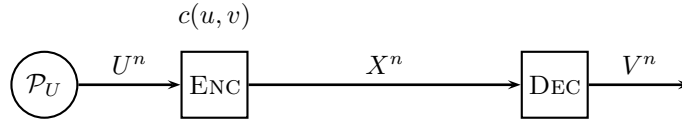


Figure III.1 – Perfect channel of restricted cardinality  $|\mathcal{X}|$ .

perfectly observes  $x^n \in \mathcal{X}^n$  and returns a sequence of actions  $v^n \in \mathcal{V}^n$ .

In this manuscript, we define the encoding and decoding functions by

$$\sigma : \mathcal{U}^n \longrightarrow \Delta(\mathcal{X}^n), \quad (\text{III.2})$$

$$\tau : \mathcal{X}^n \longrightarrow \Delta(\mathcal{V}^n). \quad (\text{III.3})$$

Note that the encoding and decoding functions are stochastic, i.e. they generate the sequences of random variables  $X^n$  and  $V^n$ . In some scenarios, e.g. in the Bayesian persuasion game of [Kamenica and Gentzkow \(2011\)](#) and in Physical Layer Security [Bloch and Barros \(2011\)](#), it is important to consider stochastic encoding and the decoding functions.

The coding scheme, depicted in Fig. III.1, induces a probability distribution over the sequences of symbols  $(U^n, X^n, V^n)$  given by

$$\Pr((U^n, X^n, V^n) = (u^n, x^n, v^n)) = \left( \prod_{t=1}^n \mathcal{P}_U(u_t) \right) \sigma(x^n | u^n) \tau(v^n | x^n), \quad \forall (u^n, x^n, v^n). \quad (\text{III.4})$$

**Definition 1** Given  $n \in \mathbb{N}^*$ , the minimum expected long-run cost is defined by

$$C_p^n = \min_{\sigma, \tau} \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n c(U_t, V_t) \right], \quad (\text{III.5})$$

where the probability distribution of the random variables  $(U_t, V_t)$ , for  $t \in \{1, \dots, n\}$ , is given by (III.4).

Note that the sequence  $(n \cdot C_p^n)_{n \in \mathbb{N}^*}$  is subadditive, i.e.  $\forall (n, m) \in \mathbb{N}^{*2}, (n+m)C_p^{n+m} \leq nC_p^n + mC_p^m$ . This property is obtained by considering the concatenation of the optimal coding scheme  $(\sigma, \tau)$  in (III.5) for some  $n \in \mathbb{N}^*$ , with the optimal coding scheme for some  $m \in \mathbb{N}^*$ .

**Definition 2** *The optimal single-letter cost is defined by*

$$C_p^* = \min_{\substack{\mathcal{Q}_{V|U}, \\ I(U;V) \leq \log_2 |\mathcal{X}|}} \mathbb{E}[c(U, V)], \quad (\text{III.6})$$

where the mutual information is defined by

$$I(U; V) = \sum_{(u,v) \in \text{supp } \mathcal{P}_U \mathcal{Q}_{V|U}} \mathcal{P}_U(u) \mathcal{Q}_{V|U}(v|u) \log_2 \frac{\mathcal{Q}_{V|U}(v|u)}{\sum_{u'} \mathcal{P}_U(u') \mathcal{Q}_{V|U}(v|u')}. \quad (\text{III.7})$$

Beside its definition in (III.7), the mutual information  $I(U; V) = D_{KL}(\mathcal{P}_U \mathcal{Q}_{V|U} \| \mathcal{P}_U \mathcal{Q}_V)$  is also the KL-divergence, see [Kullback and Leibler \(1951\)](#), between the joint distribution  $\mathcal{P}_U \mathcal{Q}_{V|U} \in \Delta(\mathcal{U} \times \mathcal{V})$  and the product of marginal distributions  $\mathcal{P}_U \mathcal{Q}_V \in \Delta(\mathcal{U}) \times \Delta(\mathcal{V})$ , where  $\mathcal{Q}_V = \sum_{u'} \mathcal{P}_U(u') \mathcal{Q}_{V|U}(\cdot|u')$ .

**Theorem 1**

$$\lim_{n \rightarrow +\infty} C_p^n = \inf_{n \in \mathbb{N}^*} C_p^n = C_p^*. \quad (\text{III.8})$$

The proof of Theorem 1, see [Shannon \(1948, 1959\)](#), consists of two parts:

1. The achievability proof, which relies on a coding construction, shows that  $\lim_{n \rightarrow +\infty} C_p^n \leq C_p^*$ .
2. The converse proof shows that  $C_p^* \leq \inf_{n \in \mathbb{N}^*} C_p^n$ .

The sequence  $(n \cdot C_p^n)_{n \in \mathbb{N}^*}$  is subadditive, [Fekete \(1923\)](#)'s Lemma shows that  $\inf_{n \in \mathbb{N}^*} C_p^n = \lim_{n \rightarrow +\infty} C_p^n$ . In conclusion, we have

$$C_p^* \leq \inf_{n \in \mathbb{N}^*} C_p^n = \lim_{n \rightarrow +\infty} C_p^n \leq C_p^*. \quad (\text{III.9})$$

Theorem 1 provides an answer to the question in the beginning of Sec. III.1.a. When  $n \in \mathbb{N}^*$  is large enough, it is possible to convey the source random variable  $U$  via the perfect channel  $\mathcal{X}$  with  $|\mathcal{X}| < |\mathcal{U}|$  if and only if  $H(U) \leq \log_2 |\mathcal{X}|$  where Shannon's entropy function is defined by

$$H(U) = \sum_{u \in \text{supp } \mathcal{P}_U} \mathcal{P}_U(u) \log_2 \frac{1}{\mathcal{P}_U(u)}. \quad (\text{III.10})$$

More precisely, we have the following proposition.

**Proposition 1** *Consider that  $\mathcal{V} = \mathcal{U}$  and  $c(u, v) = \mathbb{1}\{u \neq v\}$ .*

$$C_p^* = 0 \iff H(U) \leq \log_2 |\mathcal{X}|. \quad (\text{III.11})$$

*Proof.* [Prop. 1] We denote by  $\tilde{\mathcal{Q}}_{V|U}$  the distribution that achieves the minimum in (III.6). Therefore

$$0 = \mathbb{E}_{\tilde{\mathcal{Q}}_{V|U}}[\mathbb{1}\{U \neq V\}] = \sum_u \mathcal{P}_U(u) \sum_v \tilde{\mathcal{Q}}_{V|U}(v|u) \mathbb{1}\{u \neq v\} \iff \forall u \in \mathcal{U}, \quad \text{supp } \tilde{\mathcal{Q}}_{V|U}(\cdot|u) = \{u\}. \quad (\text{III.12})$$

This implies that the conditional entropy satisfies

$$H(U|V) = \sum_{(u,v) \in \text{supp } \mathcal{P}_U \tilde{\mathcal{Q}}_{V|U}} \mathcal{P}_U(u) \tilde{\mathcal{Q}}_{V|U}(v|u) \log_2 \frac{1}{\tilde{\mathcal{Q}}_{V|U}(v|u)} = \sum_u \mathcal{P}_U(u) \log_2 1 = 0, \quad (\text{III.13})$$

and therefore, by definition,

$$\log_2 |\mathcal{X}| \geq I(U; V) = H(U) - H(U|V) = H(U). \quad (\text{III.14})$$

The converse holds since  $I(U; V) \leq H(U)$  is true for all distribution  $\mathcal{Q}_{V|U} \in \Delta(\mathcal{V})^{|\mathcal{U}|}$ . Hence the above distribution  $\tilde{\mathcal{Q}}_{V|U}$  satisfies

$$I(U; V) \leq H(U) \leq \log_2 |\mathcal{X}|, \quad \mathbb{E}_{\tilde{\mathcal{Q}}_{V|U}} [\mathbf{1}\{U \neq V\}] = 0. \quad (\text{III.15})$$

□

**Example 1** Consider an information source with  $\mathcal{U} = \{0, 1, 2\}$ ,  $\mathcal{P}_U = (1 - \alpha, \alpha/2, \alpha/2)$  and the perfect channel  $\mathcal{X} = \{0, 1\}$ . Let  $\alpha^* \in [0, 1]$  the unique solution of  $1 - \alpha = h_b(\alpha)$  where the binary entropy function is defined by  $h_b(\alpha) = \alpha \log_2 \frac{1}{\alpha} + (1 - \alpha) \log_2 \frac{1}{1 - \alpha}$  for all  $\alpha \in ]0, 1[$  and  $h_b(\alpha) = 0$  for  $\alpha \in \{0, 1\}$ . Then for all  $\alpha \in [0, \alpha^*]$ , we have  $H(U) \leq \log_2 |\mathcal{X}|$  which implies that  $\lim_{n \rightarrow +\infty} C_p^n = \inf_{n \in \mathbb{N}^*} C_p^n = 0$ . This shows that the decoder can perfectly recover the information source.

### III.1.b General noisy channel

We consider a general noisy channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ , i.e. where  $\mathcal{T}_{Y|X} \neq \mathbf{1}\{x = y\}$ . The encoder wants to send  $n \in \mathbb{N}^*$  symbols of source  $(\mathcal{U}, \mathcal{P}_U)$  by using  $k \in \mathbb{N}^*$  times the channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ . The encoding and decoding functions are defined by

$$\sigma : \mathcal{U}^n \longrightarrow \Delta(\mathcal{X}^k), \quad (\text{III.16})$$

$$\tau : \mathcal{Y}^k \longrightarrow \Delta(\mathcal{V}^n). \quad (\text{III.17})$$

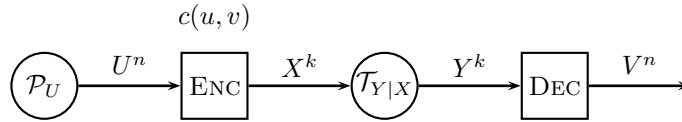


Figure III.2 – Point-to-point scenario with  $n \in \mathbb{N}^*$  uses of the source  $(\mathcal{U}, \mathcal{P}_U)$  and  $k \in \mathbb{N}^*$  uses of the channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ .

The coding scheme, depicted in Fig. III.2, induces a probability distribution given by for all  $(u^n, x^k, y^k, v^n)$

$$\Pr((U^n, X^k, Y^k, V^n) = (u^n, x^k, y^k, v^n)) = \left( \prod_{t=1}^n \mathcal{P}_U(u_t) \right) \sigma(x^k | u^n) \left( \prod_{t=1}^k \mathcal{T}_{Y|X}(y_t | x_t) \right) \tau(v^n | y^k). \quad (\text{III.18})$$

**Definition 3** Given  $n \in \mathbb{N}^*$ , the minimum expected long-run cost is defined by

$$C_c^{n,k} = \min_{\sigma, \tau} \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n c(U_t, V_t) \right]. \quad (\text{III.19})$$

**Definition 4** Given a parameter  $r \geq 0$ , we define the optimal single-letter cost by

$$C_c^*(r) = \min_{\substack{\mathcal{Q}_{V|U}, \\ I(U; V) \leq r, \max_{\mathcal{P}_X} I(X; Y)}} \mathbb{E} [c(U, V)]. \quad (\text{III.20})$$

The channel capacity is defined by

$$\max_{\mathcal{P}_X} I(X; Y) = \max_{\mathcal{P}_X} \sum_{(x,y) \in \text{supp } \mathcal{P}_X \mathcal{T}_{Y|X}} \mathcal{P}_X(x) \mathcal{T}_{Y|X}(y|x) \log_2 \frac{\mathcal{T}_{Y|X}(y|x)}{\sum_{x'} \mathcal{P}_X(x') \mathcal{T}_{Y|X}(y|x')}, \quad (\text{III.21})$$

where the maximum is taken over the set of probability distribution  $\Delta(\mathcal{X})$ . Note that when the channel is perfect, i.e.  $\mathcal{T}_{Y|X}(y|x) = \mathbb{1}\{x = y\}$  for all  $(x, y)$ , the channel capacity satisfies

$$\max_{\mathcal{P}_X} I(X; Y) = \max_{\mathcal{P}_X} H(X) = \log_2 |\mathcal{X}|, \quad (\text{III.22})$$

and we recover the result of Theorem 1.

We now introduce the notation  $\lfloor x \rfloor$  for the floor function, i.e. the greatest integer less than or equal to  $x \in \mathbb{R}$ .

**Theorem 2** 1. For all  $(n, k) \in \mathbb{N}^{*2}$ , we have

$$C_c^* \left( \frac{k}{n} \right) \leq C_c^{n,k}. \quad (\text{III.23})$$

2. Suppose that  $k = \lfloor rn \rfloor$  for some parameter  $r \geq 0$ , we have

$$\lim_{n \rightarrow +\infty} C_c^{n,k} \leq C_c^*(r). \quad (\text{III.24})$$

The Theorem 1 is referred to as the “separation result” for the joint source-channel coding problem in (El Gamal and Kim, 2011, Sec. 3.9). For  $k = \lfloor rn \rfloor$  with  $r \geq 0$ , Theorem 2(1.&2.) implies that

$$\lim_{n \rightarrow +\infty} C_c^{n,k} = \inf_{n \in \mathbb{N}^*} C_c^{n,k} = C_c^*(r). \quad (\text{III.25})$$

### III.1.c Entropy formulas

We recall the definitions of the entropy function and the mutual information. We present some properties, the proofs are detailed in (Cover and Thomas, 2006, Chap. 2), (El Gamal and Kim, 2011, Chap. 2), (Csiszár and Körner, 2011, Chap. 1).

**Definition 5** Given a distribution  $\mathcal{P}_{UV} \in \Delta(\mathcal{U} \times \mathcal{V})$  we define the mutual information  $I(U; V)$ , the conditional entropy  $H(U|V)$  and the entropy  $H(U, V)$  by

$$I(U; V) = \sum_{(u,v) \in \text{supp } \mathcal{P}_{UV}} \mathcal{P}_{UV}(u, v) \log_2 \frac{\mathcal{P}_{UV}(u, v)}{\mathcal{P}_U(u) \mathcal{P}_V(v)}, \quad (\text{III.26})$$

$$H(U|V) = \sum_{(u,v) \in \text{supp } \mathcal{P}_{UV}} \mathcal{P}_{UV}(u, v) \log_2 \frac{1}{\mathcal{P}_{U|V}(u|v)}, \quad (\text{III.27})$$

$$H(U, V) = \sum_{(u,v) \in \text{supp } \mathcal{P}_{UV}} \mathcal{P}_{UV}(u, v) \log_2 \frac{1}{\mathcal{P}_{UV}(u, v)}. \quad (\text{III.28})$$

For a distribution  $\mathcal{P}_{UVW} \in \Delta(\mathcal{U} \times \mathcal{V} \times \mathcal{W})$ , the conditional mutual information  $I(U; V|W)$  is defined by

$$I(U; V|W) = \sum_{(u,v,w) \in \text{supp } \mathcal{P}_{UVW}} \mathcal{P}_{UVW}(u, v, w) \log_2 \frac{\mathcal{P}_{UV|W}(u, v|w)}{\mathcal{P}_{U|W}(u|w) \mathcal{P}_{V|W}(v|w)}. \quad (\text{III.29})$$

These formulas are related through the chain rule for the entropy.

**Proposition 2** Given a discrete probability distribution  $\mathcal{P}_{UV} \in \Delta(\mathcal{U} \times \mathcal{V})$ , we have

$$0 \leq H(U, V) = H(V) + H(U|V) = H(U) + H(V|U) \leq \log_2 |\mathcal{U} \times \mathcal{V}|, \quad (\text{III.30})$$

$$\begin{aligned} 0 \leq I(U; V) &= H(V) - H(V|U) = H(U) - H(U|V) \\ &= H(U) + H(V) - H(U, V) = D_{KL}(\mathcal{P}_{UV} || \mathcal{P}_U \mathcal{P}_V) \leq \min(H(U), H(V)), \end{aligned} \quad (\text{III.31})$$

where  $D_{KL}(\mathcal{P}_{UV} || \mathcal{P}_U \mathcal{P}_V)$  denotes the KL-divergence *Kullback and Leibler (1951)*. We have  $H(U, V) = \log_2 |\mathcal{U} \times \mathcal{V}|$  if and only if  $\mathcal{P}_{UV}$  is the uniform distribution.

**Proposition 3** For a distribution  $\mathcal{P}_{UVW} \in \Delta(\mathcal{U} \times \mathcal{V} \times \mathcal{W})$ , we have

$$\begin{aligned} 0 \leq I(U; V|W) &= H(V|W) - H(V|U, W) = H(U|W) - H(U|V, W) \\ &= H(U|W) + H(V|W) - H(U, V|W), \end{aligned} \quad (\text{III.32})$$

$$0 \leq I(U; V, W) = I(U; W) + I(U; V|W) = I(U; V) + I(U; W|V). \quad (\text{III.33})$$

**Definition 6** A distribution  $\mathcal{P}_{UVW} \in \Delta(\mathcal{U} \times \mathcal{V} \times \mathcal{W})$  satisfies the Markov chain property  $U \dashv\vdash V \dashv\vdash W$  if

$$\mathcal{P}_{U|VW} = \mathcal{P}_{U|V} \iff \mathcal{P}_{W|UV} = \mathcal{P}_{W|V}. \quad (\text{III.34})$$

**Proposition 4** A distribution  $\mathcal{P}_{UVW} \in \Delta(\mathcal{U} \times \mathcal{V} \times \mathcal{W})$  satisfies the Markov chain property  $U \dashv\vdash V \dashv\vdash W$  if and only if

$$I(U; W|V) = 0. \quad (\text{III.35})$$

## III.2 Perfect channel of fixed capacity

### III.2.a Lossy source coding

In this section, we consider the lossy source coding problem introduced in *Shannon (1959)*. The channel is perfect and has a fixed capacity  $R \geq 0$ , also referred to as the “rate of the code”. This parameter  $R \geq 0$  is the increase rate of the number of messages  $2^{\lfloor nR \rfloor}$  used by the coding scheme.

**Definition 7** Given  $n \in \mathbb{N}^*$  and  $R \geq 0$ , a  $(R, n)$ -code is a pair of functions  $(\sigma, \tau)$  defined by

$$\sigma : \mathcal{U}^n \longrightarrow \Delta(\{1, \dots, 2^{\lfloor nR \rfloor}\}), \quad (\text{III.36})$$

$$\tau : \{1, \dots, 2^{\lfloor nR \rfloor}\} \longrightarrow \Delta(\mathcal{V}^n). \quad (\text{III.37})$$

We denote by  $\mathcal{C}(R, n)$  the set of  $(R, n)$ -codes.

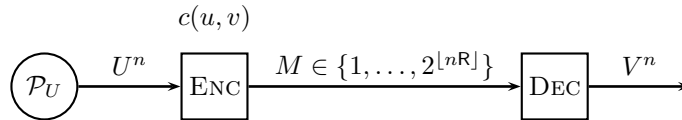


Figure III.3 – The source  $(\mathcal{U}, \mathcal{P}_U)$  and the perfect channel of rate  $R \geq 0$ .

Note that  $R = \log_2 |\mathcal{X}|$  is equivalent to  $2^{nR} = |\mathcal{X}^n|$ . Note also that  $\frac{1}{n} \lfloor nR \rfloor \longrightarrow R$ , when  $n \rightarrow +\infty$ . A  $(R, n)$ -code  $(\sigma, \tau) \in \mathcal{C}(R, n)$  induces a probability distribution given by,

$$\Pr\left((U^n, M, V^n) = (u^n, m, v^n)\right) = \left(\prod_{t=1}^n \mathcal{P}_U(u_t)\right) \sigma(m|u^n) \tau(v^n|m), \quad \forall (u^n, m, v^n). \quad (\text{III.38})$$

**Definition 8** Given  $n \in \mathbb{N}^*$ , the minimum expected long-run cost is defined by

$$C^n(\mathbb{R}) = \min_{(\sigma, \tau) \in \mathcal{C}(\mathbb{R}, n)} \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n c(U_t, V_t) \right]. \quad (\text{III.39})$$

The sequence  $(n \cdot C^n(\mathbb{R}))_{n \in \mathbb{N}^*}$  is subadditive. Since  $\lfloor n\mathbb{R} \rfloor + \lfloor m\mathbb{R} \rfloor \leq \lfloor (n+m)\mathbb{R} \rfloor$  for all  $(n, m) \in \mathbb{N}^{*2}$ , the concatenation of  $(\sigma^n, \tau^n) \in \mathcal{C}(\mathbb{R}, n)$  with  $(\sigma^m, \tau^m) \in \mathcal{C}(\mathbb{R}, m)$  generates a strategy which belongs to  $\mathcal{C}(\mathbb{R}, n+m)$ .

**Definition 9** The optimal single-letter cost is defined by

$$C^*(\mathbb{R}) = \min_{\substack{\mathcal{Q}_{V|U}, \\ I(U;V) \leq \mathbb{R}}} \mathbb{E}[c(U, V)]. \quad (\text{III.40})$$

When the cost function  $c(u, v)$  is a distortion function  $d(u, v)$ , the expression  $C^*(\mathbb{R})$  is referred to as the ‘‘distortion-rate function’’, see (Cover and Thomas, 2006, pp. 306). We recover  $C_p^*$ , resp.  $C_n^*$ , by replacing by replacing  $\mathbb{R} = \log_2 |\mathcal{X}|$ , resp.  $\mathbb{R} = \max_{\mathcal{P}_X} I(X; Y)$ , in the expression of  $C^*(\mathbb{R})$ .

**Theorem 3**

$$\lim_{n \rightarrow +\infty} C^n(\mathbb{R}) = \inf_{n \in \mathbb{N}^*} C^n(\mathbb{R}) = C^*(\mathbb{R}). \quad (\text{III.41})$$

Theorem 3 is stated in Shannon (1959), a proof can be found in (El Gamal and Kim, 2011, Sec. 3.6).

### III.2.b Rational inattention

The expression  $C^*(\mathbb{R})$  of (III.40) is related to the rational inattention problem in (Sims, 2003, Sec. 3), in the Game Theory literature. The rate parameter  $\mathbb{R} \geq 0$  captures the information-processing capacity of a decision maker who faces a signal extraction problem. She decides how much effort to put in order get a finer information about the source  $(\mathcal{U}, \mathcal{P}_U)$ . The conditional distribution  $\mathcal{Q}_{V|U}$  determines the signal extraction strategy. The decision maker is rationally inattentive when she optimizes the trade-off between the information extraction cost  $I(U; V)$  and the information exploitation to lower  $\mathbb{E}[c(U, V)]$ .

In (Steiner et al., 2017, Eq. (3)-(5)), the authors formulate the Lagrangian of  $C^*(\mathbb{R})$ . For all conditional distribution  $\mathcal{Q}_{V|U} \in \Delta(\mathcal{V})^{|\mathcal{U}|}$  and Lagrange parameter  $\lambda \geq 0$ , they define

$$\mathcal{L}(\mathcal{Q}_{V|U}, \lambda) = \mathbb{E}[c(U, V)] + \lambda I(U; V). \quad (\text{III.42})$$

Note that

$$C^*(\mathbb{R}) = \min_{\mathcal{Q}_{V|U}} \max_{\lambda \geq 0} \left( \mathcal{L}(\mathcal{Q}_{V|U}, \lambda) - \lambda \mathbb{R} \right). \quad (\text{III.43})$$

For a given Lagrange parameter  $\lambda \geq 0$ , the problem of interests writes

$$\min_{\mathcal{Q}_{V|U}} \mathcal{L}(\mathcal{Q}_{V|U}, \lambda). \quad (\text{III.44})$$

The optimality conditions are stated in (Steiner et al., 2017, Theorem 1).

**Theorem 4** We denote by  $\mathcal{Q}_{V|U}^*$  the optimal solution to problem (III.44) with marginal distribution  $\mathcal{Q}_V^* = \sum_u \mathcal{P}_U(u) \mathcal{Q}_{V|U}^*(\cdot|u) \in \Delta(\mathcal{V})$ . For all  $(u, v) \in \mathcal{U} \times \mathcal{V}$ , the optimal solution satisfies

$$\mathcal{Q}_{V|U}^*(v|u) = \frac{\mathcal{Q}_V^*(v) \exp\left(-\frac{c(u,v)}{\lambda}\right)}{\sum_{v'} \mathcal{Q}_V^*(v') \exp\left(-\frac{c(u,v')}{\lambda}\right)}. \quad (\text{III.45})$$

A similar characterization of  $\mathcal{Q}_{V|U}^*$  is proposed in (Blahut, 1972, Theorem 5) for the related problem

$$\mathbb{R}^*(C) = \min_{\substack{\mathcal{Q}_{V|U}, \\ \mathbb{E}[c(U, V)] \leq C}} I(U; V), \quad (\text{III.46})$$

which depends of a target cost value  $C \in \mathbb{R}$ .



### III.2.c Multi-objective coding problem

In this section, we consider a coding problem in which the encoder and the decoder are endowed with distinct cost functions

$$c_e : \mathcal{U} \times \mathcal{V} \longrightarrow \mathbb{R} \quad \neq \quad c_d : \mathcal{U} \times \mathcal{V} \longrightarrow \mathbb{R}. \quad (\text{III.47})$$

In this section, we assume that a unique decision-maker selects the pair of encoding and decoding strategies  $(\sigma, \tau)$ . Therefore the decision problem has no strategic feature. This problem is not a game, it is a multi-objective coding problem.

**Definition 10** *Given the pair of strategies  $(\sigma, \tau) \in \mathcal{C}(\mathbb{R}, n)$ , the encoder and decoder long-run cost functions  $c_e^n(\sigma, \tau)$  and  $c_d^n(\sigma, \tau)$  are defined by*

$$c_e^n(\sigma, \tau) = \sum_{u^n, v^n} \mathcal{P}_{U^n V^n}^{\sigma, \tau}(u^n, v^n) \cdot \left[ \frac{1}{n} \sum_{t=1}^n c_e(u_t, v_t) \right], \quad (\text{III.48})$$

$$c_d^n(\sigma, \tau) = \mathbb{E}_{\sigma, \tau} \left[ \frac{1}{n} \sum_{t=1}^n c_d(U_t, V_t) \right]. \quad (\text{III.49})$$

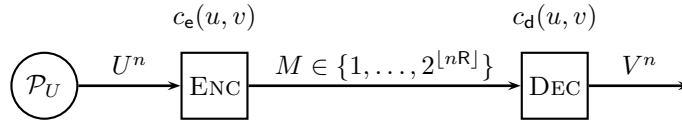


Figure III.4 – Two cost functions  $c_e : \mathcal{U} \times \mathcal{V} \longrightarrow \mathbb{R}$  and  $c_d : \mathcal{U} \times \mathcal{V} \longrightarrow \mathbb{R}$ .

**Definition 11** *The pair  $(C_e^*, C_d^*)$  is achievable if*

$$\forall \varepsilon > 0, \exists \bar{n} \in \mathbb{N}^*, \forall n \geq \bar{n}, \exists (\sigma, \tau) \in \mathcal{C}(\mathbb{R}, n), \quad |C_e^* - c_e^n(\sigma, \tau)| + |C_d^* - c_d^n(\sigma, \tau)| \leq \varepsilon. \quad (\text{III.50})$$

We denote by  $\mathcal{A}(\mathbb{R})$  the set of achievable pairs  $(C_e, C_d)$ .

**Theorem 5** *Given  $R \geq 0$ ,*

$$\mathcal{A}(\mathbb{R}) = \left\{ \left( \mathbb{E}_{\mathcal{Q}}[c_e(U, V)], \mathbb{E}_{\mathcal{Q}}[c_d(U, V)] \right), \quad \mathcal{Q} \in \mathbb{Q}(\mathbb{R}) \right\}, \quad (\text{III.51})$$

where the set of probability distribution is defined by

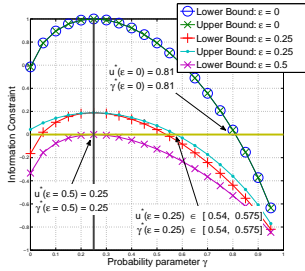
$$\mathbb{Q}(\mathbb{R}) = \left\{ \mathcal{P}_U \mathcal{Q}_{V|U}, \quad I(U; V) \leq R \right\}. \quad (\text{III.52})$$

The proof of Theorem 5 is a consequence of the lossy source coding theorem of Shannon (1959) with two cost functions. For the converse proof, we recall that the information constraint  $I(U; V) \leq R$  needs to be satisfied. For the achievability proof, given a probability distribution  $\mathcal{P}_U \mathcal{Q}_{V|U} \in \mathbb{Q}$ , we use the standard arguments to show the existence of a coding scheme that achieves the pair of costs  $(\mathbb{E}_{\mathcal{Q}}[c_e(U, V)], \mathbb{E}_{\mathcal{Q}}[c_d(U, V)])$ .

The key aspect of the solution is the set of achievable probability distributions  $\mathbb{Q}$ , which is the main object of study of Chap. IV.

# Chapter IV

## Empirical coordination with a strictly causal or a causal decoder



In the previous chapter, we have seen that the restrictions of communication are captured by an entropy constraint that affects the set of probability distributions. In a decentralized network, it is essential that the devices or the agents cooperate and coordinate their actions in order to achieve a common goal, to ensure the reliability and the stability of the network. More specifically, it is required that the sequences of actions and states follow a prescribed empirical distribution. This problem was investigated in [Gossner et al. \(2006\)](#) when the encoder actions are

perfectly observed by the decoder. We extend these results by assuming that the decoder imperfectly monitors encoder actions. Our point-to-point model involves an information source, a noisy channel and a common objective function for the encoder and the decoder. We wonder how this simple network operates when the devices have incomplete information about the state and imperfect monitoring of the actions. In this chapter, we present various empirical coordination problems under study in ([Le Treust, 2017, \[J6\]](#)), ([Le Treust, 2015b, \[IC14\]](#)), ([Le Treust, 2015c, \[IC13\]](#)), ([Le Treust, 2014, \[IC12\]](#)), ([Le Treust, 2015a, \[NC6\]](#)) ([Le Treust, 2015, \[W1\]](#)),

### Contents

---

IV.1 Introduction . . . . .	30
IV.2 Empirical coordination with strictly causal decoder . . . . .	33
IV.2.a Problem statement . . . . .	33
IV.2.b Characterization of the set of achievable distributions . . . . .	34
IV.2.c Evaluation of an arbitrary expected cost function . . . . .	35
IV.2.d Example #1: coordination game of <a href="#">Gossner et al. (2006)</a> . . . . .	36
IV.2.e Example #2: Hamming distortion and channel cost . . . . .	38
IV.3 General decoding functions . . . . .	40
IV.3.a Causal decoding with source feedforward . . . . .	40
IV.3.b Strictly causal decoding with source feedforward . . . . .	42
IV.4 State-dependent source and state-dependent channel . . . . .	43
IV.4.a Non-causal encoder and non-causal decoder . . . . .	43
IV.4.b Causal decoder and non-causal encoder . . . . .	47
IV.4.c Causal encoder and non-causal decoder . . . . .	49
IV.5 Chapter summary . . . . .	51

---

## IV.1 Introduction

A decentralized network is composed of communication devices that sense their environment and that choose autonomously the best transmission scheme to implement. The decision process is decentralized and does not require a central controller. However, it is essential that the communication devices cooperate and coordinate their actions, in order to ensure the reliability of the transmissions and the stability of the network. We investigate the problem of the coordination of two autonomous devices, by considering a point-to-point model, represented by Fig. IV.1, with an information source and a noisy channel. In the classical scenario, both encoder and decoder have the same objective: to implement a reliable transmission scheme. We wonder how this simple network operates when the devices are autonomous and try to coordinate their actions in order to achieve a common objective.

We study this problem using a two-step approach. First, we characterize the set of empirical distributions of actions and source symbols that are achievable via a coding scheme. Second, we optimize the expected common cost function with respect to the set of achievable distributions. In this chapter, we focus on the first task, i.e. determining the coordination possibilities of the encoder and the decoder by characterizing the set of achievable distributions.

Based on their observations, the encoder and the decoder choose the sequences of channel input and decoder's output. We require that the empirical distribution of all the sequences of symbols, converges to a target joint probability distribution. The aim is to determine the minimal amount of information to exchange such that the symbols of both transmitters are coordinated with the symbols of the source and of the channel. From an information theoretic point of view, this problem is closely related to the joint source-channel coding problem with two-sided state information and correlated source and state (Le Treust, 2015c, [IC13]). We characterize the set of achievable joint probability distributions using a single-letter information constraint that is related to the compression and to the transmission of information.

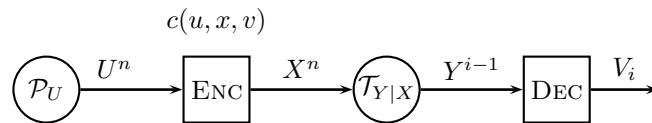


Figure IV.1 – Empirical coordination problem for the source  $(\mathcal{U}, \mathcal{P}_U)$ , the channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ , the cost function  $c : \mathcal{U} \times \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  and the decoder is strictly causal, i.e. at each instant  $t \in \{1, \dots, n\}$ , it returns a symbol  $V_t = \tau_t(Y^{t-1})$ . The encoder and the decoder implement a coding scheme such that, the empirical distribution of symbols  $Q_{UXYV}^n$  are close to the target joint probability distribution  $Q_{UXYV}$ , high probability.

We consider that the devices are equipped with a cost function, capturing their objective. The set of achievable values of the cost function is the image, by the expectation operator, of the set of achievable probability distributions. A cost value is achievable if and only if it corresponds to the expected cost, with respect to some achievable probability distribution. This approach simplifies the optimization of the long-run cost function, since the whole set of possible codes of large block-length reduces to the set of achievable target probability distributions, expressed with a single-letter formulation. As a particular case, our results boil down to the classical results of Shannon (1948) and Shannon (1959), when considering the minimal distortion for the information source or the minimal cost for the channel inputs. The main difference with the results of Shannon (1948) and Shannon (1959) is In this chapter, we consider a cost function that is common to both encoder and decoder.

### Related literature

The notion of target probability distribution has been proposed in Wyner (1975) for determining the common information of two correlated random variables. In the framework of quantum coding, the authors of Kramer and Savari (2002) and Kramer and Savari (2007) prove the existence of a code with

minimal amount of information exchanged, such that the empirical distribution of the sequences of symbols is close to the target probability distribution. The problem of empirical coordination is studied in Cover and Permuter (2007) for a three-node cascade network, for a multi-agent chain and for a multiple description setting. A stronger definition of coordination is considered in Cuff (2008) and Cuff (2013) for the problem of simulating and synthesizing a discrete memoryless channel, also related to the “reverse Shannon Theorem” Bennett et al. (2002) and “channel resolvability” Han and Verdú (1993). The concept of “coordination capacity” is introduced in Cuff et al. (2010), as a measure of the minimal amount of information transmitted, such that the nodes of a network can coordinate their actions. The authors consider two different notions of coordination, referred to as empirical coordination and strong coordination. For some networks, both notions coincide if the nodes have enough common randomness. Coordination over a network is also related to the multi-agent control problem with common randomness Anantharam and Borkar (2007). In Gohari and Anantharam (2011), the authors investigate strong coordination of the actions, assuming that the nodes have multiple rounds of noise-free communication. Empirical distribution of sub-codewords of length  $k$  of a *good code* is considered in Weissman and Ordentlich (2004) and Weissman and Ordentlich (2005), and the authors prove that it converges to the product of the optimal probability distribution Shamaï and Verdú (1997). Polar codes are under investigation for empirical coordination in Blasco-Serrano et al. (2012), and for strong coordination in Bloch et al. (2012). In Chou et al. (2015), the authors provide encoding and decoding algorithms that are based on polar codes, and that achieve the empirical and the strong coordination capacity. In (Cervia et al., 2016, [IC17]), the authors construct a polar code for empirical coordination with a noisy channel. Empirical coordination for triangular multi-terminal network is investigated in Berezhi et al. (2013). Strong coordination is studied for a multi-hop line network in Vellambi et al. (2015) and Vellambi et al. (2016), for a three-terminal line network in Satpathy and Cuff (2013) and Bloch and Kliever (2013), for a three-terminal relay network in Bloch and Kliever (2014), for two-way communication with a relay in Haddadpour et al. (2012), and for signal’s coordination in (Cervia et al., 2017b, [IC19]). The source coding problem of Ahlswede-Körner Ahlswede and Körner (1975) is investigated in Goldfeld et al. (2014), with a coordination requirement. The results for empirical coordination are extended to general alphabets in Raginsky (2010) and Raginsky (2013), by considering standard Borel spaces. The problems of zero-error coordination Abroshan et al. (2015) and of strong coordination with an evaluation function Orlitsky and Roche (2001) are both related to graph theory.

The problem of the coordination of actions is also investigated in the Game Theory literature Gossner and Vieille (2002), Gossner and Tomala (2006), Gossner and Tomala (2007), Gossner et al. (2009), Gossner et al. (2006) using the notion of *implementable probability distribution*, that is related to empirical coordination. In Gossner et al. (2006), the authors consider a point-to-point scenario with an encoder that observes the sequence of symbols of source, called “state of the nature”, and that chooses a sequence of actions. The channel is perfect and the decoder is strictly causal *i.e.*, it returns an action based on the observation of the past actions of the encoder and past symbols of the source. The objective is to coordinate the actions of both players together with the symbols of the source. The main difference with the settings described previously is that the channel inputs are also coordinated with the symbols of the source and the decoder’s actions. The encoder chooses a sequence of channel inputs that conveys some information and that is coordinated with the sequences of symbols of source. The authors characterize the set of implementable target joint probability distributions and evaluate the long-run cost function of the players, by considering the expected cost. The results of Gossner et al. (2006) are extended in Larrousse and Lasaulce (2013), Larrousse et al. (2018), by considering a noisy channel. The authors characterize the set of implementable probability distributions and apply their result to the interference channel in which the power control is used to encode embedded data about the channel state information. This approach is further applied to the two-way channel in Larrousse et al. (2015a), and to the case of causal encoding and decoding in Larrousse et al. (2015b). The results of Gossner et al. (2006) have also been extended in Cuff and Zhao (2011) by considering the notion of empirical coordination and by removing the observation by the decoder of the past symbols of source. The tools for empirical coordination with a cascade of controllers Cuff and Zhao (2011) are also used in Asnani et al. (2013), for the problem of cooperation in multi-terminal source coding with strictly causal, causal, and non-causal cribbing. In (Le Treust et al., 2011, [IC7]), the authors investigate the empirical correlation for two dependent sources and a broadcast channel with an additional secrecy constraint. The problem of empirical coordination for a joint source-

channel coding problem is solved in Cuff and Schieler (2011), for strictly causal and causal encoder with non-causal decoder. These results are based on hybrid coding Minero et al. (2011), Lim et al. (2010), and are closely related to the problem of state communication under investigation in Choudhuri et al. (2010), Choudhuri et al. (2011), Choudhuri and Mitra (2012a). The results stated in Cuff and Schieler (2011) are extended in (Le Treust, 2015b, [IC14]) by considering channel feedback available at the encoder. Channel feedback improves the coordination possibilities and simplifies the information constraint. For this problem, the authors of (Le Treust and Bloch, 2016, [IC16]) characterize of the region of achievable triple of information rate, empirical distribution and state leakage, with and without state information and noisy channel feedback. The problem of empirical coordination for non-causal encoder and decoder is not yet completely solved, but the optimal solutions are characterized in (Le Treust, 2014, [IC12]) for lossless decoding and in (Le Treust, 2015c, [IC13]) for perfect channel and for independent source and channel, based the separation result of Merhav and Shamai (2003). The duality Cover and Chiang (2002) between the channel coding of Gel'fand Pinsker Gel'fand and Pinsker (1980) and the source coding of Wyner Ziv Wyner and Ziv (1976) induces some similarities in the information constraints for lossless decoding (Le Treust, 2014, [IC12]) and for perfect channel (Le Treust, 2015c, [IC13]). This open problem is closely related to the problem of non-causal state communication, under investigation in Sutivong et al. (2005), Choudhuri et al. (2013) and Sutivong (2003). The problem of empirical coordination is a first step towards a better understanding of decentralized communication networks, in which the devices have different cost functions (Le Treust and Tomala, 2016, [IC18]) and choose autonomously the transmission power (Le Treust and Lasaulce, 2010, [J1]), Belmega et al. (2009) and the transmission rates Berry and Tse (2011), Perlaza et al. (2012), depending on their observation structure (Le Treust and Lasaulce, 2013, [J2]).

In this chapter, we investigate a point-to-point coordination problem involving an i.i.d. information source and a memoryless channel, represented by Fig. IV.1. The encoder and the decoder choose their sequences of actions *i.e.*, channel input and decoder output, so that the empirical distribution of the symbols converges to a target probability distribution. We assume that the decoder is strictly causal or causal *i.e.*, at each instant, it returns a symbol, also called an action, based on the observation of the past and current channel outputs. This *on-line* coordination assumption is related to the game theoretical framework of Gossner et al. (2006), in which the encoder and the decoder are the players that choose their actions simultaneously, based on their past observations. We characterize the set of achievable target joint probability distributions for non-causal encoder and strictly causal decoder and we relate the corresponding information constraint to the previous results from the literature, especially with the problem of source distortion and channel cost (El Gamal and Kim, 2011, pp. 47, 57 and 66). We analyze the optimization of some cost function over the set of achievable target probability distributions and we prove that this problem is convex. We also characterize the information constraint corresponding to causal decoder instead of strictly causal decoder. In that case, the actions of the decoder may also be coordinated with the *current* channel output.

This chapter is organized as follows. The coordination problem with strictly causal decoder is under study in Sec. IV.2. The system model is presented in Sec. IV.2.a. In Sec. V.2.b, we characterize the set of achievable target probability distributions for non-causal encoder and strictly causal decoder. In Sec. VII.2, we characterize the set of achievable utilities and we prove that the corresponding optimization problem is convex. We investigate two examples: the coordination game in Sec. IV.2.d and the trade-off between source distortion and channel cost in Sec. IV.2.e. In Sec. IV.3, we investigate the coordination problem for general decoding functions, with causal decoding and source feedforward in Sec. IV.3.a, with strictly causal decoding and source feedforward in Sec. IV.3.b. In Sec. IV.4, we investigate the coordination problem for state-dependent source and channel. The case of non-causal encoder and non-causal decoder under study in Sec. IV.4.a is still an open problem. We characterize the sets of achievable distributions when either the decoder is causal in Sec. IV.4.b, or the encoder is causal in Sec. IV.4.c.

## IV.2 Empirical coordination with strictly causal decoder

### IV.2.a Problem statement

We consider the setting of Fig. IV.1 in which a sequence of random variables  $U^n \in \mathcal{U}^n$  is drawn i.i.d. according to  $\mathcal{P}_U \in \Delta(\mathcal{U})$ , as in (IV.1). The non-causal encoder observes  $U^n \in \mathcal{U}^n$  and sends a sequence of channel inputs  $X^n \in \mathcal{X}^n$ . The sequence of channel outputs  $Y^n \in \mathcal{Y}^n$  is drawn i.i.d. according to  $\mathcal{T}_{Y|X} \in \Delta(\mathcal{Y})^{|\mathcal{X}|}$ , as in (IV.2).

$$\Pr(U^n = u^n) = \prod_{t=1}^n \mathcal{P}_U(u_t), \quad \forall u^n \in \mathcal{U}^n, \quad (\text{IV.1})$$

$$\Pr(Y^n = y^n | X^n = x^n) = \prod_{t=1}^n \mathcal{T}_{Y|X}(y_t | x_t), \quad \forall (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n. \quad (\text{IV.2})$$

We consider that the decoder is strictly causal. At instant  $t \in \{1, \dots, n\}$ , it observes the sequence of past channel outputs  $Y^{t-1} = (Y_1, \dots, Y_{t-1}) \in \mathcal{Y}^{t-1}$  and returns an output symbol  $V_t \in \mathcal{V}$ .

**Definition 12** Given  $n \in \mathbb{N}^*$ , we define the  $n$ -length code with non-causal encoder and strictly-causal decoder by a pair  $(\sigma, \tau)$  where  $\tau = \{\tau_t\}_{t=1}^n$  and

$$\sigma : \mathcal{U}^n \longrightarrow \mathcal{X}^n, \quad (\text{IV.3})$$

$$\tau_t : \mathcal{Y}^{t-1} \longrightarrow \mathcal{V}, \quad t \in \{1, \dots, n\}. \quad (\text{IV.4})$$

We denote by  $\mathcal{C}(n)$  the set of  $n$ -length code with non-causal encoder and strictly-causal decoder.

The code  $(\sigma, \tau) \in \mathcal{C}(n)$  induces a probability distribution  $\mathcal{P}_{U^n X^n Y^n V^n}^{\sigma\tau}$  defined, for all  $(u^n, x^n, y^n, v^n)$ , by

$$\mathcal{P}_{U^n X^n Y^n V^n}^{\sigma\tau}(u^n, x^n, y^n, v^n) = \left( \prod_{t=1}^n \mathcal{P}_U(u_t) \right) \mathbb{1}\{x^n = \sigma(u^n)\} \left( \prod_{t=1}^n \mathcal{T}_{Y|X}(y_t | x_t) \right) \left( \prod_{t=1}^n \mathbb{1}\{v_t = \tau_t(y^{t-1})\} \right). \quad (\text{IV.5})$$

We evaluate the empirical frequencies of symbols induced by the code.

**Definition 13** Given the sequences of realizations  $(u^n, x^n, y^n, v^n)$ , the empirical distribution  $q_{U^X Y^V}^n \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  is defined by

$$q_{U^X Y^V}^n(u, x, y, v) = \frac{1}{n} \sum_{t=1}^n \mathbb{1}\{(u_t, x_t, y_t, v_t) = (u, x, y, v)\}, \quad \forall (u, x, y, v) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}. \quad (\text{IV.6})$$

We denote by  $Q_{U^X Y^V}^n \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  the random variable of the empirical distribution of  $(U^n, X^n, Y^n, V^n)$ .

We aim at characterizing the asymptotic behavior of the empirical distribution  $Q_{U^X Y^V}^n$  induced by sequences of codes.

**Definition 14** A target probability distribution  $\mathcal{Q}_{U^X Y^V} \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  is achievable if there exists a sequence of codes  $(\sigma, \tau) \in \mathcal{C}(n)$  such that  $(Q_{U^X Y^V}^n)_{n \in \mathbb{N}^*}$  converges in probability to  $\mathcal{Q}_{U^X Y^V}$ , i.e.

$$\forall \delta > 0, \quad \lim_{n \rightarrow +\infty} \min_{(\sigma, \tau) \in \mathcal{C}(n)} \Pr\left(\left\| Q_{U^X Y^V}^n - \mathcal{Q}_{U^X Y^V} \right\|_1 > \delta\right) = 0. \quad (\text{IV.7})$$

We denote by  $\mathcal{A} \subset \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  the set of achievable distributions.

The concept of achievable distribution was introduced in (Cuff et al., 2010, Definition 5). In Gossner et al. (2006), a distribution  $\mathcal{Q}_{UXVY}$  is said to be *implementable* if there exists a sequence of codes  $(\sigma, \tau) \in \mathcal{C}(n)$  such that the average distribution  $(\frac{1}{n} \sum_{t=1}^n \mathcal{P}_{U_t X_t Y_t V_t}^c)_{n \in \mathbb{N}^*}$  converges to the target distribution  $\mathcal{Q}_{UXVY}$ . Note that for all  $(u, x, y, v)$ , we have

$$\frac{1}{n} \sum_{t=1}^n \mathcal{P}_{U_t X_t Y_t V_t}^c(u, x, y, v) = \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[ \mathbf{1} \{ (U_t, X_t, Y_t, V_t) = (u, x, y, v) \} \right] \quad (\text{IV.8})$$

$$= \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n \mathbf{1} \{ (U_t, X_t, Y_t, V_t) = (u, x, y, v) \} \right] \quad (\text{IV.9})$$

$$= \mathbb{E} \left[ \mathcal{Q}_{UXVY}^n(u, x, y, v) \right]. \quad (\text{IV.10})$$

The following Lemma is stated in (Larrousse et al., 2018, App. A).

**Lemma 1** *If the distribution  $\mathcal{Q}_{UXVY}$  is achievable, then it is implementable, i.e.*

$$\mathcal{Q}_{UXVY} \in \mathcal{A} \implies \lim_{n \rightarrow +\infty} \min_{(\sigma, \tau) \in \mathcal{C}(n)} \left| \mathbb{E} \left[ \mathcal{Q}_{UXVY}^n \right] - \mathcal{Q}_{UXVY} \right| = 0. \quad (\text{IV.11})$$

*Proof.* [Lemma 1] The convergence in probability implies the convergence in distribution. Therefore there exists a sequence of codes  $(\sigma, \tau) \in \mathcal{C}(n)$  such that  $(\mathbb{E}[\mathcal{Q}_{UXVY}^n])_{n \in \mathbb{N}^*}$  converges to  $\mathcal{Q}_{UXVY}$ .  $\square$

## IV.2.b Characterization of the set of achievable distributions

We characterize the set of achievable distributions for non-causal encoder and strictly-causal decoder. The following result is stated in (Le Treust, 2017, [J6], Theorem III.1).

**Theorem 6** *A target probability distribution  $\mathcal{Q}_{UXYV} \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  is achievable if and only if*

$$\mathcal{Q}_{UXYV} = \mathcal{P}_U \mathcal{Q}_{XV|U} \mathcal{T}_{Y|X}, \quad (\text{IV.12})$$

*and there exists an auxiliary random variable  $W_1 \in \mathcal{W}_1$  with  $|\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1$ , distributed according to some conditional distribution  $\tilde{\mathcal{Q}}_{W_1|UXV} \in \Delta(\mathcal{W}_1)^{|\mathcal{U} \times \mathcal{X} \times \mathcal{V}|}$  such that*

$$I(W_1; Y|V) - I(U; V, W_1) \geq 0. \quad (\text{IV.13})$$

*Note that the joint probability distribution decomposes as*

$$\mathcal{P}_U \mathcal{Q}_{XV|U} \tilde{\mathcal{Q}}_{W_1|UXV} \mathcal{T}_{Y|X} \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{W}_1 \times \mathcal{Y} \times \mathcal{V}). \quad (\text{IV.14})$$

The proof of Theorem 6 is stated in (Le Treust, 2015, [W1]). Note that the Markov chain  $Y \dashv\vdash X \dashv\vdash (U, V, W_1)$  comes from the hypothesis of strictly causal decoder. Indeed, the decoder returns a symbol  $V_t = \tau_t(Y^{t-1})$  before the channel output  $Y_t$  is generated.

**Remark 1** *As mentioned in (IV.12), an achievable distribution  $\mathcal{Q}_{UXYV}$  must have the marginals given by the source distribution  $\mathcal{P}_U$  and the channel distribution  $\mathcal{T}_{Y|X}$ . Otherwise, it is not achievable. This remark is valid for all the coding results in this document.*

**Theorem 7** *The set of achievable distributions is compact and convex, and it is characterized by*

$$\mathcal{A} = \left\{ \mathcal{P}_U \mathcal{Q}_{XV|U} \mathcal{T}_{Y|X}, \max_{\substack{\tilde{\mathcal{Q}}_{W_1|UXV}, \\ |\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1}} \left( I(W_1; Y|V) - I(U; V, W_1) \right) \geq 0 \right\}. \quad (\text{IV.15})$$

*Moreover, the function*

$$\Psi : \Delta(\mathcal{X} \times \mathcal{V})^{|\mathcal{U}|} \longrightarrow \mathbb{R}, \quad (\text{IV.16})$$

$$\mathcal{Q}_{XV|U} \longrightarrow \max_{\substack{\tilde{\mathcal{Q}}_{W_1|UXV}, \\ |\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1}} \left( I(W_1; Y|V) - I(U; V, W_1) \right), \quad (\text{IV.17})$$

*is concave in  $\mathcal{Q}_{XV|U} \in \Delta(\mathcal{X} \times \mathcal{V})^{|\mathcal{U}|}$ .*

The proof of Theorem 7 is stated in (Le Treust, 2017, [J6]). We compare the result of Theorem 6 with previous results from the literature. When, the channel is not noisy, i.e.  $\mathcal{T}_{Y|X} = \mathbb{1}\{Y = X\}$ , the set of implementable distributions of (Gossner et al., 2006, Theorems 1 and 2) coincides with the set of achievable distributions of (Cuff and Zhao, 2011, Theorem 5.1). We show that the information constraints of (Gossner et al., 2006, Eq. (1)) and of (Gossner et al., 2006, Theorem 5.1) coincide with (IV.13).

**Corollary 1** *When  $\mathcal{T}_{Y|X} = \mathbb{1}\{Y = X\}$ , we have*

$$\max_{\substack{\tilde{\mathcal{Q}}_{W_1|UXV} \\ |W_1| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1}} \left( I(W_1; Y|V) - I(U; V, W_1) \right) = H(X|V) - I(U; X, V), \quad (\text{IV.18})$$

and the maximum in the left-hand side of (IV.18) is achieved by  $W_1 = X$ .

*Proof.* [Corollary 1] When  $Y = X$ , for all distribution  $\tilde{\mathcal{Q}}_{W_1|UXV}$  we have

$$I(W_1; X|V) - I(U; V, W_1) = I(W_1; X|V, U) - I(W_1; U|V, X) - I(U; V) \quad (\text{IV.19})$$

$$\leq H(X|V, U) - I(U; V) = H(X|V) - I(U; X, V). \quad (\text{IV.20})$$

We obtain the equality by replacing  $W_1 = X$ .  $\square$

According to the separation result of Shannon (1959), the source random variables  $(U, V)$  must be independent of the channel random variables  $(X, Y)$ .

**Corollary 2** *Assume that  $(U, V)$  are independent of  $(X, Y)$ , we have*

$$\mathcal{A} \cap \Delta(\mathcal{U} \times \mathcal{V}) \times \Delta(\mathcal{X} \times \mathcal{Y}) = \left\{ \mathcal{P}_U \mathcal{Q}_{V|U} \mathcal{Q}_X \mathcal{T}_{Y|X}, \quad I(X; Y) - I(U; V) \geq 0 \right\}. \quad (\text{IV.21})$$

*Proof.* [Corollary 2] For all distribution  $\tilde{\mathcal{Q}}_{W_1|UXV}$ , we have

$$I(W_1; Y|V) - I(U; V, W_1) \leq I(W_1, V; Y) - I(U; V) \quad (\text{IV.22})$$

$$\leq I(X; Y) - I(U; V), \quad (\text{IV.23})$$

where (IV.23) comes from the Markov chain property  $Y \dashv\vdash X \dashv\vdash (U, V, W_1)$  of the channel. We have equality in (IV.22) and (IV.23) when choosing  $W_1 = X$  independent of  $(U, V)$ .  $\square$

Corollary 2 also shows that strictly causal decoding has no impact on the information constraint  $I(X; Y) - I(U; V) \geq 0$ , stated in Shannon (1959). In fact this information constraint characterizes the optimal solution for non-causal encoding and strictly causal, causal or non-causal decoding. However, when we impose that the random variables  $(X, Y)$  are correlated with  $(U, V)$ , the information constraint depends on whether the decoder works strictly causally, causally or non-causally.

**Remark 2 (Coordination is more restrictive than information transmission)** *Equations (IV.22)-(IV.23) imply that the information constraint corresponding to  $\mathcal{Q}_{XV|U}$  is stronger than the one corresponding to the product of marginals  $\mathcal{Q}_{V|U} \mathcal{Q}_X$ . We conclude that the empirical coordination is more restrictive than lossless or lossy transmission of the information source.*

### IV.2.c Evaluation of an arbitrary expected cost function

By building on the repeated game approach of Gossner et al. (2006), we consider that the encoder and the decoder implement a coding scheme in order to minimize an arbitrary cost function  $\phi$  defined by

$$\phi : \mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V} \longrightarrow \mathbb{R}. \quad (\text{IV.24})$$

This cost function  $\phi$  may capture the different objectives of the encoder and the decoder. We may also consider multiple cost functions  $(\phi_1, \phi_2, \dots, \phi_K)$ . For example, we consider a distortion function  $d : \mathcal{U} \times \mathcal{V} \longrightarrow \mathbb{R}$ , a channel input cost function  $c : \mathcal{X} \longrightarrow \mathbb{R}$ . Note that the maximization of the payoff function  $\pi : \mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V} \longrightarrow \mathbb{R}$  of a repeated game as in Gossner et al. (2006), corresponds to the minimization of the cost function  $-\pi$ . Given a code  $(\sigma, \tau) \in \mathcal{C}(n)$ , we define the associated  $n$ -stage cost.



**Definition 15** The  $n$ -stage cost  $\phi^n(c)$  of the code  $(\sigma, \tau) \in \mathcal{C}(n)$  is defined by

$$\phi^n(c) = \mathbb{E} \left[ \frac{1}{n} \cdot \sum_{t=1}^n \phi(U_t, X_t, Y_t, V_t) \right]. \quad (\text{IV.25})$$

A cost value  $\phi \in \mathbb{R}$  is achievable if for all  $\varepsilon > 0$ , there exists a  $\bar{n} \in \mathbb{N}^*$ , such that for all  $n \geq \bar{n}$ , there exists a code  $c \in \mathcal{C}(n)$  such that

$$\left| \phi - \mathbb{E} \left[ \frac{1}{n} \cdot \sum_{t=1}^n \phi(U_t, X_t, Y_t, V_t) \right] \right| \leq \varepsilon. \quad (\text{IV.26})$$

We denote by  $\Phi$  the set of achievable cost values.

**Theorem 8** We have

$$\Phi = \left\{ \phi \in \mathbb{R}, \exists Q_{UXYV} \in \mathcal{A}, \mathbb{E}[\phi(U, X, Y, V)] = \phi \right\}. \quad (\text{IV.27})$$

Moreover, the optimization problem

$$\min_{Q_{UXYV} \in \mathcal{A}} \mathbb{E}[\phi(U, X, Y, V)], \quad (\text{IV.28})$$

is a convex optimization problem.

The proof of Theorem 8 is stated in (Le Treust, 2017, [J6]). In Sec. IV.2.e, we study the coordination game of Gossner et al. (2006) and we investigate the trade-off between the source distortion  $d : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}$ , and the channel input cost function  $c : \mathcal{X} \rightarrow \mathbb{R}$ .

#### IV.2.d Example #1: coordination game of Gossner et al. (2006)

We consider the binary sets of symbols  $\mathcal{U} = \mathcal{X} = \mathcal{Y} = \mathcal{V} = \{0, 1\}$ , a binary information source  $\mathcal{P}_U = (p, 1-p)$ , with  $p \in [0, 1]$  and a binary symmetric channel  $\mathcal{T}_{Y|X} = (1-\varepsilon)\mathbb{1}\{Y=X\} + \varepsilon\mathbb{1}\{Y \neq X\}$ , with cross-over probability parameter  $\varepsilon \in [0, 1]$ , as depicted in Fig. IV.2.

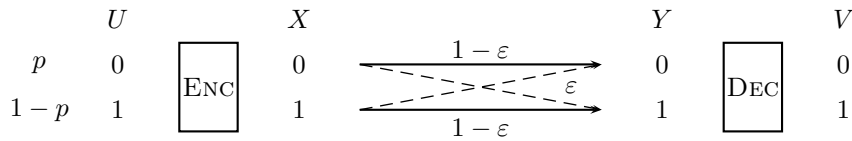


Figure IV.2 – Binary information source  $\mathcal{P}_U$  and binary symmetric channel  $\mathcal{T}_{Y|X}$ .

The encoder and the decoder seek to coordinate their actions  $X$  and  $V$  with the information source  $U$ , in order to maximize the utility function  $\pi : \mathcal{U} \times \mathcal{X} \times \mathcal{V} \mapsto \mathbb{R}$  defined by Fig. IV.3.

	$V = 0$	$V = 1$		$V = 0$	$V = 1$
$X = 0$	1	0	$X = 0$	0	0
$X = 1$	0	0	$X = 1$	0	1
	$U = 0$			$U = 1$	

Figure IV.3 – Utility function  $\pi : \mathcal{U} \times \mathcal{X} \times \mathcal{V} \mapsto \mathbb{R}$  of Gossner et al. (2006).

	$V = 0$	$V = 1$		$V = 0$	$V = 1$
$X = 0$	$\frac{\gamma}{2}$	$\frac{1-\gamma}{6}$	$X = 0$	$\frac{1-\gamma}{6}$	$\frac{1-\gamma}{6}$
$X = 1$	$\frac{1-\gamma}{6}$	$\frac{1-\gamma}{6}$	$X = 1$	$\frac{1-\gamma}{6}$	$\frac{\gamma}{2}$
	$U = 0$			$U = 1$	

 Figure IV.4 – The optimal distribution for  $\pi$ , depending on  $\gamma \in [0, 1]$ .

Since the decoder is strictly causal, it is clear from [Gossner et al. \(2006\)](#), that the distribution  $\mathcal{P}_U \mathbb{1}\{X = V = U\}$  which maximizes  $\mathbb{E}[\pi(U, X, V)]$  is not achievable. We suppose that the source parameter is  $p = \frac{1}{2}$ . By symmetry, the optimal achievable distribution  $\mathcal{Q}_{XV|U}$  is of the form of [Fig. IV.3](#) which depends on some parameter  $\gamma \in [0, 1]$ . We consider lower and upper bounds on the information constraint ([IV.13](#)) that do not involve an auxiliary random variable  $W_1$ .

**Proposition 5** For all  $\mathcal{Q}_{XV|U} \in \Delta(\mathcal{X} \times \mathcal{V})^{|\mathcal{U}|}$ , we have

$$I(X; Y|V) - I(U; V, X) \leq \max_{\substack{\mathcal{Q}_{W_1|U, X, V} \\ |\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1}} \left( I(W_1; Y|V) - I(U; V, W_1) \right) \leq I(X; Y|U, V) - I(U; V). \quad (\text{IV.29})$$

The difference between the upper bound and the lower bound in ([IV.29](#)) is equal to  $I(U; X|V, Y)$ . When the channel is not noisy  $\mathcal{T}_{Y|X} = \mathbb{1}\{Y = X\}$ , the two bounds coincides since  $I(U; X|V, Y) = 0$ , and correspond to the information constraint  $H(X|V) - I(U; V, X) \geq 0$  of [Gossner et al. \(2006\)](#) and [Cuff and Zhao \(2011\)](#).

**Proposition 6** For a distribution  $\mathcal{Q}_{XV|U} \in \Delta(\mathcal{X} \times \mathcal{V})^{|\mathcal{U}|}$  of the form of [Fig. IV.4](#) with  $\gamma \in [0, 1]$ , we have

$$I(X; Y|V) - I(U; V, X) = h_b(\gamma) + (1 - \gamma) \cdot \log_2(3) - 1 - h_b(\varepsilon) - h_b\left(\frac{2}{3}(1 - \gamma)\right) + h_b\left(\frac{2}{3}(1 - \gamma) + \varepsilon \cdot \frac{4\gamma - 1}{3}\right), \quad (\text{IV.30})$$

$$I(X; Y|U, V) - I(U; V) = h_b(\gamma) + (1 - \gamma) \cdot \log_2(3) - 1 - h_b(\varepsilon) + \frac{2\gamma + 1}{3} \left( h_b\left(\left(1 - \varepsilon\right) \cdot \frac{3\gamma}{2\gamma + 1} + \varepsilon \cdot \frac{1 - \gamma}{2\gamma + 1}\right) - h_b\left(\frac{3\gamma}{2\gamma + 1}\right) \right). \quad (\text{IV.31})$$

**Remark 3** Note that when  $\varepsilon = 0$ , the information constraints ([IV.30](#)) and ([IV.31](#)) reduce to the one stated in [Gossner et al. \(2006\)](#) and [Cuff and Zhao \(2011\)](#):

$$H(X|V) - I(U; V, X) = h_b(\gamma) + (1 - \gamma) \cdot \log_2(3) - 1. \quad (\text{IV.32})$$

[Fig. IV.5](#) and [IV.6](#) represent the lower and the upper bounds of equations ([IV.30](#)) and ([IV.31](#)), depending on  $\gamma \in [0, 1]$ , for different values of  $\varepsilon \in [0, 1]$ . The maximum of the information constraint is achieved by parameter  $\gamma = 0.25$ , that corresponds to the uniform probability distribution over the symbols  $\mathcal{U} \times \mathcal{X} \times \mathcal{V}$ . The maximum of the cost is achieved by a distribution  $\mathcal{Q}_{XV|U}^*$  of the form of [Fig. IV.4](#) with parameter  $\gamma^* \in [0, 1]$  such that  $\Psi(\mathcal{Q}_{XV|U}^*) = 0$ .

- If the channel is perfect *i.e.*,  $\varepsilon = 0$ , the optimal solution corresponds to the one stated in [Gossner et al. \(2006\)](#) where the optimal distribution  $\mathcal{Q}_{XV|U}^*$  has parameter  $\gamma^* \simeq 0.81$  which is the solution of  $h_b(\gamma) + (1 - \gamma) \cdot \log_2(3) = 1$ .

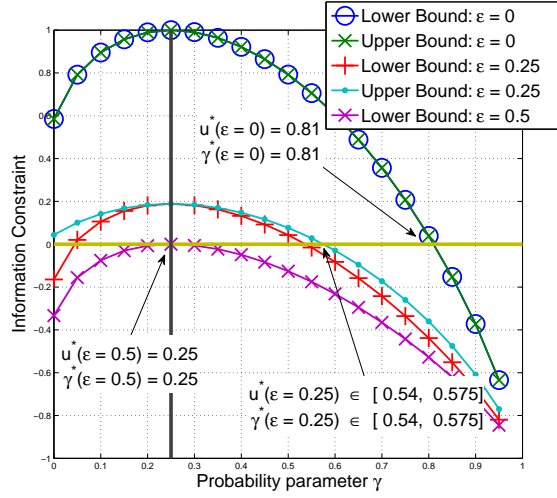


Figure IV.5 – Equations (IV.30) and (IV.31) as functions of  $\gamma \in [0, 1]$  for  $\varepsilon \in \{0, 0.25, 0.5\}$ .

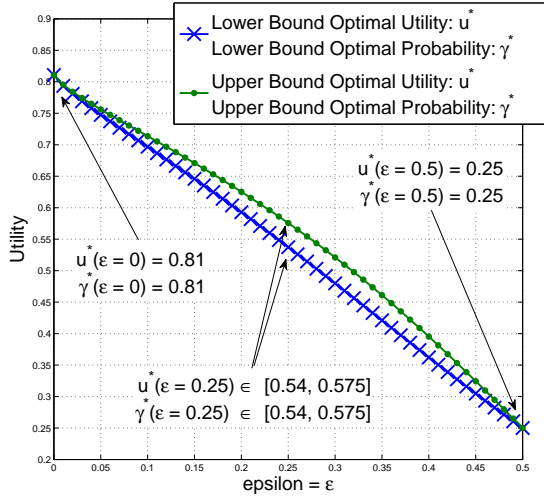


Figure IV.6 – Parameter  $\gamma^* \in [0, 1]$  solution to  $\Psi(Q_{XV|U}^*) = 0$ , depending on  $\varepsilon \in [0, 0.5]$ .

- If the channel parameter is  $\varepsilon = 0.5$ , then the channel outputs are statistically independent of the channel inputs. The optimal utility is obtained when  $U$ ,  $X$  and  $V$  are uniformly distributed. In that case, no information is transmitted.
- If the channel parameter is  $\varepsilon = 0.25$ , the optimal cost belongs to the interval  $[0.54, 0.575]$ . Even if the channel is noisy, the symbols of the encoder, the decoder and the source are perfectly coordinated more than half of the time.

#### IV.2.e Example #2: Hamming distortion and channel cost

We investigate the relationship between the result stated in Theorem 6 for empirical coordination and the separation result of Shannon (1959) for the joint source-channel coding problem, see also (El Gamal and Kim, 2011, pp. 47, 57 and 66). We consider a distortion function  $d : \mathcal{U} \times \mathcal{V} \mapsto \mathbb{R}$  and a cost function  $c : \mathcal{X} \mapsto \mathbb{R}$ .

**Definition 16** *The pair of distortion-cost  $(D^*, C^*)$  is achievable if*

$$\lim_{n \rightarrow +\infty} \min_{(\sigma, \tau) \in \mathcal{C}(n)} \left( \left| \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n d(U_t, V_t) \right] - D^* \right| + \left| \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n c(X_t) \right] - C^* \right| \right) = 0. \quad (\text{IV.33})$$

Without loss of generality, we consider exact distortion and cost rather than upper bounds on the distortion  $\mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n d(U_t, V_t) \right] \leq D^* + \varepsilon$ , and cost  $\mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n c(X_t) \right] \leq C^* + \varepsilon$ , as in (El Gamal and Kim, 2011, pp. 47 and 57). Theorem 6 provides a characterization of the achievable pairs of distortion-cost  $(D^*, C^*)$ .

**Corollary 3** *The following assertions are equivalent:*

$$1. \quad \text{The pair of distortion-cost } (D^*, C^*) \text{ is achievable,} \quad (\text{IV.34})$$

$$2. \quad \exists \mathcal{Q}_{UXYV} \in \mathcal{A}, \quad \mathbb{E}_{\mathcal{Q}}[d(U, V)] = D^*, \quad \mathbb{E}_{\mathcal{Q}}[c(X)] = C^*, \quad (\text{IV.35})$$

$$3. \quad \max_{\substack{\mathcal{P}_X \\ \mathbb{E}_{\mathcal{Q}}[c(X)] = C^*}} I(X; Y) - \min_{\substack{\mathcal{Q}_{V|U} \\ \mathbb{E}_{\mathcal{Q}}[d(U, V)] = D^*}} I(U; V) \geq 0. \quad (\text{IV.36})$$

**Remark 4** *Unlike in (El Gamal and Kim, 2011, pp. 43, Remark 3.5), Corollary 3 establishes that the pair of distortion-cost  $(D^*, C^*)$  is achievable when equation (IV.36) is also equal to zero. More details are provided in (Le Treust, 2017, [J6]).*

*Proof.* [Corollary 3] The equivalence between 1. and 2. follows from Theorem 8. We prove the equivalence between 2. and 3. We consider a distribution  $\mathcal{Q}_{UXYV} \in \mathcal{A}$  that satisfy (IV.35) and we define the product of the marginal distributions  $\tilde{\mathcal{Q}}_{UXYV} = \mathcal{Q}_{UV} \mathcal{Q}_{XY}$ . Then,  $\mathbb{E}_{\tilde{\mathcal{Q}}}[d(U, V)] = \mathbb{E}_{\mathcal{Q}}[d(U, V)] = D^*$  and  $\mathbb{E}_{\tilde{\mathcal{Q}}}[c(X)] = \mathbb{E}_{\mathcal{Q}}[c(X)] = C^*$ . Moreover,

$$0 \leq \max_{\substack{\mathcal{Q}_{W_1|UXV} \\ |\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1}} \left( I_{\mathcal{Q}}(W_1; Y|V) - I_{\mathcal{Q}}(U; V, W_1) \right) \leq I_{\mathcal{Q}}(X; Y) - I_{\mathcal{Q}}(U; V) = I_{\tilde{\mathcal{Q}}}(X; Y) - I_{\tilde{\mathcal{Q}}}(U; V). \quad (\text{IV.37})$$

Therefore, (IV.36) is satisfied.

Now we consider the distributions  $\mathcal{Q}_X$  and  $\mathcal{Q}_{V|U}$  that achieve respectively the maximum and the minimum in (IV.36). By Corollary 2, we have  $\mathcal{P}_U \mathcal{Q}_{V|U} \mathcal{Q}_X \mathcal{T}_{Y|X} \in \mathcal{A} \cap \Delta(\mathcal{U} \times \mathcal{V}) \times \Delta(\mathcal{U} \times \mathcal{V}) \subset \mathcal{A}$ . Moreover,  $\mathbb{E}_{\mathcal{Q}}[d(U, V)] = D^*$  and  $\mathbb{E}_{\mathcal{Q}}[c(X)] = C^*$ , which imply (IV.35).  $\square$

**Example 2 (Trade-off distortion-cost)** *We consider the coordination problem of Fig. IV.2 with  $p \in [0, 1]$ ,  $\varepsilon \in [0, 1]$  and the distortion function is  $d(u, v) = \mathbb{1}\{u \neq v\}$  and cost function is  $c(x) = \mathbb{1}\{x = 0\}$ . We consider the distributions  $\mathcal{Q}_X = (\alpha, 1 - \alpha)$ , with  $\alpha \in [0, 1]$  and  $\mathcal{Q}_{V|U} = (1 - \beta)\mathbb{1}\{V = U\} + \beta\mathbb{1}\{V \neq U\}$ , with  $\beta \in [0, 1]$ . The expected distortion and the expected cost are given by:*

$$\mathbb{E}[d(U, V)] = p \cdot \beta + (1 - p) \cdot \beta = \beta, \quad (\text{IV.38})$$

$$\mathbb{E}[c(X)] = \sum_x \mathcal{Q}_X(x) \cdot \mathbb{1}\{x = 0\} = \alpha. \quad (\text{IV.39})$$

Note that a target pair of distortion-cost  $(D^*, C^*)$  induces a unique pair of parameters  $(\beta, \alpha)$ . The information constraint (IV.36) of Corollary 3 is equal to:

$$I(X; Y) - I(U; V) = h_b(\alpha \cdot \varepsilon + (1 - \alpha) \cdot (1 - \varepsilon)) + h_b(\beta) - h_b(\varepsilon) - h_b(\beta \cdot p + (1 - \beta) \cdot (1 - p)). \quad (\text{IV.40})$$

Fig. IV.7 represents the regions of achievable pairs of exact distortion-cost, for parameters  $(\varepsilon, p) \in \{(0.05, 0.5); (0.25, 0.25); (0.25, 0.5)\}$ . This illustrates the trade-off between minimal source distortion and minimal channel cost. The boundary of the three regions corresponds to the pairs of distortion and cost  $(D^*, C^*)$ , that satisfy the equality  $I(X; Y) = I(U; V)$ .

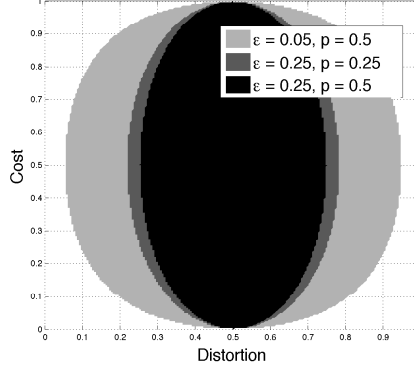


Figure IV.7 – The set of achievable Distortion-Cost ( $D^*$ ,  $C^*$ ) depending on source and channel parameters  $(\varepsilon, p) \in \{(0.05, 0.5); (0.25, 0.25); (0.25, 0.5)\}$ .

### IV.3 General decoding functions

We investigate the empirical coordination problem for general decoding functions.

#### IV.3.a Causal decoding with source feedforward

We consider the setting depicted in Fig. IV.8 in which the decoder observes causally the channel output symbols, i.e. the past and the current symbols  $Y^t = (Y^{t-1}, Y_t)$ , and we consider four scenarios of source feedforward, i.e. the decoder observes the source symbols either strictly causally  $U^{t-1}$ , or causally  $U^t$ , or non-causally  $U^n$ , or does not observe any source symbol  $\emptyset$ .

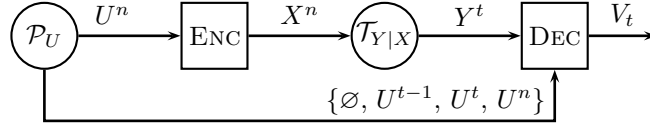


Figure IV.8 – Causal channel output  $Y^t$  and four different scenarios of source feedforward  $\{\emptyset, U^{t-1}, U^t, U^n\}$ .

**Definition 17** Given  $n \in \mathbb{N}^*$ , we define the non-causal encoder  $\sigma$  and four different decoding functions  $\tau_t^c, \tau_t^{cs}, \tau_t^{cc}, \tau_t^{cn}$ , for  $t \in \{1, \dots, n\}$ .

$$\sigma : \mathcal{U}^n \longrightarrow \mathcal{X}^n, \quad (\text{IV.41})$$

$$\tau_t^c : \mathcal{Y}^t \longrightarrow \mathcal{V}, \quad t \in \{1, \dots, n\}, \quad (\text{IV.42})$$

$$\tau_t^{cs} : \mathcal{Y}^t \times \mathcal{U}^{t-1} \longrightarrow \mathcal{V}, \quad t \in \{1, \dots, n\}, \quad (\text{IV.43})$$

$$\tau_t^{cc} : \mathcal{Y}^t \times \mathcal{U}^t \longrightarrow \mathcal{V}, \quad t \in \{1, \dots, n\}, \quad (\text{IV.44})$$

$$\tau_t^{cn} : \mathcal{Y}^t \times \mathcal{U}^n \longrightarrow \mathcal{V}, \quad t \in \{1, \dots, n\}. \quad (\text{IV.45})$$

We denote by  $\mathcal{C}^{cs}(n)$  (resp.  $\mathcal{C}^{cc}(n)$ ,  $\mathcal{C}^{cn}(n)$ ,  $\mathcal{C}^c(n)$ ) the set of  $n$ -length code with causal channel output and strictly causal (resp. causal, non-causal, absent) source feedforward.

The sets of achievable target probability distributions  $\mathcal{A}^c \subset \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  (resp.  $\mathcal{A}^{cc}$ ,  $\mathcal{A}^{cn}$ ,  $\mathcal{A}^c$ ) are defined similarly as in Definition 14, with respect to the set of code  $\mathcal{C}^{cs}(n)$  (resp.  $\mathcal{C}^{cc}(n)$ ,  $\mathcal{C}^{cn}(n)$ ,  $\mathcal{C}^c(n)$ ). By definition, we have  $\mathcal{A}^c \subset \mathcal{A}^{cs} \subset \mathcal{A}^{cc} \subset \mathcal{A}^{cn} \subset \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ .

When considering a code  $(\sigma, \tau) \in \mathcal{C}^{\text{cn}}(n)$ , both encoder and decoder have non-causal observation of the source sequence  $U^n$ . Therefore, the decoder can replicate the encoding function and select a sequence  $V^n$  that is empirically correlated with  $(U^n, X^n, Y^n)$ . Thus, the set of achievable target probability distributions is given by

$$\mathcal{A}^{\text{cn}} = \left\{ \mathcal{P}_U \mathcal{Q}_{X|U} \mathcal{T}_{Y|X} \mathcal{Q}_{V|UXY} \right\}. \quad (\text{IV.46})$$

In the following, we characterize the sets  $\mathcal{A}^{\text{cs}}$ ,  $\mathcal{A}^{\text{cc}}$  and  $\mathcal{A}^{\text{c}}$ , that are subsets of  $\mathcal{A}^{\text{cn}}$ . The following results are stated in (Le Treust, 2015b, [IC14], Sec. IV), (Le Treust, 2017, [J6], Theorem 4), the proof is also stated in (Le Treust, 2015, [W1]).

**Theorem 9** 1. A distribution  $\mathcal{Q}_{UXYV} \in \mathcal{A}^{\text{cc}}$  if and only if  $\mathcal{Q}_{UXYV} = \mathcal{P}_U \mathcal{Q}_{X|U} \mathcal{T}_{Y|X} \mathcal{Q}_{V|UXY}$  and there exists an auxiliary random variable  $W_2 \in \mathcal{W}_2$  where  $|\mathcal{W}_2| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1$  and a joint distribution

$$\tilde{\mathcal{Q}}_{UXW_2YV} = \mathcal{P}_U \mathcal{Q}_{X|U} \tilde{\mathcal{Q}}_{W_2|UX} \mathcal{T}_{Y|X} \tilde{\mathcal{Q}}_{V|UYW_2}, \quad (\text{IV.47})$$

that satisfies

$$\mathcal{Q}_{V|UXY} = \sum_{w_2} \tilde{\mathcal{Q}}_{W_2|UX}(w_2|\cdot, \cdot) \tilde{\mathcal{Q}}_{V|UYW_2}(\cdot|\cdot, w_2), \quad (\text{IV.48})$$

$$I(X; Y|U, W_2) - I(W_2; U) \geq 0. \quad (\text{IV.49})$$

2. A distribution  $\mathcal{Q}_{UXYV} \in \mathcal{A}^{\text{cs}}$  if and only if  $\mathcal{Q}_{UXYV} = \mathcal{P}_U \mathcal{Q}_{X|U} \mathcal{T}_{Y|X} \mathcal{Q}_{V|UXY}$  and there exists an auxiliary random variable  $W_2 \in \mathcal{W}_2$  where  $|\mathcal{W}_2| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1$  and a joint distribution

$$\tilde{\mathcal{Q}}_{UXW_2YV} = \mathcal{P}_U \mathcal{Q}_{X|U} \tilde{\mathcal{Q}}_{W_2|UX} \mathcal{T}_{Y|X} \tilde{\mathcal{Q}}_{V|YW_2}, \quad (\text{IV.50})$$

that satisfies

$$\mathcal{Q}_{V|UXY} = \sum_{w_2} \tilde{\mathcal{Q}}_{W_2|UX}(w_2|\cdot, \cdot) \tilde{\mathcal{Q}}_{V|YW_2}(\cdot|\cdot, w_2), \quad (\text{IV.51})$$

$$I(X; Y|U, W_2) - I(W_2; U) \geq 0. \quad (\text{IV.52})$$

3. A distribution  $\mathcal{Q}_{UXYV} \in \mathcal{A}^{\text{c}}$  if and only if  $\mathcal{Q}_{UXYV} = \mathcal{P}_U \mathcal{Q}_{X|U} \mathcal{T}_{Y|X} \mathcal{Q}_{V|UXY}$  and there exists a pair of auxiliary random variables  $(W_1, W_2) \in \mathcal{W}_1 \times \mathcal{W}_2$  where  $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1$  and a joint distribution

$$\tilde{\mathcal{Q}}_{UXW_1W_2YV} = \mathcal{P}_U \mathcal{Q}_{X|U} \tilde{\mathcal{Q}}_{W_1W_2|UX} \mathcal{T}_{Y|X} \tilde{\mathcal{Q}}_{V|YW_2}, \quad (\text{IV.53})$$

that satisfies

$$\mathcal{Q}_{V|UXY} = \sum_{(w_1, w_2)} \tilde{\mathcal{Q}}_{W_1W_2|UX}(w_1, w_2|\cdot, \cdot) \tilde{\mathcal{Q}}_{V|YW_2}(\cdot|\cdot, w_2), \quad (\text{IV.54})$$

$$I(W_1; Y|W_2) - I(W_1, W_2; U) \geq 0. \quad (\text{IV.55})$$

The only difference between the sets  $\mathcal{A}^{\text{cc}}$  and  $\mathcal{A}^{\text{cs}}$  comes from the conditional distributions  $\tilde{\mathcal{Q}}_{V|UYW_2}$  and  $\tilde{\mathcal{Q}}_{V|YW_2}$ . This is due to the observation of the current symbol  $U_t$  in the decoding functions  $\tau_t^{\text{cc}}$ , which is not the case for the decoding functions  $\tau_t^{\text{cs}}$ , for  $t \in \{1, \dots, n\}$ . When the decoder has no source feedforward, another auxiliary random variable  $W_1 \in \mathcal{W}_1$  is needed in order to tune the state-dependent channel, as in Gel'fand and Pinsker (1980).

### IV.3.b Strictly causal decoding with source feedforward

For the seek of completeness, we characterize the set of achievable distributions for strictly causal decoding with source feedforward, similarly as in Sec. IV.3.a. We consider the setting depicted in Fig. IV.9 in which the decoder observes strictly causally the channel output symbols  $Y^{t-1}$ , and we consider four scenarios of source feedforward, i.e. the decoder observes the source symbols either strictly causally  $U^{t-1}$ , or causally  $U^t$ , or non-causally  $U^n$ , or does not observe any source symbol  $\emptyset$ .

**Definition 18** Given  $n \in \mathbb{N}^*$ , we define the non-causal encoder  $\sigma$  and four different decoding functions  $\tau_t^s, \tau_t^{ss}, \tau_t^{sc}, \tau_t^{sn}$ , for  $t \in \{1, \dots, n\}$ .

$$\sigma : \mathcal{U}^n \longrightarrow \mathcal{X}^n, \quad (\text{IV.56})$$

$$\tau_t^s : \mathcal{Y}^{t-1} \longrightarrow \mathcal{V}, \quad t \in \{1, \dots, n\}, \quad (\text{IV.57})$$

$$\tau_t^{ss} : \mathcal{Y}^{t-1} \times \mathcal{U}^{t-1} \longrightarrow \mathcal{V}, \quad t \in \{1, \dots, n\}, \quad (\text{IV.58})$$

$$\tau_t^{sc} : \mathcal{Y}^{t-1} \times \mathcal{U}^t \longrightarrow \mathcal{V}, \quad t \in \{1, \dots, n\}, \quad (\text{IV.59})$$

$$\tau_t^{sn} : \mathcal{Y}^{t-1} \times \mathcal{U}^n \longrightarrow \mathcal{V}, \quad t \in \{1, \dots, n\}. \quad (\text{IV.60})$$

We denote by  $\mathcal{C}^{sc}(n)$  (resp.  $\mathcal{C}^{ss}(n)$ ,  $\mathcal{C}^{sn}(n)$ ,  $\mathcal{C}^s(n)$ ) the set of  $n$ -length code with strictly causal channel output and causal (resp. strictly causal, non-causal, absent) source feedforward.

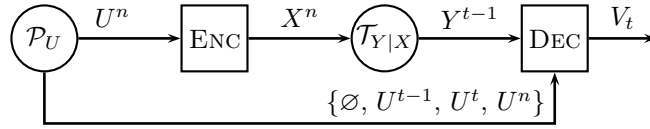


Figure IV.9 – Strictly causal channel output  $Y^{t-1}$  and four different scenarios of source feedforward  $\{\emptyset, U^{t-1}, U^t, U^n\}$ .

The sets of achievable target probability distributions  $\mathcal{A}^{sc} \subset \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  (resp.  $\mathcal{A}^{ss}$ ,  $\mathcal{A}^{sn}$ ,  $\mathcal{A}^s$ ) are defined similarly as in Definition 14, with respect to the set of code  $\mathcal{C}^{sc}(n)$  (resp.  $\mathcal{C}^{ss}(n)$ ,  $\mathcal{C}^{sn}(n)$ ,  $\mathcal{C}^s(n)$ ). By definition, we have  $\mathcal{A}^s \subset \mathcal{A}^{ss} \subset \mathcal{A}^{sc} \subset \mathcal{A}^{sn} \subset \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ .

When considering a code  $(\sigma, \tau) \in \mathcal{C}^{sn}(n)$ , both encoder and decoder have non-causal observation of the source sequence  $U^n$ . Therefore, the decoder can replicate the encoding function and therefore select a sequence  $V^n$  that is empirically correlated with  $(U^n, X^n)$ . For all  $t \in \{1, \dots, n\}$ , the decoder must return  $V_t$  before it observes  $Y_t$ . Thus, we have the Markov chain property  $Y \ominus X \ominus (U, V)$  and the set of achievable distributions is given by

$$\mathcal{A}^{sn} = \left\{ \mathcal{P}_U \mathcal{Q}_{XV|U} \mathcal{T}_{Y|X} \right\}. \quad (\text{IV.61})$$

In the following, we characterize the sets  $\mathcal{A}^s$ ,  $\mathcal{A}^{ss}$  and  $\mathcal{A}^{sc}$ .

**Theorem 10** 1. A distribution  $\mathcal{Q}_{UXYV} \in \mathcal{A}^{sc}$  if and only if  $\mathcal{Q}_{UXYV} = \mathcal{P}_U \mathcal{Q}_{XV|U} \mathcal{T}_{Y|X}$  and there exists an auxiliary random variable  $W_2 \in \mathcal{W}_2$  where  $|\mathcal{W}_2| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1$  and a joint distribution

$$\tilde{\mathcal{Q}}_{UXW_2YV} = \mathcal{P}_U \mathcal{Q}_{X|U} \tilde{\mathcal{Q}}_{W_2|UX} \tilde{\mathcal{Q}}_{V|UW_2} \mathcal{T}_{Y|X}, \quad (\text{IV.62})$$

that satisfies

$$\mathcal{Q}_{V|UX} = \sum_{w_2} \tilde{\mathcal{Q}}_{W_2|UX}(w_2|\cdot, \cdot) \tilde{\mathcal{Q}}_{V|UW_2}(\cdot|w_2), \quad (\text{IV.63})$$

$$I(X; Y|U, W_2) - I(W_2; U) \geq 0. \quad (\text{IV.64})$$

2. A distribution  $\mathcal{Q}_{UXYV} \in \mathcal{A}^{\text{ss}}$  if and only if  $\mathcal{Q}_{UXYV} = \mathcal{P}_U \mathcal{Q}_{XV|U} \mathcal{T}_{Y|X}$  and

$$I(X; Y|U, V) - I(V; U) \geq 0. \quad (\text{IV.65})$$

3. A distribution  $\mathcal{Q}_{UXYV} \in \mathcal{A}^{\text{s}}$  if and only if  $\mathcal{Q}_{UXYV} = \mathcal{P}_U \mathcal{Q}_{XV|U} \mathcal{T}_{Y|X}$  and there exists an auxiliary random variables  $W_1 \in \mathcal{W}_1$  where  $|\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{V}| + 1$  and a joint distribution

$$\tilde{\mathcal{Q}}_{UXW_1YV} = \mathcal{P}_U \mathcal{Q}_{XV|U} \tilde{\mathcal{Q}}_{W_1|UXV} \mathcal{T}_{Y|X}, \quad (\text{IV.66})$$

that satisfies

$$I(W_1; Y|V) - I(W_1, V; U) \geq 0. \quad (\text{IV.67})$$

The Theorem 10 (2.) was stated for achievable distributions in (Le Treust, 2017, [J6], Corollary 3) and for implementable distributions in (Gossner et al., 2006, Eq. (11)), and (Larrousse et al., 2018, Theorem 4 and Corollary 12). The Theorem 10 (3.) corresponds to the Theorem 7 in Sec. IV.2.

**Remark 5** An inner bound to the set  $\mathcal{A}^{\text{sc}}$  was stated in (Larrousse et al., 2015a, Theorem 2) with a pair of auxiliary random variables  $(W_1, W_2)$ . When using our notations for  $(S_0, S_1, S_2, U, V, X_1, X_2, Y_1, Y_2) = (\emptyset, U, U, W_2, W_1, X, V, Y, \emptyset)$ , we obtain the decomposition of the distribution

$$\tilde{\mathcal{Q}}_{UXW_1W_2VY} = \mathcal{P}_U \mathcal{Q}_{X|U} \tilde{\mathcal{Q}}_{W_1W_2|UX} \tilde{\mathcal{Q}}_{V|UW_2} \mathcal{T}_{Y|X}, \quad (\text{IV.68})$$

and the information constraint

$$I(W_1; Y|W_2) - I(W_1, W_2; U) \geq 0. \quad (\text{IV.69})$$

However, the decoder may exploit the causal observation of the source which may lead to a weaker information constraint

$$I(W_1; Y|W_2) - I(W_1, W_2; U) \leq I(W_1; U, Y|W_2) - I(W_1, W_2; U) = I(W_1; Y|W_2, U) - I(W_2; U), \quad (\text{IV.70})$$

for which it is optimal to take  $W_1 = X$ .

## IV.4 State-dependent source and state-dependent channel

In this section, we investigate the empirical coordination problem for state-dependent source and channels. The state-dependent source generates a triple of sequences  $(U^n, S^n, Z^n) \in \mathcal{U}^n \times \mathcal{S}^n \times \mathcal{Z}^n$ , drawn i.i.d. according to  $\mathcal{P}_{USZ} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z})$ , as in (IV.71). We denote by  $U^n \in \mathcal{U}^n$  the sequence of source symbols,  $S^n \in \mathcal{S}^n$  the sequence of channel states symbols and  $Z^n \in \mathcal{Z}^n$  the sequence of decoder state information symbols. The state-dependent channel generates an output sequence  $Y^n \in \mathcal{Y}^n$  drawn i.i.d. according to  $\mathcal{T}_{Y|XS} \in \Delta(\mathcal{Y})^{|\mathcal{X} \times \mathcal{S}|}$ , as in (IV.72).

$$\Pr((U^n, S^n, Z^n) = (u^n, s^n, z^n)) = \prod_{t=1}^n \mathcal{P}_{USZ}(u_t, s_t, z_t), \quad \forall (u^n, s^n, z^n) \in \mathcal{U}^n \times \mathcal{S}^n \times \mathcal{Z}^n, \quad (\text{IV.71})$$

$$\Pr(Y^n = y^n | (X^n, S^n) = (x^n, s^n)) = \prod_{t=1}^n \mathcal{T}_{Y|XS}(y_t | x_t, s_t), \quad \forall (x^n, s^n, y^n) \in \mathcal{X}^n \times \mathcal{S}^n \times \mathcal{Y}^n. \quad (\text{IV.72})$$

In this section, we consider strictly causal, causal and non-causal encoding and decoding functions.

### IV.4.a Non-causal encoder and non-causal decoder

We consider the setting depicted in Fig. IV.10.



**Definition 19** Given  $n \in \mathbb{N}^*$ , we define the  $n$ -length code with non-causal encoder and non-causal decoder by a pair  $(\sigma, \tau)$  of deterministic functions

$$\sigma : \mathcal{U}^n \times \mathcal{S}^n \longrightarrow \mathcal{X}^n, \quad (\text{IV.73})$$

$$\tau : \mathcal{Y}^n \times \mathcal{Z}^n \longrightarrow \mathcal{V}^n. \quad (\text{IV.74})$$

We denote by  $\mathcal{C}^{\text{nc}}(n)$  the set of  $n$ -length code with non-causal encoder and non-causal decoder.

The code  $(\sigma, \tau) \in \mathcal{C}^{\text{nc}}(n)$  induces a probability distribution  $\mathcal{P}_{\mathcal{U}^n \mathcal{S}^n \mathcal{Z}^n \mathcal{X}^n \mathcal{Y}^n \mathcal{V}^n}^c$  defined, for all  $(u^n, s^n, z^n, x^n, y^n, v^n)$ , by

$$\begin{aligned} & \mathcal{P}_{\mathcal{U}^n \mathcal{S}^n \mathcal{Z}^n \mathcal{X}^n \mathcal{Y}^n \mathcal{V}^n}^c(u^n, s^n, z^n, x^n, y^n, v^n) \\ &= \left( \prod_{t=1}^n \mathcal{P}_{USZ}(u_t, s_t, z_t) \right) \mathbf{1}\{x^n = \sigma(u^n, s^n)\} \left( \prod_{t=1}^n \mathcal{T}_{Y|XS}(y_t | x_t, s_t) \right) \mathbf{1}\{v^n = \tau(y^n, z^n)\}. \end{aligned} \quad (\text{IV.75})$$

We evaluate the empirical frequencies of symbols induced by the code.

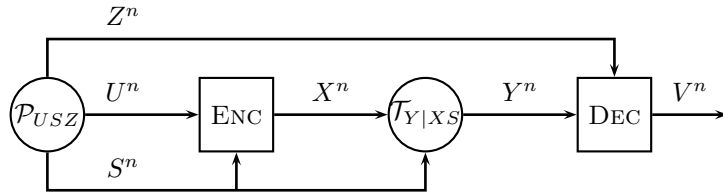


Figure IV.10 – Non-Causal Encoding function  $f : \mathcal{U}^n \times \mathcal{S}^n \rightarrow \mathcal{X}^n$  and Decoding function  $g : \mathcal{Y}^n \times \mathcal{Z}^n \rightarrow \mathcal{V}^n$ .

We refer to the Definition 13 for the random empirical distribution  $Q_{USZXYV}^n$ . We define the notion of achievable distribution similarly to Def. 14.

**Definition 20** A target probability distribution  $\mathcal{Q}_{USZXYV} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  is achievable if there exists a sequence of codes  $(\sigma, \tau) \in \mathcal{C}^{\text{nc}}(n)$  such that the sequence of empirical distributions  $(Q_{USZXYV}^n)_{n \in \mathbb{N}^*}$  converges in probability to  $\mathcal{Q}_{USZXYV}$ , i.e.

$$\forall \delta > 0, \quad \lim_{n \rightarrow +\infty} \min_{(\sigma, \tau) \in \mathcal{C}^{\text{nc}}(n)} \Pr \left( \left\| Q_{USZXYV}^n - \mathcal{Q}_{USZXYV} \right\|_1 > \delta \right) = 0. \quad (\text{IV.76})$$

We denote by  $\mathcal{A}^{\text{nc}} \subset \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  the set of achievable distributions.

We aim at characterizing the asymptotic behavior of the empirical distribution  $Q_{USZXYV}^n$  induced by sequences of codes.

**Theorem 11** 1. A target probability distribution  $\mathcal{Q}_{USZXYV} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  is achievable if

$$\mathcal{Q}_{USZXYV} = \mathcal{P}_{USZ} \mathcal{Q}_{X|US} \mathcal{T}_{Y|XS} \mathcal{Q}_{V|USZXY}, \quad (\text{IV.77})$$

and there exist an auxiliary random variable  $W \in \mathcal{W}$  with  $|\mathcal{W}| \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y}| + 2$  such that

$$\max_{\substack{\mathcal{Q}_{USZWXYZV} \\ \in \mathcal{Q}^{\text{nc}}(\mathcal{Q}_{V|USZXY})}} \left( I(W; Y, Z) - I(W; U, S) \right) \geq 0, \quad (\text{IV.78})$$

where for every  $\mathcal{Q}_{V|USZXY}$ ,

$$\begin{aligned} \mathcal{Q}^{\text{nc}}(\mathcal{Q}_{V|USZXY}) &= \left\{ \tilde{\mathcal{Q}}_{USZWXYZV} = \mathcal{P}_{USZ} \mathcal{Q}_{X|US} \tilde{\mathcal{Q}}_{W|USX} \mathcal{T}_{Y|XS} \tilde{\mathcal{Q}}_{V|YZW}, \right. \\ & \left. \mathcal{Q}_{V|USZXY} = \sum_{w \in \mathcal{W}} \tilde{\mathcal{Q}}_{W|USX}(w | \cdot, \cdot, \cdot) \tilde{\mathcal{Q}}_{V|YZW}(\cdot | \cdot, \cdot, w) \right\}. \end{aligned} \quad (\text{IV.79})$$

2. If a target probability distribution  $\mathcal{Q}_{USZXYV} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  is achievable, then

$$\mathcal{Q}_{USZXYV} = \mathcal{P}_{USZ} \mathcal{Q}_{X|US} \mathcal{T}_{Y|XS} \mathcal{Q}_{V|USZXY}, \quad (\text{IV.80})$$

and

$$\max_{\substack{\tilde{\mathcal{Q}}_{USZWXV} \\ \in \mathbb{Q}^{\text{nc}}(\mathcal{Q}_{V|USZXY})}} \left( \min (I(X, U, S; Y, Z), I(X, S; Y) + I(U; Z)) - I(W; U, S) \right) \geq 0. \quad (\text{IV.81})$$

The Theorem 11 (1.) is stated in (Le Treust, 2015c, [IC13], Theorem III.1) and in (Le Treust, 2015, [W1]). The Theorem 11 (2.) is stated in (Cervia et al., 2020, [J9], Theorem 2). When there is no state information  $S = Z = \emptyset$ , Theorem 11 (1.) boils down to (Cuff and Schieler, 2011, Theorem 1). Any distribution  $\tilde{\mathcal{Q}}_{USZWXV} \in \mathbb{Q}^{\text{nc}}(\mathcal{Q}_{V|USZXY})$  satisfies the Markov chains

$$Y \text{---} \ominus (X, S) \text{---} \ominus (U, Z, W), \quad (\text{IV.82})$$

$$Z \text{---} \ominus (U, S) \text{---} \ominus (X, Y, W), \quad (\text{IV.83})$$

$$V \text{---} \ominus (Y, Z, W) \text{---} \ominus (U, S, X). \quad (\text{IV.84})$$

**Remark 6** The mutual informations in (IV.81) and (IV.81) are continuous over the set of probability distributions  $\mathbb{Q}^{\text{nc}}(\mathcal{Q}_{V|USZXY})$ , for all  $\mathcal{Q}_{V|USZXY}$ . Moreover,  $\mathbb{Q}^{\text{nc}}(\mathcal{Q}_{V|USZXY})$  is compact since  $|\mathcal{W}| < +\infty$ . As mentioned in (Cuff, 2013, pp. 7083) and in (Wyner and Ziv, 1976, pp. 9), the supremum is a maximum in (IV.81) and (IV.81).

### Perfect Channel

We consider that the state-dependent channel is not noisy, i.e.  $\mathcal{T}_{Y|XS} = \mathbb{1}\{Y = X\}$  depicted in Fig. IV.11. We characterize the set of achievable distributions.

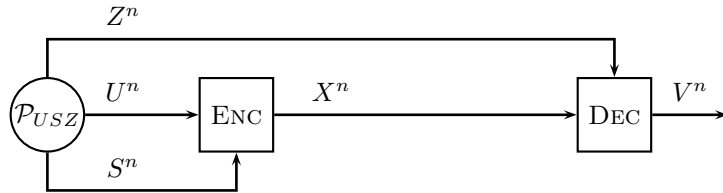


Figure IV.11 – The perfect channel is defined by  $\mathcal{T}_{Y|XS} = \mathbb{1}\{Y = X\}$  and the decoding is lossy.

**Theorem 12** A target probability distribution  $\mathcal{Q}_{USZXV} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V})$  is achievable if and only if

$$\mathcal{Q}_{USZXYV} = \mathcal{P}_{USZ} \mathcal{Q}_{X|US} \mathcal{Q}_{V|USZX}, \quad (\text{IV.85})$$

and there exist an auxiliary random variable  $W_2 \in \mathcal{W}_2$  with  $|\mathcal{W}_2| \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 2$  such that

$$\max_{\substack{\tilde{\mathcal{Q}}_{USZW_2XV} \\ \in \mathbb{Q}^{\text{P}}(\mathcal{Q}_{V|USZX})}} \left( I(W_2; Z|X) + H(X) - I(X, W_2; U, S) \right) \geq 0, \quad (\text{IV.86})$$

where for every  $\mathcal{Q}_{V|USZX}$ ,

$$\mathbb{Q}^{\text{P}}(\mathcal{Q}_{V|USZX}) = \left\{ \tilde{\mathcal{Q}}_{USZW_2XV} = \mathcal{P}_{USZ} \mathcal{Q}_{X|US} \tilde{\mathcal{Q}}_{W_2|USX} \tilde{\mathcal{Q}}_{V|XZW_2}, \right. \\ \left. \mathcal{Q}_{V|USZX} = \sum_{w_2 \in \mathcal{W}_2} \tilde{\mathcal{Q}}_{W_2|USX}(w_2|\cdot, \cdot, \cdot) \tilde{\mathcal{Q}}_{V|XZW_2}(\cdot|\cdot, \cdot, w_2) \right\}. \quad (\text{IV.87})$$

This result is stated in (Le Treust, 2015c, [IC13], Theorem IV.1) and in (Le Treust, 2015, [W1]). Any distribution  $\tilde{\mathcal{Q}}_{USZW_2XV} \in \mathbb{Q}^{\mathcal{P}}(\mathcal{Q}_{V|USZX})$  satisfies the Markov chains

$$Z \circlearrowleft (U, S) \circlearrowleft (X, W), \quad (\text{IV.88})$$

$$V \circlearrowleft (X, Z, W) \circlearrowleft (U, S). \quad (\text{IV.89})$$

**Remark 7** This result generalizes the coding theorem for state-dependent source of Wyner and Ziv (1976).

### Lossless reconstruction at the decoder

We restrict our attention to the case where the decoder must recover the source symbol, as depicted in Fig. IV.12. We characterize the set of achievable distributions  $\mathcal{Q}_{USZXYV}$  for which  $\mathcal{Q}_{V|USZXY} = \mathbb{1}\{V = U\}$ .

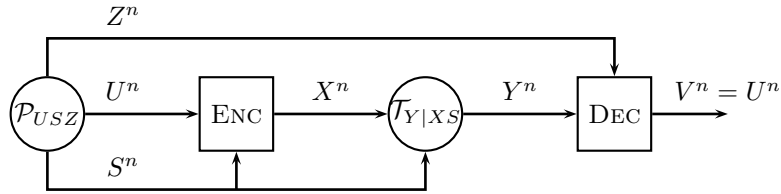


Figure IV.12 – The decoder must recover the source symbol, i.e.  $\mathcal{Q}_{V|USZXY} = \mathbb{1}\{V = U\}$ .

**Theorem 13** A target probability distribution  $\mathcal{Q}_{USZXY} \mathbb{1}\{V = U\} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$  is achievable if and only if

$$\mathcal{Q}_{USZXY} \mathbb{1}\{V = U\} = \mathcal{P}_{USZ} \mathcal{Q}_{X|US} \mathcal{T}_{Y|XS} \mathbb{1}\{V = U\}, \quad (\text{IV.90})$$

and there exist an auxiliary random variable  $W_1 \in \mathcal{W}_1$  with  $|\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 2$  such that

$$\max_{\tilde{\mathcal{Q}}_{USZW_1XYV} \in \mathbb{Q}^{\mathcal{I}}} \left( I(U, W_1; Y, Z) - I(W_1; S|U) - H(U) \right) \geq 0, \quad (\text{IV.91})$$

where

$$\mathbb{Q}^{\mathcal{I}} = \left\{ \tilde{\mathcal{Q}}_{USZW_2XYV} = \mathcal{P}_{USZ} \mathcal{Q}_{X|US} \tilde{\mathcal{Q}}_{W_1|USX} \mathcal{T}_{Y|XS} \mathbb{1}\{V = U\} \right\}. \quad (\text{IV.92})$$

This result is stated in (Le Treust, 2015c, [IC13], Eq. (11)) and in (Le Treust, 2015, [W1]). Any distribution  $\tilde{\mathcal{Q}}_{USZW_1XYV} \in \mathbb{Q}^{\mathcal{I}}$  satisfies the Markov chains

$$Y \circlearrowleft (X, S) \circlearrowleft (U, Z, W_1), \quad (\text{IV.93})$$

$$Z \circlearrowleft (U, S) \circlearrowleft (X, Y, W_1). \quad (\text{IV.94})$$

Another formulation is stated in (Le Treust, 2014, [IC12]) with a more restrictive lossless decoding constraint:  $\mathcal{P}(\hat{U}^n \neq U^n) \rightarrow 0$ .

**Remark 8** This result generalizes the coding theorem for state-dependent channel of Gel'fand and Pinsker (1980).

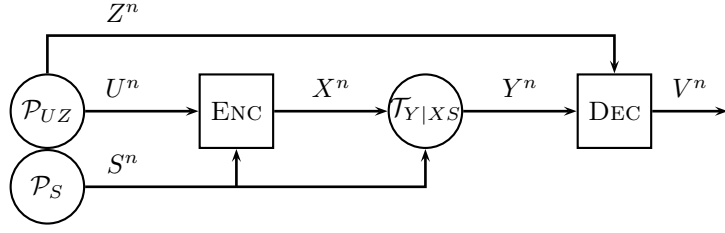


Figure IV.13 – The source random variables  $(U, Z, V)$  are independent of the channel random variables  $(X, S, Y)$ .

### Separation between Source and Channel

We restrict our attention to the case where the source distribution satisfies  $\mathcal{P}_{USZ} = \mathcal{P}_{UZ}\mathcal{P}_S$  and we investigate the achievability of target distributions such that the random variables of the source  $(U, Z, V)$  are independent of the random variables of the channel  $(X, S, Y)$ , as depicted in Fig. IV.13. We characterize the set of achievable distributions of the form  $\mathcal{Q}_{UZV}\mathcal{Q}_{XSY}$ .

**Theorem 14** *A target probability distribution  $\mathcal{Q}_{UZV}\mathcal{Q}_{XSY} \in \Delta(\mathcal{U} \times \mathcal{Z} \times \mathcal{V}) \times \Delta(\mathcal{X} \times \mathcal{S} \times \mathcal{Y})$  is achievable if and only if*

$$\mathcal{Q}_{UZV} = \mathcal{P}_{UZ}\mathcal{Q}_{V|UZ}, \quad \mathcal{Q}_{XSY} = \mathcal{P}_S\mathcal{Q}_{X|S}\mathcal{T}_{Y|XS}, \quad (\text{IV.95})$$

and there exist a pair of auxiliary random variables  $(W_1, W_2) \in \mathcal{W}_1 \times \mathcal{W}_2$  with  $|\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{Z} \times \mathcal{V}| + 2$  and  $|\mathcal{W}_2| \leq |\mathcal{S} \times \mathcal{X} \times \mathcal{Y}| + 2$  such that

$$\max_{\substack{\tilde{\mathcal{Q}}_{USZW_1W_2XYV} \\ \in \mathcal{Q}^s(\mathcal{Q}_{V|UZ})}} \left( I(W_1; Y) - I(W_1; S) + I(W_2; Z) - I(W_2; U) \right) \geq 0, \quad (\text{IV.96})$$

where for every  $\mathcal{Q}_{V|UZ}$ ,

$$\begin{aligned} \mathcal{Q}^s(\mathcal{Q}_{V|UZ}) &= \left\{ \tilde{\mathcal{Q}}_{USZW_1W_2XYV} = \mathcal{P}_{UZ}\tilde{\mathcal{Q}}_{W_2|U}\tilde{\mathcal{Q}}_{V|ZW_2}\mathcal{P}_S\mathcal{Q}_{X|S}\tilde{\mathcal{Q}}_{W_1|XS}\mathcal{T}_{Y|XS}, \right. \\ &\quad \left. \mathcal{Q}_{V|UZ} = \sum_{w_2 \in \mathcal{W}_2} \tilde{\mathcal{Q}}_{W_2|U}(w_2|\cdot)\tilde{\mathcal{Q}}_{V|ZW_2}(\cdot|w_2) \right\}. \end{aligned} \quad (\text{IV.97})$$

This result is stated in (Le Treust, 2015c, [IC13], Theorem IV.2) and (Le Treust, 2015, [W1]). Any distribution  $\tilde{\mathcal{Q}}_{USZW_1W_2XYV} \in \mathcal{Q}^s$  satisfies  $(U, Z, W_2, V)$  is independent of  $(S, X, W_1, Y)$  and the Markov chains

$$Z \text{ -- } U \text{ -- } W_2, \quad (\text{IV.98})$$

$$V \text{ -- } (Z, W_2) \text{ -- } U, \quad (\text{IV.99})$$

$$Y \text{ -- } (X, S) \text{ -- } W_1. \quad (\text{IV.100})$$

**Remark 9** *This separation result was already stated in Merhav and Shamai (2003), given a distortion level and a channel cost.*

### IV.4.b Causal decoder and non-causal encoder

We consider the case where the decoder is causal, as depicted in Fig. IV.14.

**Definition 21** Given  $n \in \mathbb{N}^*$ , we define the  $n$ -length code with non-causal encoder and causal decoder by a pair  $(\sigma, \tau)$  where  $\tau = \{\tau_t\}_{t \in \{1, \dots, n\}}$  of deterministic functions

$$\sigma : \mathcal{U}^n \times \mathcal{S}^n \longrightarrow \mathcal{X}^n, \quad (\text{IV.101})$$

$$\tau_t : \mathcal{Y}^t \times \mathcal{Z}^t \longrightarrow \mathcal{V}, \quad \forall t \in \{1, \dots, n\}. \quad (\text{IV.102})$$

We denote by  $\mathcal{C}^{\text{cd}}(n)$  the set of  $n$ -length code with non-causal encoder and causal decoder.

The code  $(\sigma, \tau) \in \mathcal{C}^{\text{cd}}(n)$  induces a probability distribution  $\mathcal{P}_{U^n S^n Z^n X^n Y^n V^n}^c$  defined, for all  $(u^n, s^n, z^n, x^n, y^n, v^n)$ , by

$$\begin{aligned} & \mathcal{P}_{U^n S^n Z^n X^n Y^n V^n}^c(u^n, s^n, z^n, x^n, y^n, v^n) \\ &= \left( \prod_{t=1}^n \mathcal{P}_{USZ}(u_t, s_t, z_t) \right) \mathbb{1}\{x^n = \sigma(u^n, s^n)\} \left( \prod_{t=1}^n \mathcal{T}_{Y|XS}(y_t | x_t, s_t) \right) \left( \prod_{t=1}^n \mathbb{1}\{v_t = \sigma_t(y^t, z^t)\} \right). \end{aligned} \quad (\text{IV.103})$$

We refer to the Definitions 13 and 14 for the random empirical distribution  $Q_{USZXYV}^n$  and the notion of achievable distribution. We denote by  $\mathcal{A}^{\text{cd}}$  the set of target distributions that are achievable by using a code  $(\sigma, \tau) \in \mathcal{C}^{\text{cd}}(n)$  with non-causal encoder and causal decoder.

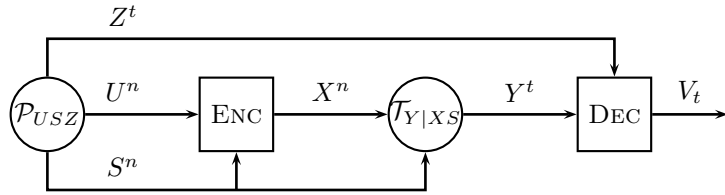


Figure IV.14 – The encoding function  $\sigma$  is non-causal and the decoding function is causal  $\tau_t$  for  $t \in \{1, \dots, n\}$ .

**Theorem 15** A target probability distribution  $Q_{USZXYV} \in \mathcal{A}^{\text{cd}}$  if and only if

$$Q_{USZXYV} = \mathcal{P}_{USZ} Q_{X|US} \mathcal{T}_{Y|XS} Q_{V|USZXY}, \quad (\text{IV.104})$$

and there exist a pair of auxiliary random variables  $(W_1, W_2) \in \mathcal{W}_1 \times \mathcal{W}_2$  with  $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 2$  such that

$$\max_{\substack{Q_{USZW_1W_2XYV} \\ \in \mathcal{Q}^{\text{cd}}(Q_{V|USZXY})}} \left( I(W_1; Y, Z | W_2) - I(W_1, W_2; U, S) \right) \geq 0, \quad (\text{IV.105})$$

where for every  $Q_{V|USZXY}$ ,

$$\begin{aligned} \mathcal{Q}^{\text{cd}}(Q_{V|USZXY}) &= \left\{ \tilde{Q}_{USZW_1W_2XYV} = \mathcal{P}_{USZ} Q_{X|US} \tilde{Q}_{W_1W_2|USX} \mathcal{T}_{Y|XS} \tilde{Q}_{V|YZW_2}, \right. \\ & \left. Q_{V|USZXY} = \sum_{w_2 \in \mathcal{W}_2} \tilde{Q}_{W_2|USX}(w_2 | \cdot, \cdot, \cdot) \tilde{Q}_{V|YZW_2}(\cdot | \cdot, \cdot, w) \right\}. \end{aligned} \quad (\text{IV.106})$$

This result is stated in (Le Treust, 2015c, [IC13], Eq. (16)) and in (Le Treust, 2015, [W1]). Any distribution  $\tilde{Q}_{USZW_1W_2XYV} \in \mathcal{Q}^{\text{cd}}$  satisfies the Markov chains

$$Y \text{---} (X, S) \text{---} (U, Z, W_1, W_2), \quad (\text{IV.107})$$

$$Z \text{---} (U, S) \text{---} (X, Y, W_1, W_2), \quad (\text{IV.108})$$

$$V \text{---} (Y, Z, W_2) \text{---} (U, S, X, W_1). \quad (\text{IV.109})$$

**Strictly causal decoder and non-causal encoder**

**Definition 22** Given  $n \in \mathbb{N}^*$ , we define the  $n$ -length code with non-causal encoder and strictly causal decoder by a pair  $(\sigma, \tau)$  where  $\tau = \{\tau_t\}_{t \in \{1, \dots, n\}}$  of deterministic functions

$$\sigma : \mathcal{U}^n \times \mathcal{S}^n \longrightarrow \mathcal{X}^n, \quad (\text{IV.110})$$

$$\tau_t : \mathcal{Y}^{t-1} \times \mathcal{Z}^{t-1} \longrightarrow \mathcal{V}, \quad \forall t \in \{1, \dots, n\}. \quad (\text{IV.111})$$

We denote by  $\mathcal{C}^{\text{sd}}(n)$  the set of  $n$ -length code with non-causal encoder and strictly causal decoder.

We denote by  $\mathcal{A}^{\text{sd}}$ , the set of target distributions that are achievable by using a code  $(\sigma, \tau) \in \mathcal{C}^{\text{sd}}(n)$  with non-causal encoder and strictly causal decoder.

**Theorem 16** A target probability distribution  $Q_{USZXYV} \in \mathcal{A}^{\text{sd}}$  if and only if

$$Q_{USZXYV} = P_{USZ} Q_{XV|US} \mathcal{T}_{Y|XS} \quad (\text{IV.112})$$

and there exist an auxiliary random variables  $W_1 \in \mathcal{W}_1$  with  $|\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 2$  such that

$$\max_{\tilde{Q}_{USZW_1XYV} \in \mathcal{Q}^{\text{sd}}} \left( I(W_1; Y, Z|V) - I(W_1, V; U, S) \right) \geq 0, \quad (\text{IV.113})$$

where

$$\mathcal{Q}^{\text{sd}} = \left\{ \tilde{Q}_{USZW_1XYV} = P_{USZ} Q_{XV|US} \tilde{Q}_{W_1|USXV} \mathcal{T}_{Y|XS} \right\}. \quad (\text{IV.114})$$

This result is stated in (Le Treust, 2015c, [IC13], Eq. (16)) and in (Le Treust, 2015, [W1]). The solution of (IV.113) can be obtained by replacing  $W_2$  by  $V$  in the solution of (IV.105) in Theorem 15. Any distribution  $\tilde{Q}_{USZW_1XYV} \in \mathcal{Q}^{\text{sd}}$  satisfies the Markov chains

$$Y \circlearrowleft (X, S) \circlearrowleft (U, Z, W_1, V), \quad (\text{IV.115})$$

$$Z \circlearrowleft (U, S) \circlearrowleft (X, Y, W_1, V). \quad (\text{IV.116})$$

**IV.4.c Causal encoder and non-causal decoder**

We consider the case where the decoder is causal, as depicted in Fig. IV.15.

**Definition 23** Given  $n \in \mathbb{N}^*$ , we define the  $n$ -length code with causal encoder and non-causal decoder by a pair  $(\sigma, \tau)$  where  $\sigma = \{\sigma_t\}_{t \in \{1, \dots, n\}}$  of deterministic functions

$$\sigma_t : \mathcal{U}^t \times \mathcal{S}^t \longrightarrow \mathcal{X}, \quad \forall t \in \{1, \dots, n\}, \quad (\text{IV.117})$$

$$\tau : \mathcal{Y}^n \times \mathcal{Z}^n \longrightarrow \mathcal{V}^n, \quad (\text{IV.118})$$

We denote by  $\mathcal{C}^{\text{ce}}(n)$  the set of  $n$ -length code with causal encoder and non-causal decoder.

The code  $(\sigma, \tau) \in \mathcal{C}^{\text{ce}}(n)$  induces a probability distribution  $\mathcal{P}_{U^n S^n Z^n X^n Y^n V^n}^c$  defined, for all  $(u^n, s^n, z^n, x^n, y^n, v^n)$ , by

$$\begin{aligned} & \mathcal{P}_{U^n S^n Z^n X^n Y^n V^n}^c(u^n, s^n, z^n, x^n, y^n, v^n) \\ &= \left( \prod_{t=1}^n \mathcal{P}_{USZ}(u_t, s_t, z_t) \right) \left( \prod_{t=1}^n \mathbb{1}\{x_t = \sigma_t(u^t, s^t)\} \right) \left( \prod_{t=1}^n \mathcal{T}_{Y|XS}(y_t|x_t, s_t) \right) \mathbb{1}\{v^n = \tau(y^n, z^n)\}. \end{aligned} \quad (\text{IV.119})$$

We refer to the Definitions 13 and 14 for the random empirical distribution  $Q_{USZXYV}^n$  and the notion of achievable distribution. We denote by  $\mathcal{A}^{\text{ce}}$  the set of target distributions that are achievable by using a code  $(\sigma, \tau) \in \mathcal{C}^{\text{ce}}(n)$  with non-causal encoder and causal decoder.

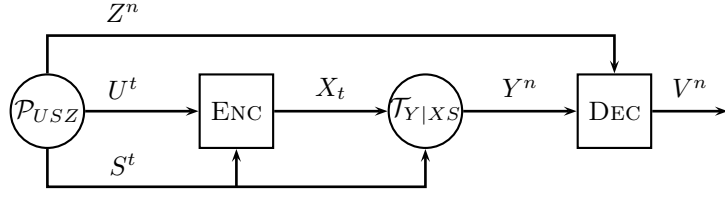


Figure IV.15 – The encoding function is causal  $\sigma_t$  for  $t \in \{1, \dots, n\}$  and the decoding function  $\tau$  is non-causal.

**Theorem 17** A target probability distribution  $\mathcal{Q}_{USZXYV} \in \mathcal{A}^{\text{ce}}$  if and only if

$$\mathcal{Q}_{USZXYV} = \mathcal{P}_{USZ} \mathcal{Q}_{X|US} \mathcal{T}_{Y|XS} \mathcal{Q}_{V|USZXY}, \quad (\text{IV.120})$$

and there exist a pair of auxiliary random variables  $(W_1, W_2) \in \mathcal{W}_1 \times \mathcal{W}_2$  with  $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 2$  such that

$$\max_{\substack{\tilde{\mathcal{Q}}_{USZW_1W_2XYV} \\ \in \mathcal{Q}^{\text{ce}}(\mathcal{Q}_{USZXYV})}} \left( I(W_1, W_2; Y, Z) - I(W_2; U, S|W_1) \right) \geq 0, \quad (\text{IV.121})$$

where for every  $\mathcal{Q}_{USZXYV}$ ,

$$\mathcal{Q}^{\text{ce}}(\mathcal{Q}_{USZXYV}) = \left\{ \begin{aligned} \tilde{\mathcal{Q}}_{USZW_1W_2XYV} &= \mathcal{P}_{USZ} \tilde{\mathcal{Q}}_{W_1} \tilde{\mathcal{Q}}_{X|USW_1} \tilde{\mathcal{Q}}_{W_2|USW_1} \mathcal{T}_{Y|XS} \tilde{\mathcal{Q}}_{V|YZW_1W_2}, \\ \mathcal{Q}_{USZXYV} &= \sum_{(w_1, w_2) \in \mathcal{W}_1 \times \mathcal{W}_2} \tilde{\mathcal{Q}}_{USZW_1W_2XYV}. \end{aligned} \right. \quad (\text{IV.122})$$

This result is stated in (Le Treust, 2015c, [IC13], Theorem V.1) and in (Le Treust, 2015, [W1]). Any distribution  $\tilde{\mathcal{Q}}_{USZW_1W_2XYV} \in \mathcal{Q}^{\text{ce}}$  satisfies the properties

$$(U, S) \text{ are independent of } W_1, \quad (\text{IV.123})$$

$$X \perp\!\!\!\perp (U, S, W_1) \perp\!\!\!\perp W_2, \quad (\text{IV.124})$$

$$Y \perp\!\!\!\perp (X, S) \perp\!\!\!\perp (U, Z, W_1, W_2), \quad (\text{IV.125})$$

$$Z \perp\!\!\!\perp (U, S) \perp\!\!\!\perp (X, Y, W_1, W_2), \quad (\text{IV.126})$$

$$V \perp\!\!\!\perp (Y, Z, W_1, W_2) \perp\!\!\!\perp (U, S, X). \quad (\text{IV.127})$$

### Strictly causal encoder and non-causal decoder

**Definition 24** Given  $n \in \mathbb{N}^*$ , we define the  $n$ -length code with causal encoder and non-causal decoder by a pair  $(\sigma, \tau)$  where  $\sigma = \{\sigma_t\}_{t \in \{1, \dots, n\}}$  of deterministic functions

$$\sigma_t : \mathcal{U}^{t-1} \times \mathcal{S}^{t-1} \longrightarrow \mathcal{X}, \quad \forall t \in \{1, \dots, n\}, \quad (\text{IV.128})$$

$$\tau : \mathcal{Y}^n \times \mathcal{Z}^n \longrightarrow \mathcal{V}^n, \quad (\text{IV.129})$$

We denote by  $\mathcal{C}^{\text{se}}(n)$  the set of  $n$ -length code with causal encoder and non-causal decoder.

We denote by  $\mathcal{A}^{\text{se}}$ , the set of target distributions that are achievable by using a code  $(\sigma, \tau) \in \mathcal{C}^{\text{se}}(n)$  with non-causal encoder and strictly causal decoder.

**Theorem 18** A target probability distribution  $\mathcal{Q}_{USZXYV} \in \mathcal{A}^{\text{se}}$  if and only if

$$\mathcal{Q}_{USZXYV} = \mathcal{P}_{USZ} \mathcal{Q}_X \mathcal{T}_{Y|XS} \mathcal{Q}_{V|USZXY} \quad (\text{IV.130})$$

and there exist an auxiliary random variables  $W_2 \in \mathcal{W}_2$  with  $|\mathcal{W}_2| \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 2$  such that

$$\max_{\tilde{Q}_{USZW_2XYV} \in \mathbb{Q}^{\text{se}}} \left( I(X, W_2; Y, Z) - I(W_2; U, S|X) \right) \geq 0, \quad (\text{IV.131})$$

where

$$\mathbb{Q}^{\text{se}} = \left\{ \tilde{Q}_{USZW_2XYV} = \mathcal{P}_{USZ} \mathcal{Q}_X \tilde{Q}_{W_2|USX} \mathcal{T}_{Y|XS} \tilde{Q}_{V|YZXW_2}, \right. \\ \left. \tilde{Q}_{V|USZXY} = \sum_{w_2 \in \mathcal{W}_2} \tilde{Q}_{W_2|USX}(w_2|\cdot, \cdot, \cdot) \tilde{Q}_{V|YZXW_2}(\cdot|\cdot, \cdot, \cdot, w) \right\}. \quad (\text{IV.132})$$

This result is stated in (Le Treust, 2015c, [IC13], Theorem V.1) and in (Le Treust, 2015, [W1]). The solution of (IV.131) can be obtained by replacing  $W_1$  by  $X$  in the solution of (IV.121) in Theorem 17. Any distribution  $\tilde{Q}_{USZW_1XYV} \in \mathbb{Q}^{\text{se}}$  satisfies

$$(U, S) \text{ are independent of } X, \quad (\text{IV.133})$$

$$Y \perp\!\!\!\perp (X, S) \perp\!\!\!\perp (U, Z, W_2), \quad (\text{IV.134})$$

$$Z \perp\!\!\!\perp (U, S) \perp\!\!\!\perp (X, Y, W_2), \quad (\text{IV.135})$$

$$V \perp\!\!\!\perp (Y, Z, X, W_2) \perp\!\!\!\perp (U, S). \quad (\text{IV.136})$$

## IV.5 Chapter summary

In this chapter, we investigate the coordination problem posed by Gossner et al. (2006), and revisited in the Information Theory literature by Cuff et al. (2010). We characterize the optimal solutions to various point-to-point scenarios, e.g. when the decoder or the encoder are strictly causal, causal, non-causal, when the source and the channel are state-dependent, and with source feedforward and channel output feedback.

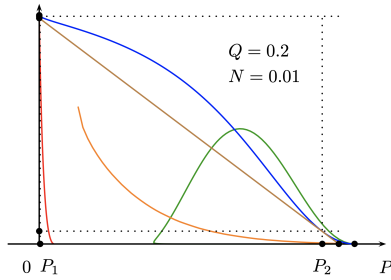
One setting remains unsolved, when both encoder and decoder perform non-causally. This open question seems related to the distributed lossy source coding problem, in which the pair of random variables  $(U, Y)$  are correlated information sources observed by the encoder and the decoder. Inner and outer bounds for the distributed lossy source coding problem are stated in Berger (1978) and Tung (1978).

In the next chapter, we present several extensions and applications of the coordination results.



# Chapter V

## Impact of empirical coordination results



In this chapter, we explore the consequences of the results of Chap. IV. The first section presents the contributions for the channel state masking problem, stated in (Le Treust and Bloch, 2021, [J10]), (Le Treust and Bloch, 2016, [IC16]), (Le Treust and Bloch, 2019, [NC9]). In the second section, we use coordination coding for a decentralized control problem called the Witsenhausen’s counterexample. This corresponds to (Le Treust and Oechtering, 2021, [IC27]), (Oechtering and Le Treust, 2019, [IC23]), (Le Treust and Oechtering, 2018, [IC22]), (Le Treust and Oechtering, 2022, [W9]). The third section presents some of Giulia Cervia’s PhD contributions on strong coordination problems, i.e. when the probability distribution induced by the coding converges to a target i.i.d. probability distribution. This corresponds to (Cervia et al., 2020, [J9]), (Cervia et al., 2018, [IC21]), (Cervia et al., 2017b, [IC19]), (Cervia et al., 2016, [IC17]), (Cervia et al., 2017a, [NC7]).

### Contents

---

V.1	State leakage and coordination with causal state knowledge at the encoder . . . . .	53
V.1.a	Problem statement . . . . .	54
V.1.b	Characterization of achievable triples of rate, leakage and distribution . . . . .	55
V.1.c	Channel state estimation via cost function . . . . .	57
V.1.d	Extensions to more general scenarios . . . . .	59
V.1.e	Strictly causal encoding . . . . .	61
V.2	Decentralized control: Witsenhausen’s counterexample . . . . .	62
V.2.a	System model . . . . .	63
V.2.b	Coding Result . . . . .	64
V.2.c	Control schemes of particular interest . . . . .	65
V.2.d	Numerical Results . . . . .	68
V.2.e	Discussion . . . . .	69
V.3	Strong coordination of signals and actions . . . . .	70
V.3.a	Inner and outer bounds for the strong coordination region . . . . .	72
V.3.b	Strong coordination region with state and side information . . . . .	73
V.3.c	Characterization of the strong coordination region for special cases . . . . .	75
V.3.d	Is separation optimal? . . . . .	78
V.4	Chapter summary . . . . .	81

---

## V.1 State leakage and coordination with causal state knowledge at the encoder

The study of state-dependent channels can be traced back to the early works of Shannon (1958) and Gel'fand and Pinsker (1980), which identified optimal coding strategies to transmit reliably in the presence of a state known at the encoder causally or non-causally, respectively. The insights derived from the models have since proved central to the study of diverse topics including wireless communications, see Costa (1983); Keshet et al. (2008), information-hiding and watermarking as in Moulin and O'Sullivan (2003), and information transmission in repeated games in Gossner et al. (2006). The present work relates to the latter application and studies state-dependent channels with causal state knowledge from the perspective of *empirical coordination* Cuff et al. (2010).

Previous studies that have explored the problem of not only decoding messages at the receiver but also estimating the channel state, are particularly relevant to the present work. The *state masking* formulation

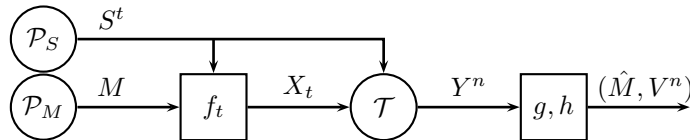


Figure V.1 – The memoryless channel  $\mathcal{T}_{Y|XS}$  depends on the state drawn i.i.d. according to  $\mathcal{P}_S$ . The encoding function is causal  $f_t : \mathcal{M} \times \mathcal{S}^t \rightarrow \mathcal{X}$ , for all  $t \in \{1, \dots, n\}$  and the decoding functions  $g : \mathcal{Y}^n \rightarrow \mathcal{M}$  and  $h : \mathcal{Y}^n \rightarrow \Delta(\mathcal{V}^n)$  are non-causal.

of the problem, see Merhav and Shamai (2007), aims at characterizing the trade-off between the rate of reliable communication and the minimal leakage about the channel state. The rate-leakage capacity region of state masking has been successfully characterized for both causal and non-causal state knowledge. The *state amplification* formulation of Kim et al. (2008), in which the state is conveyed to the receiver instead of being masked, aims at characterizing the trade-off between the rate of reliable communication and the reduction of uncertainty about the state. The rate-uncertainty reduction capacity region of state amplification has also been successfully characterized for causal and non-causal state knowledge. The state amplification formulation was subsequently extended in the causal case by replacing the reduction of uncertainty about the state by an average cost function in Choudhuri et al. (2013) (this model was dubbed *causal state communication*). Note that, in such a scenario, the channel output feedback at the encoder increases the region of achievable rate-cost pairs, as in Bross and Lapidot (2018). The rate-cost capacity region of state communication has been successfully characterized for causal and strictly causal state knowledge, and has been characterized for noiseless and noisy non-causal state knowledge in the case of Gaussian channels with a quadratic cost, see Sutivong et al. (2005); Tian et al. (2015). Both formulations have been combined in Koyluoglu et al. (2016) to study the trade-off between amplification and leakage rates in a channel with two receivers having opposite objectives. The amplification-leakage capacity region has been investigated for non-causal state knowledge, via generally non-matching inner and outer bounds. As a perhaps more concrete example, in Tutuncuoglu et al. (2014), the authors have studied the trade-off between amplification and leakage in the context of an energy harvesting scenario. An extreme situation of state masking, called state obfuscation, in which the objective is to make the channel output sequence nearly independent of the channel states, has recently been investigated in Wang and Wornell (2020).

We revisit the problems of state masking and state amplification with causal and strictly causal state knowledge through the lens of *empirical coordination*, see Cuff et al. (2010), Kramer and Savari (2007). Empirical coordination refers to the control of the joint histograms of the various sequences such as states, codewords, that appear in channel models, and is related to the coordination of autonomous decision makers in game theory as in Gossner et al. (2006). Specifically, the study of empirical coordination over state-dependent channels is a proxy for characterizing the utility of autonomous decision makers playing a repeated game in the presence of an environment variable (the state), random Gossner et al. (2006); Gossner and Vieille (2002) or adversarial Gossner et al. (2009); Gossner and Tomala (2006, 2007),

and of an observation structure (the channel) describing how agents observe each other's actions. The characterization of the empirical coordination capacity requires the design of coding schemes in which the actions of the decision makers are sequences that embed coordination information. The empirical coordination capacity has been studied for state-dependent channels under different constraints including strictly causal and causal encoding Cuff and Schieler (2011), for perfect channel Cuff and Zhao (2011), for strictly causal and causal decoding (Le Treust, 2017, [J6]), with source feedforward Larrousse et al. (2018), for lossless decoding (Le Treust, 2014, [IC12]), with secrecy constraint Schieler and Cuff (2014), with two-sided state information (Le Treust, 2015c, [IC13]) and with channel feedback (Le Treust, 2015b, [IC14]). Empirical coordination is also a powerful tool for controlling the Bayesian posterior beliefs of the decoder, e.g. in the problems of Bayesian persuasion (Le Treust and Tomala, 2019, [J8]) and strategic communication (Le Treust and Tomala, 2019, [W3]).

The main contribution of the present work is to show that empirical coordination provides a natural framework in which to jointly study the problems of reliable communication, state masking, and state amplification. This connection highlights some of the benefits of empirical coordination beyond those already highlighted in earlier works Cuff and Schieler (2011)–(Le Treust, 2015b, [IC14]). In particular, we obtain the following.

- We introduce and characterize the notion of *core of the receiver's knowledge*, which captures what the decoder can exactly know about the other variables involved in the system. For instance, this allows us to characterize the rate-leakage-coordination region for the causal state-dependent channel (Theorem 19). Our definition of leakage refines previous work by exactly characterizing the leakage rate instead of only providing a single-sided bound. When specialized, our result (Theorem 20) simultaneously recovers the constraints already established both in (Merhav and Shamai, 2007, Section V) and (Kim et al., 2008, Theorem 2).
- We revisit the problem of causal state communication and characterize the normalized KL-divergence between the decoder's posterior beliefs and a target belief induced by coordination (Theorem 21). This allows us to characterize the rate-cost trade-off for a zero-sum game, in which the decoder attempts to estimate the state while the encoder tries to mask it (Theorem 22).
- We extend the results to other models, including two-sided state information (Theorem 23), noisy feedback (Theorem 24), and strictly causal encoding (Theorem 25).

The rest of the section is organized as follows. In Section V.1.a, we formally introduce the model, along with necessary definitions and notation, and we state our main results. In Section V.1.c, we investigate the channel state estimation problem by introducing the KL-divergence and the decoder's posterior beliefs. In Section V.1.d and Section V.1.e, we present some extensions of our results to different scenarios.

### V.1.a Problem statement

The problem under investigation is illustrated in Figure V.1. A uniformly distributed message represented by the random variable  $M \in \mathcal{M}$  is to be transmitted over a state dependent memoryless channel characterized by the conditional distribution  $\mathcal{T}_{Y|XS}$  and a channel state  $S \in \mathcal{S}$  drawn according to the i.i.d. distribution  $\mathcal{P}_S$ . For  $n \in \mathbb{N}^*$ , the message  $M$  and the state sequence  $S^n$  are encoded into a codeword  $X^n \in \mathcal{X}^n$  using an encoder, subject to causal constraints to be precised later. Upon observing the output  $Y^n \in \mathcal{Y}^n$  of the noisy channel, the receiver uses a decoder to form an estimate  $\hat{M} \in \mathcal{M}$  of  $M$  and to generate actions  $V^n \in \mathcal{V}^n$ , whose exact role will be precised shortly. For now,  $V^n$  can be thought of as an estimate of the state sequence  $S^n$  but more generally captures the ability of the receiver to coordinate with the transmitter and the channel state. Both  $\mathcal{T}_{Y|XS}$  and  $\mathcal{P}_S$  are assumed known to all parties. We are specifically interested in causal encoders formally defined as follows.

**Definition 25** A code with causal encoding  $(f, g, h)$  consists of

$$f_t : \mathcal{M} \times \mathcal{S}^t \longrightarrow \Delta(\mathcal{X}), \quad \forall t \in \{1, \dots, n\}, \quad (\text{V.1})$$

$$g : \mathcal{Y}^n \longrightarrow \mathcal{M}, \quad (\text{V.2})$$

$$h : \mathcal{Y}^n \longrightarrow \Delta(\mathcal{V}^n), \quad (\text{V.3})$$

where  $f = (f_t)_{t \in \{1, \dots, n\}}$  are stochastic encoding functions,  $g$  is a deterministic decoding function and  $h$  is a stochastic receiver action function. The set of codes with causal encoding with length  $n$  and message set  $\mathcal{M}$  is denoted  $\mathcal{C}_c(n, \mathcal{M})$ .

A code  $(f, g, h) \in \mathcal{C}_c(n, \mathcal{M})$ , the uniform distribution of the messages  $\mathcal{P}_M$ , the source  $\mathcal{P}_S$  and the channel  $\mathcal{T}_{Y|XS}$ , induce a distribution on  $(M, S^n, X^n, Y^n, V^n, \hat{M})$  given by

$$\mathcal{P}_M \prod_{t=1}^n \left[ \mathcal{P}_{S_t} f_{X_t|S^t M} \mathcal{T}_{Y_t|X_t S_t} \right] h_{V^n|Y^n} \mathbb{1}\{\hat{M} = g(Y^n)\}. \quad (\text{V.4})$$

Since the sequences  $(S^n, X^n, Y^n, V^n)$  are random, the empirical distribution  $Q_{SXYV}^n$  is also a random variable. The performance of codes is measured as follows.

**Definition 26** Fix a target rate  $R \geq 0$ , a target state leakage  $E \geq 0$  and a target distribution  $Q_{SXYV}$ . The triple  $(R, E, Q_{SXYV})$  is achievable if for all  $\varepsilon > 0$ , there exists  $\bar{n} \in \mathbb{N}^*$ , for all  $n \geq \bar{n}$ , there exists a code  $(f, g, h) \in \mathcal{C}_c(n, \mathcal{M})$  that satisfies

$$\begin{aligned} \frac{\log_2 |\mathcal{M}|}{n} &\geq R - \varepsilon, \\ \left| \mathcal{L}_\varepsilon(c) - E \right| &\leq \varepsilon, \quad \text{with} \quad \mathcal{L}_\varepsilon(c) = \frac{1}{n} I(S^n; Y^n), \\ \mathcal{P}_\varepsilon(c) &= \Pr(M \neq \hat{M}) + \Pr\left(\left\| Q_{SXYV}^n - Q_{SXYV} \right\|_1 > \varepsilon\right) \leq \varepsilon. \end{aligned}$$

We denote by  $\mathcal{A}_c$  the set of achievable triples  $(R, E, Q_{SXYV})$ .

In layman's term, performance is captured along three metrics: i) the rate at which the message  $M$  can be reliably transmitted; ii) the information leakage rate about the state sequence  $S^n$  at the receiver; and iii) the ability of the encoder to coordinate with the receiver, captured by the empirical coordination with respect to  $Q_{SXYV}$ . The need to coordinate with receiver action  $V$  is motivated by problems in which the terminals represent decision makers that choose actions  $(X, V)$  as a function of the system state  $S$ , as in Gossner et al. (2006). The state can also be used to represent a system to control, in which case coordination also ties to the Witsenhausen's counterexample Grover et al. (2015), (Oechtering and Le Treust, 2019, [IC23]).

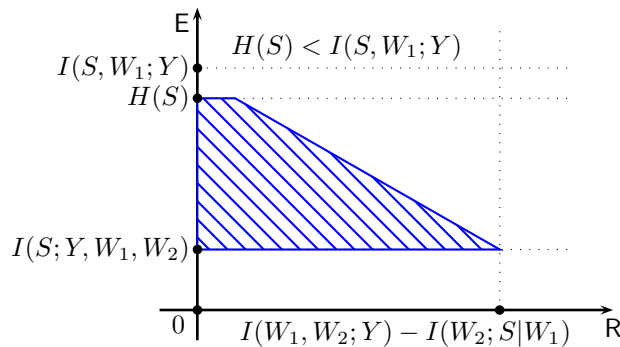


Figure V.2 – The region of achievable  $(R, E) \in \mathcal{A}_c$  for a given distribution  $Q_{SW_1W_2XYV}$  for which  $H(S) < I(S, W_1; Y)$ .

### V.1.b Characterization of achievable triples of rate, leakage and distribution

We now state the main result.

**Theorem 19** Consider a target distribution  $\mathcal{Q}_{SXYV}$  that decomposes as  $\mathcal{Q}_{SXYV} = \mathcal{P}_S \mathcal{Q}_{X|S} \mathcal{T}_{Y|XS} \mathcal{Q}_{V|SXY}$ . Then,  $(R, E, \mathcal{Q}_{SXYV}) \in \mathcal{A}_c$  if and only if there exist two auxiliary random variables  $(W_1, W_2)$  with distribution  $\mathcal{Q}_{SW_1W_2XYV} \in \mathcal{Q}_c$  satisfying

$$I(S; W_1, W_2, Y) \leq E \leq H(S), \quad (\text{V.5})$$

$$R + E \leq I(W_1, S; Y), \quad (\text{V.6})$$

where  $\mathcal{Q}_c$  is the set of distributions  $\mathcal{Q}_{SW_1W_2XYV}$  with marginal  $\mathcal{Q}_{SXYV}$  that decompose as

$$\mathcal{P}_S \mathcal{Q}_{W_1} \mathcal{Q}_{W_2|SW_1} \mathcal{Q}_{X|SW_1} \mathcal{T}_{Y|XS} \mathcal{Q}_{V|YW_1W_2}, \quad (\text{V.7})$$

and such that  $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq |\mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 1$ .

The achievability and converse proofs are provided in (Le Treust and Bloch, 2021, [J10]). The key idea behind the achievability proof is the following. The encoder operates in a Block-Markov fashion to ensure that the transmitted signals, the state, the received sequence, and the receiver actions are coordinated subject to the causal constraint at the encoder. This requires the use of two auxiliary codebooks, captured by the auxiliary random variables  $W_1$  and  $W_2$ , where the first codebook is used for reliable communication while the second one is used to coordinate with the state. Simultaneously, the encoder quantizes the channel state and transmits carefully chosen bin indices on top of its messages to finely control how much the receiver can infer about the channel state. The region of achievable pairs  $(R, E)$  is depicted in Fig. V.2 for a given distribution  $\mathcal{Q}_{SW_1W_2XYV}$ , assuming  $H(S) < I(S, W_1; Y)$ .

**Remark 10** Equation (V.6) and the first inequality of (V.5) imply the information constraints of (Choudhuri et al., 2013, Theorem 3) for causal state communication and of (Cuff and Schieler, 2011, Theorem 2) for empirical coordination.

$$R \leq I(W_1, W_2; Y) - I(W_2; S|W_1). \quad (\text{V.8})$$

Indeed, both Markov chains  $X \circlearrowleft (S, W_1) \circlearrowleft W_2$  and  $Y \circlearrowleft (X, S) \circlearrowleft (W_1, W_2)$  imply  $Y \circlearrowleft (W_1, S) \circlearrowleft W_2$ .

Theorem 19 has several important consequences. First, the coordination of both encoder and decoder actions according to  $\mathcal{P}_S \mathcal{Q}_{X|S} \mathcal{T}_{Y|XS} \mathcal{Q}_{V|SXY}$  is compatible with the reliable transmission of additional information at rate  $R \geq 0$ . Second, the case of equality in the right-hand-side inequality of (V.5) corresponds to the full disclosure of the channel state  $S$  to the decoder. Third, for any  $(R, \mathcal{Q}_{SXYV})$ , the minimal state leakage  $E^*(R, \mathcal{Q}_{SXYV})$  such that  $(R, E^*(R, \mathcal{Q}_{SXYV}), \mathcal{Q}_{SXYV}) \in \mathcal{A}_c$ , if it exists, is given by

$$E^*(R, \mathcal{Q}_{SXYV}) = \min_{\substack{\mathcal{Q}_{SW_1W_2XYV} \in \mathcal{Q}_c, \\ \text{s.t. } R \leq I(W_1, W_2; Y) - I(W_2; S|W_1)}} I(S; W_1, W_2, Y). \quad (\text{V.9})$$

The reliable transmission of information requires the decoder to know the encoding function, from which it can estimate the channel state  $S$ . In Section V.1.c, we investigate the relationship between the state leakage  $\mathcal{L}_e(c)$  and the decoder's posterior belief  $\mathcal{P}_{S^n|Y^n}$  induced by the encoding process.

### Special case without receiver actions

We now assume that the decoder does not return an action  $V$  coordinated with the other symbols  $(S, X, Y)$ , in order to compare our setting with the problems of “state masking” (Merhav and Shamai, 2007, Section V) and “state amplification” (Kim et al., 2008, Section IV). Note that these earlier works involve slightly different notions of achievable state leakage. In Merhav and Shamai (2007), the state leakage is upper bounded by  $\mathcal{L}_e(c) = \frac{1}{n} I(S^n; Y^n) \leq E + \varepsilon$ . In Kim et al. (2008), the decoder forms a list  $L_n(Y^n) \subseteq \mathcal{S}^n$  with cardinality  $\log_2 |L_n(Y^n)| = H(S) - E$  such that the list decoding error probability  $\Pr(S^n \notin L_n(Y^n)) \leq \varepsilon$  is small, hence reducing the uncertainty about the state. Here, we require the leakage  $\mathcal{L}_e(c) = \frac{1}{n} I(S^n; Y^n)$  induced by the code to be controlled by  $|\mathcal{L}_e(c) - E| \leq \varepsilon$ . Nevertheless, we shall see that our definition allows us to obtain the rate constraints of Kim et al. (2008); Merhav and Shamai (2007) as extreme cases.

**Definition 27** A code without receiver actions consists of stochastic encoding functions  $f_t : \mathcal{M} \times \mathcal{S}^t \rightarrow \Delta(\mathcal{X})$ ,  $\forall t \in \{1, \dots, n\}$  and a deterministic decoding function  $g : \mathcal{Y}^n \rightarrow \mathcal{M}$ . The set of such codes with length  $n$  and message set  $\mathcal{M}$  is denoted  $\mathcal{C}_d(n, \mathcal{M})$ . The corresponding set of achievable triples  $(R, E, \mathcal{Q}_{SXY})$  is defined as in Definition 26 and is denoted  $\mathcal{A}_d$ .

Note that the target distribution is here restricted to  $\mathcal{Q}_{SXY} \in \Delta(\mathcal{S} \times \mathcal{X} \times \mathcal{Y})$  since the receiver does not take an action.

**Theorem 20** Consider a target distribution  $\mathcal{Q}_{SXY}$  that decomposes as  $\mathcal{Q}_{SXY} = \mathcal{P}_S \mathcal{Q}_{X|S} \mathcal{T}_{Y|XS}$ . Then,  $(R, E, \mathcal{Q}_{SXY}) \in \mathcal{A}_d$  if and only if there exists an auxiliary random variable  $W_1$  with distribution  $\mathcal{Q}_{SW_1XY} \in \mathbb{Q}_d$  that satisfies

$$I(S; W_1, Y) \leq E \leq H(S), \quad (\text{V.10})$$

$$R + E \leq I(W_1, S; Y), \quad (\text{V.11})$$

where  $\mathbb{Q}_d$  is the set of distributions  $\mathcal{Q}_{SW_1XY}$  with marginal  $\mathcal{Q}_{SXY}$  that decompose as

$$\mathcal{P}_S \mathcal{Q}_{W_1} \mathcal{Q}_{X|SW_1} \mathcal{T}_{Y|XS}, \quad (\text{V.12})$$

and such that  $|W_1| \leq |\mathcal{S} \times \mathcal{Y}| + 1$ .

The achievability proof is obtained from Theorem 19 by setting  $W_2 = \emptyset$  and by considering a single block coding instead of block-Markov coding. The converse proof is similar to the converse of Theorem 19 and is provided in the Supplementary Materials.

**Remark 11** When setting  $W_2 = \emptyset$ , (V.8) in Remark 10 simplifies to

$$R \leq I(W_1; Y), \quad (\text{V.13})$$

which, together with the first inequality in (V.10), coincides with the information constraints of (Merhav and Shamai, 2007, pp. 2260). Furthermore, (V.13), (V.11) and the second inequality of (V.10) correspond to the region  $\mathcal{R}_0$  stated in (Kim et al., 2008, Lemma 3). Formally, the region characterized by Theorem 20 is the intersection of the regions identified in (Merhav and Shamai, 2007, pp. 2260) and (Kim et al., 2008, Lemma 3).

### V.1.c Channel state estimation via cost function

#### Decoder posterior belief

In this section, we provide an upper bound on the KL-divergence between the decoder posterior belief  $\mathcal{P}_{S^n|Y^n}$  induced by an encoding, and the target conditional distribution  $\mathcal{Q}_{S|YW_1W_2}$ .

**Theorem 21 (Channel state estimation)** Assume that the distribution  $\mathcal{Q}_{SW_1W_2XY}$  has full support. For any conditional distribution  $\mathcal{P}_{W_1^n W_2^n X^n | S^n}$ , we have

$$\frac{1}{n} D \left( \mathcal{P}_{S^n | Y^n} \left\| \prod_{t=1}^n \mathcal{Q}_{S_t | Y_t W_{1,t} W_{2,t}} \right\| \right) \quad (\text{V.14})$$

$$\leq \mathcal{L}_\epsilon(c) - I(S; W_1, W_2, Y) + \alpha_1 \delta + \alpha_2 \Pr \left( \left\| \mathcal{Q}_{SW_1W_2Y}^n - \mathcal{Q}_{SW_1W_2Y} \right\|_1 > \delta \right), \quad (\text{V.15})$$

where the constant parameters  $\alpha_1 = \sum_{s, w_1, w_2, y} \log_2 \frac{1}{\mathcal{Q}(s|w_1, w_2, y)}$  and  $\alpha_2 = \log_2 \frac{1}{\min_{s, y, w_1, w_2} \mathcal{Q}(s|y, w_1, w_2)}$  are strictly positive.

The proof of Theorem 21 is given in (Le Treust and Bloch, 2021, [J10]). Consider a target leakage  $E = I(S; W_1, W_2, Y)$  and a pair  $(R, \mathcal{Q}_{SXYV})$ , and assume there exists a distribution  $\mathcal{Q}_{SW_1W_2XYV} \in \mathbb{Q}_c$  with full support, satisfying (V.5) and (V.6). By Theorem 19, for all  $\epsilon > 0$  and for all  $\delta > 0$ , there

exists  $\bar{n} \in \mathbb{N}^*$  such that for all  $n \geq \bar{n}$  there exists a code  $(f, g, h) \in \mathcal{C}(n, \mathcal{M})$  with two auxiliary sequences  $(W_1^n, W_2^n)$ , such that

$$\begin{aligned} & \left| \mathcal{L}_\epsilon(c) - I(S; W_1, W_2, Y) \right| \leq \epsilon \quad \text{and} \\ & \Pr\left(\left\| Q_{SW_1W_2Y}^n - \mathcal{Q}_{SW_1W_2Y} \right\|_1 > \delta\right) \leq \epsilon. \end{aligned} \quad (\text{V.16})$$

Hence, by Theorem 21 we have

$$\frac{1}{n} D\left(\mathcal{P}_{S^n|Y^n} \left\| \prod_{t=1}^n \mathcal{Q}_{S_t|Y_t W_{1,t} W_{2,t}}\right.\right) \leq \epsilon + \alpha_1 \delta + \alpha_2 \epsilon, \quad (\text{V.17})$$

where  $\epsilon$  and  $\delta$  may go to zero when  $n$  goes to infinity. This shows that the posterior belief  $\mathcal{P}_{S^n|Y^n}$  approaches the single-letter distribution  $\mathcal{Q}_{S|YW_1W_2}$ . Based on the triple of symbols  $(Y, W_1, W_2)$ , the decoder generates action  $V$  using the conditional distribution  $\mathcal{Q}_{V|YW_1W_2}$ , and infers the channel state  $S$  according to the conditional distribution  $\mathcal{Q}_{S|YW_1W_2}$ . In that regard, the random variables  $(Y, W_1, W_2)$  capture the "core of the receiver's knowledge," regarding other random variables  $S$  and  $V$ . The bound on the KL-divergence in (V.15) relates to the notion of strategic distance (Gossner and Vieille, 2002, Section 5.2), later used in several articles on repeated game Gossner and Tomala (2006), Gossner and Tomala (2007), Gossner et al. (2009), on Bayesian persuasion (Le Treust and Tomala, 2019, [J8]) and on strategic communication (Le Treust and Tomala, 2019, [W3]).

### Channel state estimation zero-sum game

We now introduce a cost function  $c : \mathcal{S} \times \mathcal{V} \rightarrow \mathbb{R}$  and the channel state estimation zero-sum game, in which the encoder and decoder are *opponents*, i.e. the decoding is selected in order to minimize the cost function, which is maximize by the encoder. Although the encoder and the decoder cooperate in transmitting reliably at rate  $R$ , the encoder seeks to prevent the decoder from returning a good estimate  $v \in \mathcal{V}$  of the channel state  $s \in \mathcal{S}$  by maximizing the expected long-run cost, while the decoder attempts to minimize it.

**Definition 28** *A target rate  $R \geq 0$  and a target cost  $C \geq 0$  are achievable if for all  $\epsilon > 0$ , there exists  $\bar{n} \in \mathbb{N}^*$  such that for all  $n \geq \bar{n}$ , there exists a code in  $\mathcal{C}_d(n, \mathcal{M})$  such that*

$$\frac{\log_2 |\mathcal{M}|}{n} \geq R - \epsilon, \quad (\text{V.18})$$

$$\Pr\left(M \neq \hat{M}\right) \leq \epsilon, \quad (\text{V.19})$$

$$\left| \min_{h_{V^n|Y^n}} \frac{1}{n} \sum_{t=1}^n \mathbb{E}\left[c(S_t, V_t)\right] - C \right| \leq \epsilon. \quad (\text{V.20})$$

We denote by  $\mathcal{A}_g$  the set of achievable pairs  $(R, C) \in \mathcal{A}_g$ .

**Theorem 22 (Zero-sum game)** *A pair of rate and cost  $(R, C) \in \mathcal{A}_g$  is achievable if and only if there exists an auxiliary random variable  $W_1$  with distribution  $\mathcal{Q}_{SW_1XY} \in \mathcal{Q}_d$  that satisfies*

$$R \leq I(W_1; Y), \quad (\text{V.21})$$

$$C = \min_{\mathcal{P}_{V|W_1Y}} \mathbb{E}\left[c(S, V)\right], \quad (\text{V.22})$$

where the set  $\mathcal{Q}_d$  is defined in Theorem 20.

The proof of Theorem 22 is provided in (Le Treust and Bloch, 2021, [J10]). The achievability proof is a consequence of Theorem 20 and Theorem 21, and of (Le Treust and Tomala, 2019, [J8], Lemma A.8, Lemma A.21).

**Remark 12 (Maximin-minimax result)** *The optimal cost-rate function  $C^*(R)$  reformulates as a maximin problem*

$$\begin{aligned} C^*(R) &= \max_{\substack{\mathcal{Q}_{W_1}, \mathcal{Q}_{X|SW_1} \\ R \leq I(W_1; Y)}} \min_{\mathcal{P}_{V|W_1Y}} \mathbb{E}[c(S, V)] \\ &= \min_{\mathcal{P}_{V|W_1Y}} \max_{\substack{\mathcal{Q}_{W_1}, \mathcal{Q}_{X|SW_1} \\ R \leq I(W_1; Y)}} \mathbb{E}[c(S, V)]. \end{aligned} \quad (\text{V.23})$$

*The maximum and the minimum are taken over compact and convex sets and the cost function is linear. Hence by Sion's Theorem [Sion \(1958\)](#) the maximin is equal to the minimax and the value of this channel state estimation zero-sum game is  $C^*(R)$ .*

**Remark 13 (One auxiliary random variable)** *The formulation of [Theorem 22](#) is based on the set of distributions  $\mathcal{Q}_d$  with only one auxiliary random variable  $W_1$ , instead of the two random variables  $(W_1, W_2)$  of the set  $\mathcal{Q}_c$ . When the encoder tries to mask the channel state, it does not require the auxiliary random variable  $W_2$  anymore, since*

$$C^\circ = \max_{\substack{\mathcal{Q}_{W_1}, \mathcal{Q}_{X|SW_1}, \mathcal{Q}_{W_2|SW_1} \\ R \leq I(W_1, W_2; Y) - I(W_2; S|W_1)}} \min_{\mathcal{P}_{V|W_1W_2Y}} \mathbb{E}[c(S, V)] \quad (\text{V.24})$$

$$\leq \max_{\substack{\mathcal{Q}_{W_1}, \mathcal{Q}_{X|SW_1}, \mathcal{Q}_{W_2|SW_1} \\ R \leq I(W_1, W_2; Y) - I(W_2; S|W_1)}} \min_{\mathcal{P}_{V|W_1Y}} \mathbb{E}[c(S, V)] \quad (\text{V.25})$$

$$\leq \max_{\substack{\mathcal{Q}_{W_1}, \mathcal{Q}_{X|SW_1} \\ R \leq I(W_1; Y)}} \min_{\mathcal{P}_{V|W_1Y}} \mathbb{E}[c(S, V)] = C^*, \quad (\text{V.26})$$

*where [\(V.25\)](#) comes from taking the minimum over  $\mathcal{P}_{V|W_1Y}$  instead of  $\mathcal{P}_{V|W_1W_2Y}$ ; [\(V.26\)](#) comes from the Markov chain  $Y \ominus (S, W_1) \ominus W_2$  stated in [\(V.7\)](#), that ensures  $I(W_1, W_2; Y) - I(W_2; S|W_1) \leq I(W_1; Y)$ . Hence, the information constraint  $R \leq I(W_1, W_2; Y) - I(W_2; S|W_1)$  is more restrictive than  $R \leq I(W_1; Y)$ .*

**Remark 14 (Zero rate case)** *In the special case  $R = 0$ , which corresponds to a channel estimation game without communication, the encoding functions reduce to  $f_{X_t|S^t}$  instead of  $f_{X_t|S^tM}$ . The channel state estimation zero-sum game becomes the maximin problem*

$$\max_{\{f_{X_t|S^t}\}_{t \in \{1, \dots, n\}}} \min_{h_{V^n|Y^n}} \mathbb{E}\left[\frac{1}{n} \sum_{t=1}^n c(S_t, V_t)\right], \quad (\text{V.27})$$

*in which the encoder chooses  $\{f_{X_t|S^t}\}_{t \in \{1, \dots, n\}}$  and the decoder chooses  $h_{V^n|Y^n}$ . [Theorem 22](#) shows that the single-letter solution is  $\max_{\mathcal{Q}_{W_1}, \mathcal{Q}_{X|SW_1}} \min_{\mathcal{P}_{V|W_1Y}} \mathbb{E}[c(S, V)]$ .*

*If the objectives of both encoder and decoder were aligned, i.e., they would both try to minimize the long term average cost*

$$\min_{\substack{\{f_{X_t|S^t}\}_{t \in \{1, \dots, n\}}, \\ h_{V^n|Y^n}}} \mathbb{E}\left[\frac{1}{n} \sum_{t=1}^n c(S_t, V_t)\right], \quad (\text{V.28})$$

*the problem would become the causal channel state communication studied in [Choudhuri et al. \(2013\)](#).*

## V.1.d Extensions to more general scenarios

### Two-sided state information

The case of two-sided state information is illustrated in [Fig. V.3](#). The channel state  $S^n$ , information source  $U^n$  and decoder state information  $Z^n$  are jointly distributed according to the i.i.d. distribution  $\mathcal{P}_{USZ} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z})$ .



**Definition 29** A code with two-sided state information consists of stochastic functions  $f_t : \mathcal{M} \times \mathcal{U}^t \times \mathcal{S}^t \rightarrow \Delta(\mathcal{X}) \forall t \in \{1, \dots, n\}$ , a deterministic decoding function  $g : \mathcal{Y}^n \times \mathcal{Z}^n \rightarrow \mathcal{M}$ , and a stochastic receiver action function  $h : \mathcal{Y}^n \times \mathcal{Z}^n \rightarrow \Delta(\mathcal{V}^n)$ . The set of codes with causal encoding with length  $n$  and message set  $\mathcal{M}$  is denoted  $\mathcal{C}_s(n, \mathcal{M})$ .

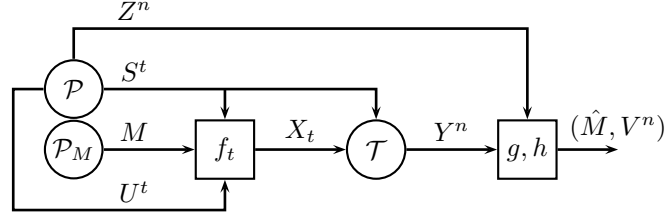


Figure V.3 – The causal encoding function is  $f_t : \mathcal{M} \times \mathcal{U}^t \times \mathcal{S}^t \rightarrow \mathcal{X}$ , for all  $t \in \{1, \dots, n\}$  and the non-causal decoding functions are  $g : \mathcal{Y}^n \times \mathcal{Z}^n \rightarrow \mathcal{M}$  and  $h : \mathcal{Y}^n \times \mathcal{Z}^n \rightarrow \Delta(\mathcal{V}^n)$ .

A code  $(f, g, h) \in \mathcal{C}_s(n, \mathcal{M})$ , the uniform distribution of the messages  $\mathcal{P}_M$ , the source  $\mathcal{P}_{USZ}$  and the channel  $\mathcal{T}_{Y|XS}$  induce a distribution on  $(M, U^n, S^n, Z^n, X^n, Y^n, V^n, \hat{M})$  given by

$$\mathcal{P}_M \prod_{t=1}^n \left[ \mathcal{P}_{U_t S_t Z_t} f_{X_t|U^t S^t M} \mathcal{T}_{Y_t|X_t S_t} \right] h_{V^n|Y^n Z^n} \mathbb{1}\{\hat{M} = g(Y^n, Z^n)\}. \quad (\text{V.29})$$

We denote by  $\mathcal{A}_s$  the set of achievable triples  $(R, E, \mathcal{Q}_{USZXYV})$ , defined similarly as in Definition 26.

**Theorem 23 (Two-sided state information)** Consider a target distribution  $\mathcal{Q}_{USZXYV}$  that decomposes as  $\mathcal{Q}_{USZXYV} = \mathcal{P}_{USZ} \mathcal{Q}_{X|US} \mathcal{T}_{Y|XS} \mathcal{Q}_{V|USZXY}$ . Then,  $(R, E, \mathcal{Q}_{USZXYV}) \in \mathcal{A}_s$  if and only if there exist two auxiliary random variables  $(W_1, W_2)$  with distribution  $\mathcal{Q}_{USZW_1W_2XYV} \in \mathcal{Q}_s$  satisfying

$$I(U, S; W_1, W_2, Y, Z) \leq E \leq H(U, S), \quad (\text{V.30})$$

$$R + E \leq I(W_1, U, S; Y, Z), \quad (\text{V.31})$$

where  $\mathcal{Q}_s$  is the set of distributions  $\mathcal{Q}_{USZW_1W_2XYV}$  that decompose as

$$\mathcal{P}_{USZ} \mathcal{Q}_{W_1} \mathcal{Q}_{W_2|USW_1} \mathcal{Q}_{X|USW_1} \mathcal{T}_{Y|XS} \mathcal{Q}_{V|YZW_1W_2}, \quad (\text{V.32})$$

and such that  $\max(|W_1|, |W_2|) \leq d + 1$  with  $d = |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}|$ .

The achievability proof follows directly from the proof of Theorem 19, by replacing the random variable of the channel state  $S$  by the pair  $(U, S)$  and the random variable of the channel output  $Y$  by the pair  $(Y, Z)$ . The converse proof is provided in the Supplementary Materials.

**Remark 15** The Markov chains  $X \circlearrowleft (U, S, W_1) \circlearrowleft W_2$ ,  $Y \circlearrowleft (X, S) \circlearrowleft (U, Z, W_1, W_2)$  and  $Z \circlearrowleft (U, S) \circlearrowleft (X, Y, W_1, W_2)$  imply another Markov chain property  $(Y, Z) \circlearrowleft (W_1, U, S) \circlearrowleft W_2$ . Indeed, for all  $(u, s, z, w_1, w_2, x, y)$  we have

$$\begin{aligned} & \mathcal{P}(y, z|w_1, w_2, u, s) \\ &= \sum_{x \in \mathcal{X}} \mathcal{Q}(x|u, s, w_1) \mathcal{T}(y|x, s) \mathcal{P}(z|u, s) = \mathcal{P}(y, z|w_1, u, s). \end{aligned}$$

By combining (V.30) and (V.31) with the Markov chain  $(Y, Z) \circlearrowleft (W_1, U, S) \circlearrowleft W_2$ , we recover the information constraint of (Le Treust, 2015c, [IC13], Theorem V.1):

$$R \leq I(W_1, W_2; Y, Z) - I(W_2; U, S|W_1). \quad (\text{V.33})$$

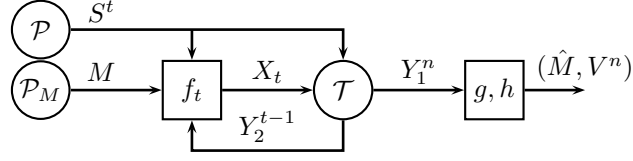


Figure V.4 – The noisy feedback sequence  $Y_2^{t-1}$  is drawn i.i.d. according to  $\mathcal{T}_{Y_1 Y_2 | X S}$ . The encoding is  $f_t : \mathcal{M} \times \mathcal{S}^t \times \mathcal{Y}_2^{t-1} \rightarrow \mathcal{X}$ ,  $\forall t \in \{1, \dots, n\}$ .

### Noisy channel feedback observed by the encoder

In this section, we consider that the encoder has noisy feedback  $Y_2$  from the state-dependent channel  $\mathcal{T}_{Y_1 Y_2 | X S}$ , as depicted in Fig. V.4. The encoding function becomes  $f_t : \mathcal{M} \times \mathcal{S}^t \times \mathcal{Y}_2^{t-1} \rightarrow \mathcal{X}$ ,  $\forall t \in \{1, \dots, n\}$  while the decoding functions and the definition of the state leakage remain unchanged. The corresponding set of achievable triples  $(R, E, \mathcal{Q}_{S X Y_1 Y_2 V})$  is denoted  $\mathcal{A}_f$ .

**Theorem 24 (Noisy channel feedback)** *We consider a target distribution  $\mathcal{Q}_{S X Y_1 Y_2 V}$  that decomposes as  $\mathcal{Q}_{S X Y_1 Y_2 V} = \mathcal{P}_S \mathcal{Q}_{X | S} \mathcal{T}_{Y_1 Y_2 | X S} \mathcal{Q}_{V | S X Y_1 Y_2}$ . Then,  $(R, E, \mathcal{Q}_{S X Y_1 Y_2 V}) \in \mathcal{A}_f$  if and only if there exist two auxiliary random variables  $(W_1, W_2)$  with distribution  $\mathcal{Q}_{S W_1 W_2 X Y_1 Y_2 V} \in \mathcal{Q}_f$  that satisfy*

$$R \leq I(W_1, W_2; Y_1) - I(W_2; S, Y_2 | W_1), \quad (\text{V.34})$$

$$I(S; W_1, W_2, Y_1) \leq E \leq H(S), \quad (\text{V.35})$$

$$R + E \leq I(W_1, S; Y_1), \quad (\text{V.36})$$

where  $\mathcal{Q}_f$  is the set of distributions with marginal  $\mathcal{Q}_{S W_1 W_2 X Y_1 Y_2 V}$  that decompose as

$$\mathcal{P}_S \mathcal{Q}_{W_1} \mathcal{Q}_{X | S W_1} \mathcal{T}_{Y_1 Y_2 | X S} \mathcal{Q}_{W_2 | S W_1 Y_2} \mathcal{Q}_{V | Y_1 W_1 W_2},$$

and such that  $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq d + 1$  with  $d = |\mathcal{S} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{V}|$ .

The achievability proof of Theorem 24 follows directly from the proof of Theorem 19, by replacing the pair  $(S^n, W_1^n)$  by the triple  $(S^n, W_1^n, Y_2^n)$  in order to select  $W_2^n$ . The decoding functions and the leakage analysis remain unchanged. The converse proof is stated in the Supplementary Materials.

**Remark 16 (Noisy feedback improves coordination)** *The channel feedback increases the set of achievable triples, i.e.  $\mathcal{A}_c \subset \mathcal{A}_f$ , since the conditional distribution  $\mathcal{Q}_{W_2 | S W_1 Y_2}$  depends on channel outputs  $Y_2$ . The information constraints of Theorem 24 are reduced to that of Theorem 19 since  $\mathcal{Q}_{W_2 | S W_1 Y_2} = \mathcal{Q}_{W_2 | S W_1} \iff W_2 \oplus (S, W_1) \oplus Y_2 \iff I(W_2; Y_2 | S, W_1) = 0$ . This was already pointed out for the coordination problem in (Le Treust, 2015b, [IC14]), and for the rate-and-state capacity problem in Bross and Lapidoth (2018).*

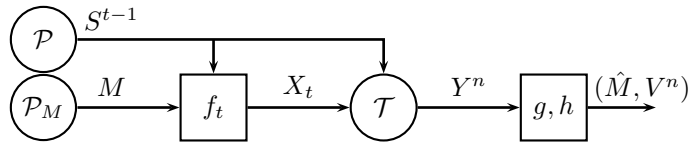


Figure V.5 – The strictly causal encoding function is  $f_t : \mathcal{M} \times \mathcal{S}^{t-1} \rightarrow \Delta(\mathcal{X})$ , for all  $t \in \{1, \dots, n\}$  and the non-causal decoding functions are  $g : \mathcal{Y}^n \rightarrow \mathcal{M}$  and  $h : \mathcal{Y}^n \rightarrow \Delta(\mathcal{V}^n)$ .

### V.1.e Strictly causal encoding

**Definition 30** *A code with strictly causal encoding consists of stochastic encoding functions  $f_t : \mathcal{M} \times \mathcal{S}^{t-1} \rightarrow \Delta(\mathcal{X})$   $\forall t \in \{1, \dots, n\}$ , a deterministic decoding function  $g : \mathcal{Y}^n \rightarrow \mathcal{M}$ , and a stochastic*

receiver action function  $h : \mathcal{Y}^n \rightarrow \Delta(\mathcal{V}^n)$ . The set of codes with strictly causal encoding with length  $n$  and message set  $\mathcal{M}$  is denoted  $\mathcal{C}_{\text{sc}}(n, \mathcal{M})$ . The corresponding set of achievable triples  $(R, E, \mathcal{Q}_{SXYV})$  is defined similarly as in Definition 26 and is denoted  $\mathcal{A}_{\text{sc}}$ .

**Theorem 25 (Strictly causal encoding)** Consider a target distribution  $\mathcal{Q}_{SXYV}$  that decomposes as  $\mathcal{Q}_{SXYV} = \mathcal{P}_S \mathcal{Q}_X \mathcal{T}_{Y|XS} \mathcal{Q}_{V|SXY}$ . Then,  $(R, E, \mathcal{Q}_{SXYV}) \in \mathcal{A}_{\text{sc}}$  if and only if there exists an auxiliary random variable  $W_2$  with distribution  $\mathcal{Q}_{SW_2XYV} \in \mathcal{Q}_{\text{sc}}$  that satisfies

$$I(S; X, W_2, Y) \leq E \leq H(S), \quad (\text{V.37})$$

$$R + E \leq I(X, S; Y), \quad (\text{V.38})$$

where  $\mathcal{Q}_{\text{sc}}$  is the set of distributions  $\mathcal{Q}_{SW_2XYV}$  with marginal  $\mathcal{Q}_{SW_2XYV}$  that decompose as

$$\mathcal{Q}_{SW_2XYV} = \mathcal{P}_S \mathcal{Q}_X \mathcal{Q}_{W_2|SX} \mathcal{T}_{Y|XS} \mathcal{Q}_{V|XYW_2} \quad (\text{V.39})$$

and such that  $|\mathcal{W}_2| \leq |\mathcal{S} \times \mathcal{X} \times \mathcal{Y}| + 1$ .

The achievability proof is obtained from Theorem 19 by replacing the auxiliary random variable  $W_1$  by the channel input  $X$ . The converse proof is provided in the Supplementary Materials.

**Remark 17** Equation (V.38), the first inequality of (V.37), the Markov chain  $Y \text{---} (X, S) \text{---} W_2$ , and the independence between  $S$  and  $X$  imply

$$R \leq I(X, W_2; Y) - I(W_2; S|X). \quad (\text{V.40})$$

**Corollary 4 (Without receiver's outputs)** A pair of rate and state leakage  $(R, E)$  is achievable if and only if there exists a distribution  $\mathcal{Q}_X$  that satisfies

$$I(S; Y|X) \leq E \leq H(S), \quad (\text{V.41})$$

$$R + E \leq I(X, S; Y). \quad (\text{V.42})$$

The achievability proof of Corollary 4 comes from the achievability proof of Theorem 25. The converse proof is based on standard arguments. Equations (V.41) and (V.42) imply  $R \leq I(X; Y)$ .

## V.2 Decentralized control: Witsenhausen's counterexample

In this section, we investigate the impact of the coordination results of Chap. IV in decentralized control problem. In Witsenhausen (1968), the author introduce a famous counterexample that shows that the best affine policy is outperformed by non-linear policies. Since then the example serves as study object illustrating the importance of the information pattern in distributed decision making, see Yüksel and Başar (2013) for a comprehensive discussion.

The first vector-valued extension considering a non-causal setup was studied in Grover and Sahai (2008) followed by a series of works, e.g. Grover et al. (2013); Grover and Sahai (2010); Grover et al. (2015). A comprehensive overview on the corresponding results is provided in Le Treust and Oechtering (2018), where we also discuss its close relation to the coordination problem. Optimal coding schemes for relevant setups are derived in Le Treust (2017), which also provides a review on the related literature. Most of the results were derived for finite alphabet setup where the concept of strong typicality provides the Markov Lemma. A rigorous extension to the Gaussian case has been done in Grover and Sahai (2010). The main result was proved to be optimal in Choudhuri and Mitra (2012b).

In this work we extend the finite alphabet coding scheme based on the concept of weak typicality, as in Cover and Thomas (2006) that we extend so that the need of the Markov Lemma can be avoided. Conceptually, the extension is similar to an extension as done in Vu et al. (2019), where also the approach of Wyner (1978) on how to analyze the average estimation error has been used. We extend this approach in this work since the second decision maker estimates the interim state and not the i.i.d. source.

In the following we show that also in this vector-valued setup the best affine policies can be outperformed by non-linear policies. In more detail, there exists parameter configurations where our coordination coding outperforms a simple *amplification* strategy.

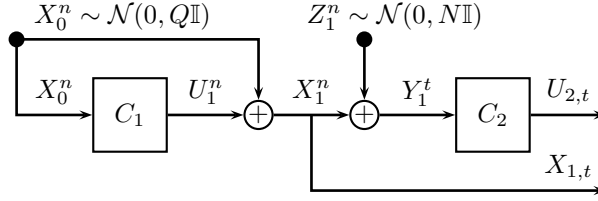


Figure V.6 – The state and the channel noise are drawn according to the i.i.d. Gaussian distributions  $X_0^n \sim \mathcal{N}(0, Q\mathbb{I})$  and  $Z_1^n \sim \mathcal{N}(0, N\mathbb{I})$ .

### V.2.a System model

We consider the vector-valued Witsenhausen setup depicted in Fig. V.6. The notations  $\mathcal{X}_0, \mathcal{U}_1, \mathcal{X}_1, \mathcal{Y}_1, \mathcal{U}_2$  stand for the sets of states, channel inputs, interim states, channel outputs, receiver outputs, that are all equal to the real line  $\mathbb{R}$ . For  $n \in \mathbb{N}^*$ , the  $n$ -time Cartesian product of sets is denoted by  $\mathcal{X}_0^n$ . The sequences of states and channel noises are drawn independently according to the i.i.d. Gaussian distributions  $X_0^n \sim \mathcal{N}(0, Q\mathbb{I})$  and  $Z_1^n \sim \mathcal{N}(0, N\mathbb{I})$  with  $\min(Q, N) > 0$ , where  $\mathbb{I} \in \mathbb{R}^{n \times n}$  denotes the identity matrix. We denote by  $X_1$  the *interim state* and  $Y_1$  the output of the noisy channel, defined by

$$X_1 = X_0 + U_1 \quad \text{with } X_0 \sim \mathcal{N}(0, Q), \quad (\text{V.43})$$

$$Y_1 = X_1 + Z_1 = X_0 + U_1 + Z_1 \quad \text{with } Z_1 \sim \mathcal{N}(0, N). \quad (\text{V.44})$$

We denote by  $\mathcal{P}_{X_0} = \mathcal{N}(0, Q)$  the Gaussian probability distribution of the state random variable  $X_0$ , and we denote by  $\mathcal{P}_{X_1, Y_1 | X_0, U_1}$  the conditional probability distribution corresponding to (V.43) and (V.44).

**Definition 31** For  $n \in \mathbb{N}^*$ , a “control design” with non-causal encoder and causal decoder is a tuple of stochastic functions  $c = (f, \{g_t\}_{t \in \{1, \dots, n\}})$  defined by

$$f : \mathcal{X}_0^n \longrightarrow \mathcal{U}_1^n, \quad g_t : \mathcal{Y}_1^t \longrightarrow \mathcal{U}_2, \quad \forall t \in \{1, \dots, n\}, \quad (\text{V.45})$$

which induces a distribution over the sequences of symbols given by

$$\left( \prod_{t=1}^n \mathcal{P}_{X_{0,t}} \right) f_{U_1^n | X_0^n} \left( \prod_{i=t}^n \mathcal{P}_{X_{1,t}, Y_{1,t} | X_{0,t}, U_{1,t}} \right) \left( \prod_{t=1}^n g_{U_{2,t} | Y_1^t} \right). \quad (\text{V.46})$$

We denote by  $\mathcal{C}_d(n)$  the set of control designs with non-causal encoder and causal decoder.

The counterexample of Witsenhausen (1968) allows to investigate the trade-off between two cost functions, a power cost for the channel input  $U_1$ , and a decoder estimation cost of the interim state  $X_1$ . We evaluate these two costs by considering their respective averages over the sequences of symbols.

**Definition 32** We define the  $n$ -stage costs associated with control design  $c \in \mathcal{C}_d(n)$  by

$$\gamma_p^n(c) = \begin{cases} \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n U_{1,t}^2 \right] & \text{if it exists,} \\ +\infty & \text{otherwise,} \end{cases} \quad (\text{V.47})$$

$$\gamma_s^n(c) = \begin{cases} \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n (X_{1,t} - U_{2,t})^2 \right] & \text{if it exists,} \\ +\infty & \text{otherwise.} \end{cases} \quad (\text{V.48})$$

The pair of costs  $(P, S) \in \mathbb{R}^2$  is achievable if for all  $\varepsilon > 0$ , there exists  $\bar{n} \in \mathbb{N}^*$  such that for all  $n \geq \bar{n}$ , there exists a control design  $c \in \mathcal{C}_d(n)$  such that

$$\left| P - \gamma_p^n(c) \right| + \left| S - \gamma_s^n(c) \right| \leq \varepsilon. \quad (\text{V.49})$$

The goal is to characterize the set of achievable pair of costs  $(P, S) \in \mathbb{R}^2$ , which we call the Witsenhausen costs.

### V.2.b Coding Result

In this section, we extend the coordination coding result of (Le Treust, 2017, Theorem 4) to the case where state and channel noise are Gaussian random variables, other random variables are real-valued, and we consider the power and estimation cost functions of the vector-valued Witsenhausen counterexample. We provide a characterization of the achievable pairs of costs  $(P, S)$ .

**Theorem 26** *The pair of Witsenhausen costs  $(P, S)$  is achievable if and only if there exists a joint probability distribution that decomposes according to*

$$\mathcal{P}_{X_0} \mathcal{Q}_{U_1 W_1 W_2 | X_0} \mathcal{P}_{X_1 Y_1 | X_0 U_1} \mathcal{Q}_{U_2 | W_2 Y_1}, \quad (\text{V.50})$$

such that

$$I(W_1; Y_1 | W_2) - I(W_1, W_2; X_0) \geq 0, \quad (\text{V.51})$$

$$P = \mathbb{E}_{\mathcal{Q}}[U_1^2], \quad S = \mathbb{E}_{\mathcal{Q}}[(X_1 - U_2)^2], \quad (\text{V.52})$$

where  $W_1$  and  $W_2$  are discrete or continuous auxiliary random variables.

The achievability proof of Theorem 26 relies on a block-Markov coding scheme and an adequate notion of weak typicality, inspired from the techniques of Vu et al. (2019). The achievability and converse proofs are stated in (Le Treust and Oechtering, 2022, App. A, [W9]). Entropy and mutual information are defined using the Radon-Nikodym derivative, which is briefly recapitulated in the preliminaries.

**Remark 18** *The probability distributions in (V.50) satisfy the Markov chains*

$$\begin{cases} (X_1, Y_1) \text{---} (X_0, U_1) \text{---} (W_1, W_2), \\ U_2 \text{---} (Y_1, W_2) \text{---} (X_0, X_1, U_1, W_1). \end{cases} \quad (\text{V.53})$$

The causality condition prevents the controller  $C_2$  to recover  $W_1$  which induces the second Markov chain of (V.53). The first and second Markov chains are induced by the network topology.

**Remark 19** *The information constraint (V.51) reformulates as*

$$I(W_1; Y_1, W_2) - I(W_1; X_0, W_2) \geq I(X_0; W_2). \quad (\text{V.54})$$

The terms  $I(X_0; W_2)$  corresponds to the quantization of the state  $X_0$  via the auxiliary random variable  $W_2$ . The expression  $I(W_1; Y_1, W_2) - I(W_1; X_0, W_2)$  stands for the capacity of a two-sided state dependent channel where the encoder observes  $(X_0, W_2)$  and the decoder observes  $W_2$ . Intuitively,  $W_1$  is used to tune the state-dependent channel so as to increase its capacity, as in Gel'fand and Pinsker (1980), in order to refine the quantization of the state  $X_0$ , via the auxiliary random variable  $W_2$ .

In order to investigate the region of achievable pairs of Witsenhausen costs  $(P, S)$ , we focus on its boundary. We fix the power cost to some parameter  $P \geq 0$ , and we investigate the optimal estimation cost at the decoder.

**Definition 33** *Given a power cost parameter  $P \geq 0$ , the optimal estimation cost  $\text{MMSE}(P)$  is the solution of the optimization problem*

$$\text{MMSE}(P) = \inf_{\mathcal{Q} \in \mathcal{Q}(P)} \mathbb{E}_{\mathcal{Q}}[(X_1 - U_2)^2], \quad (\text{V.55})$$

$$\mathcal{Q}(P) = \left\{ (\mathcal{Q}_{U_1 W_1 W_2 | X_0}, \mathcal{Q}_{U_2 | W_2 Y_1}) \text{ s.t. } P = \mathbb{E}_{\mathcal{Q}}[U_1^2], \right. \\ \left. I(W_1; Y_1 | W_2) - I(W_1, W_2; X_0) \geq 0 \right\}. \quad (\text{V.56})$$

The notation  $\text{MMSE}(P)$  in (V.55) recalls that the decoder estimation cost is the Minimum Mean Square Error estimation cost. For such objective, the decoder optimal decision policy is well known and given by the conditional expectation stated in the following lemma.

**Lemma 2** *Given a power cost parameter  $P \geq 0$ , the optimal estimation cost  $\text{MMSE}(P)$  satisfies*

$$\text{MMSE}(P) = \inf_{\mathcal{Q} \in \mathcal{Q}_c(P)} \mathbb{E}_{\mathcal{Q}} \left[ (X_1 - \mathbb{E}[X_1|W_2, Y_1])^2 \middle| W_2, Y_1 \right], \quad (\text{V.57})$$

$$\mathcal{Q}_c(P) = \left\{ \mathcal{Q}_{U_1 W_1 W_2 | X_0} \text{ s.t. } P = E_{\mathcal{Q}}[U_1^2], \right. \\ \left. I(W_1; Y_1 | W_2) - I(W_1, W_2; X_0) \geq 0 \right\}. \quad (\text{V.58})$$

*Proof.* [Lemma 2] For all probability distribution  $\mathcal{Q}_{X_1 W_2 Y_1}$ , the random variable  $U_2 = \mathbb{E}[X_1|W_2, Y_1]$  minimizes  $\mathbb{E}_{\mathcal{Q}}[(X_1 - U_2)^2]$ .  $\square$

Note that the remaining optimization problem (V.57) is difficult to solve since the domain is the set of real-valued distributions. In the next section, we investigate this optimization problem by considering additional assumptions that restrict the set of conditional distributions  $\mathcal{Q}_{U_1 W_1 W_2 | X_0}$ .

### V.2.c Control schemes of particular interest

We restrict our attention to specific choices for the auxiliary random variables  $W_1$  and  $W_2$ . We consider both  $W_1$  and  $W_2$  are Gaussian in Sec. V.2.c, whereas in Sec. V.2.c, the random variable  $W_2$  is discrete and  $W_1$  is Gaussian.

#### Gaussian auxiliary random variables $W_1$ and $W_2$

We focus our attention to the class of jointly Gaussian random variables  $(X_0, U_1, W_1, W_2, X_1, Y_1, U_2)$ .

**Definition 34** *Given a power cost parameter  $P \geq 0$ , we define the optimal estimation cost obtained with Gaussian random variables*

$$\text{MMSE}_G(P) = \inf_{\mathcal{Q} \in \mathcal{Q}_G(P)} \mathbb{E}_{\mathcal{Q}} \left[ (X_1 - \mathbb{E}[X_1|W_2, Y_1])^2 \middle| W_2, Y_1 \right], \quad (\text{V.59})$$

$$\mathcal{Q}_G(P) = \left\{ \mathcal{Q}_{U_1 W_1 W_2 | X_0} \text{ is conditionally Gaussian and } P = E_{\mathcal{Q}}[U_1^2], \right. \\ \left. I(W_1; Y_1 | W_2) - I(W_1, W_2; X_0) \geq 0 \right\}. \quad (\text{V.60})$$

Note that if  $(X_0, U_1, W_2)$  are Gaussian, then  $\mathbb{E}[X_1|W_2, Y_1]$  is also Gaussian.

**Definition 35** *Given a power cost parameter  $P \geq 0$ , we consider the linear scheme defined by*

$$U_{1,\ell}(P) = \begin{cases} -\sqrt{\frac{P}{Q}} X_0 & \text{if } P \in [0, Q], \\ -X_0 + \sqrt{P-Q} & \text{otherwise.} \end{cases} \quad (\text{V.61})$$

The linear estimation cost function given by

$$\text{MMSE}_{\ell}(P) = \begin{cases} \frac{(\sqrt{Q}-\sqrt{P})^2 \cdot N}{(\sqrt{Q}-\sqrt{P})^2 + N} & \text{if } P \in [0, Q], \\ 0 & \text{otherwise.} \end{cases} \quad (\text{V.62})$$

Note that if  $P \geq Q$ , the interim state  $X_1$  can be canceled and the offset  $\sqrt{P-Q}$  is only included to meet the power constraint with equality, as in (V.49). In the linear scheme  $U_{1,\ell}(P)$ , the channel input is used to cancel the state  $X_0$ . The next lemma is a reformulation of (Witsenhausen, 1968, Lemma 11), which shows that  $\text{MMSE}_{\ell}(P)$  is obtained by using the best linear scheme.

**Lemma 3** *We consider the linear strategy*

$$U_1 = a \cdot X_0 + b, \quad (\text{V.63})$$

with parameters  $(a, b) \in \mathbb{R}^2$ , such that to match the power cost constraint  $\mathbb{E}[U_1^2] = a^2Q + b^2 = P$ . The optimal estimation cost is given by

$$\inf_{\substack{(a,b) \in \mathbb{R}^2, \\ a^2Q + b^2 = P}} \mathbb{E} \left[ (X_1 - \mathbb{E}[X_1|Y_1])^2 \middle| Y_1 \right] = \text{MMSE}_\ell(P), \quad (\text{V.64})$$

which is achieved by the strategy  $U_{1,\ell}(P)$  defined in (V.61).

For the sake of clarity, we also provide the proof of Lemma 3, in (Le Treust and Oechtering, 2022, App. B, [W9]).

**Theorem 27** *Suppose that  $Q > 4N$ , we define the parameters*

$$P_1 = \frac{1}{2} \left( Q - 2N - \sqrt{Q \cdot (Q - 4N)} \right), \quad (\text{V.65})$$

$$P_2 = \frac{1}{2} \left( Q - 2N + \sqrt{Q \cdot (Q - 4N)} \right). \quad (\text{V.66})$$

The optimal estimation cost obtained with Gaussian random variables is given by

$$\text{MMSE}_G(P) = \begin{cases} \frac{N \cdot (Q - N - P)}{Q} & \text{if } Q > 4N \text{ and } P \in [P_1, P_2], \\ \text{MMSE}_\ell(P) & \text{otherwise.} \end{cases} \quad (\text{V.67})$$

The proof of Theorem 27 is stated in (Le Treust and Oechtering, 2022, App. C, [W9]). The estimation cost in (V.67) can be obtained by using, either a time sharing strategy between the two linear schemes  $U_{1,\ell}(P_1)$  and  $U_{1,\ell}(P_2)$ , when  $Q > 4N$  and  $P \in [P_1, P_2]$ , and otherwise with the linear scheme  $U_{1,\ell}(P)$ . This result shows that, under the Gaussian assumption, memoryless policies are optimal so that these policies are also optimal for the original scalar Witsenhausen counterexample setup restricted to Gaussian random variables. However, as pointed out by Witsenhausen in Witsenhausen (1968), the Gaussian assumption is a strong restriction in the original scalar model which induces control designs that are generally not optimal.

### Gaussian auxiliary random variable $W_1$ and discrete $W_2$

In this section, we assume that  $P \leq Q$  and we consider that  $W_2$  is a discrete auxiliary random variable, equal to the sign of the interim random variable  $X_1$ ,

$$W_2 = \text{sign}(X_1). \quad (\text{V.68})$$

We assume that the random variables  $(X_0, U_1)$  are centered jointly Gaussian, distributed according to  $\mathcal{N}(0, K)$ , with covariance matrix

$$K = \begin{pmatrix} Q & \rho\sqrt{PQ} \\ \rho\sqrt{PQ} & P \end{pmatrix}, \quad (\text{V.69})$$

depending on the correlation parameter  $\rho \in [-1, 1]$ .

Given a correlation parameter  $\rho \in [-1, 1]$ , we reformulate the pair of correlated Gaussian random variables  $(X_0, U_1)$  into a pair of independent Gaussian random variables  $(\tilde{S}, \tilde{X})$  such that the sum is preserved, i.e.  $X_0 + U_1 = \tilde{S} + \tilde{X}$ , and the auxiliary channel input  $\tilde{X}$  is independent of the auxiliary channel state  $\tilde{S}$ . Since  $(X_0, U_1) \sim \mathcal{N}(0, K)$ , we have  $U_1 = \rho\sqrt{\frac{P}{Q}}X_0 + \tilde{X}$  with  $\tilde{X} \sim \mathcal{N}(0, P(1 - \rho^2))$  and  $\tilde{X} \perp X_0$ .

Therefore, we introduce two Gaussian random variables  $\tilde{X} \sim \mathcal{N}(0, P(1-\rho^2))$  and  $\tilde{S} \sim \mathcal{N}(0, (\sqrt{Q} + \rho\sqrt{P})^2)$  such that

$$\tilde{S} = \frac{\sqrt{Q} + \rho\sqrt{P}}{\sqrt{Q}} \cdot X_0, \quad (\text{V.70})$$

$$\tilde{X} \perp (\tilde{S}, X_0), \quad (\text{V.71})$$

$$X_1 = X_0 + U_1 = \tilde{X} + \tilde{S}. \quad (\text{V.72})$$

Then, the state-dependent channel reformulates

$$Y_1 = X_0 + U_1 + Z = \tilde{X} + \tilde{S} + Z, \quad (\text{V.73})$$

for which Costa's auxiliary random variable for Dirty Paper Coding (DPC), see [Costa \(1983\)](#), writes

$$W_1 = \tilde{X} + \alpha\tilde{S}, \quad \text{with} \quad \alpha = \frac{P(1-\rho^2)}{P(1-\rho^2) + N}. \quad (\text{V.74})$$

By combining [\(V.70\)](#) and [\(V.74\)](#), we reformulate the auxiliary random variable

$$W_1 = \tilde{X} + \frac{P(1-\rho^2)(\sqrt{Q} + \rho\sqrt{P})}{(P(1-\rho^2) + N)\sqrt{Q}} \cdot X_0, \quad \text{where } X_0 \perp \tilde{X} \sim \mathcal{N}(0, P(1-\rho^2)). \quad (\text{V.75})$$

Note that the correlation parameter  $\rho \in [-1, 1]$  is a free parameter that we use to minimize the decoder estimation cost.

**Definition 36** Given  $P \geq 0$ , we consider the auxiliary random variables  $(W_1, W_2)$  defined by [\(V.75\)](#) and [\(V.68\)](#), the optimal estimation cost is defined by

$$\text{MMSE}_{\text{coord}}(P) = \min_{\rho \in [-1, 1]} \mathbb{E} \left[ \left( X_1 - \mathbb{E}[X_1 | W_2, Y_1] \right)^2 \middle| W_2, Y_1 \right], \quad (\text{V.76})$$

$$\text{s.t.} \quad I(W_1; Y_1 | W_2) - I(W_1; X_0 | W_2) \geq I(X_0; W_2). \quad (\text{V.77})$$

In the information constraint [\(V.77\)](#), the quantization rate  $I(X_0; W_2)$  must be smaller than the state dependent channel capacity  $I(W_1; Y_1 | W_2) - I(W_1; X_0 | W_2)$ .

**Proposition 7** Given  $P \geq 0$ , we consider the auxiliary random variables  $(W_1, W_2)$  defined by [\(V.75\)](#) and [\(V.68\)](#) and we use the change of variable  $T = P + Q + 2\rho\sqrt{PQ}$ . We have

$$\text{MMSE}_{\text{coord}}(P) = \min_{\rho \in [-1, 1]} \frac{TN}{T+N} \cdot \left( 1 - \frac{2}{\sqrt{T+N}} \cdot \frac{1}{2\pi} \int \frac{\phi\left(y_1 \cdot \sqrt{\frac{2T+N}{N(T+N)}}\right)}{\Phi\left(y_1 \cdot \sqrt{\frac{T}{N(T+N)}}\right)} dy_1 \right), \quad (\text{V.78})$$

$$\begin{aligned} \text{s.t.} \quad & \frac{1}{2} \log_2 \left( 1 + \frac{P(1-\rho^2)}{N} \right) - \Psi \left( \sqrt{\frac{T}{N}} \right) \\ & + \Psi \left( \sqrt{\frac{T(\sqrt{Q} + \rho\sqrt{P})^2 N + P(1-\rho^2)(T+N)^2}{(\sqrt{Q} + \rho\sqrt{P})^2 N^2}} \right) \geq 1, \end{aligned} \quad (\text{V.79})$$

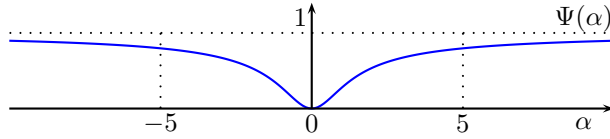
where the entropy reduction function  $\Psi : \mathbb{R} \rightarrow [0, 1]$  is defined by

$$\Psi(\alpha) = \int 2\Phi(\alpha \cdot x) \cdot \log_2 \left( 2\Phi(\alpha \cdot x) \right) \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) dx, \quad (\text{V.80})$$

and  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left( -\frac{t^2}{2} \right) dt$ , is the Gaussian cumulative distribution function.

The proof of Prop. 7 is stated in ([Le Treust and Oechtering, 2022](#), App. D, [W9]). The first term of [\(V.79\)](#) corresponds to the capacity of a Gaussian state-dependent channel with channel input power  $P(1-\rho^2)$  and noise variance  $N$ . The entropy reduction function  $\Psi(\alpha)$  corresponds to the entropy penalty term of skew normal distribution with the skewness factor  $\alpha \in \mathbb{R}$ .




 Figure V.7 – Entropy reduction function  $\Psi(\alpha)$  defined in (V.80).

## V.2.d Numerical Results

We compare the performances of the control schemes of Sec. V.2.c and Sec. V.2.c, to the Witsenhausen two-point strategy (Witsenhausen, 1968, Sec. 5). In Grover and Sahai (2010), Grover and Sahai investigate a vector-version of the Witsenhausen counterexample in which the decoder is non-causal and implements the DPC scheme of Costa (1983), for a specific channel state. We compare the control schemes of Sec. V.2.c and Sec. V.2.c to the DPC based scheme of Grover and Sahai (2010).

### Witsenhausen two-point strategy

**Proposition 8 (Two-point strategy)** *For some parameter  $a \geq 0$ , Witsenhausen two-point strategy is defined by*

$$U_1 = a \cdot \text{sign}(X_0) - X_0. \quad (\text{V.81})$$

The power and estimation costs are given by

$$P_{\text{two}}(a) = Q + a \left( a - 2\sqrt{\frac{2Q}{\pi}} \right), \quad (\text{V.82})$$

$$\text{MMSE}_{\text{two}}(a) = \sqrt{\frac{2\pi}{N}} a^2 \phi\left(\frac{a}{\sqrt{N}}\right) \int \frac{\phi\left(\frac{y}{\sqrt{N}}\right)}{\cosh\left(\frac{ay}{N}\right)} dy, \quad (\text{V.83})$$

where  $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$  and the optimal receiver's strategy is given by  $\mathbb{E}[X_1|Y = y] = a \tanh\left(\frac{ay}{N}\right)$ .

For the sake of clarity, we also provide the proof of Proposition 8, in (Le Treust and Oechtering, 2022, App. E, [W9]). By letting  $N = 1$  and  $a = \sqrt{2Q}$ , we recover the equations in the proof of (Witsenhausen, 1968, Theorem 2). Note that the function  $P_{\text{two}}(a)$  decreases over the interval  $\left[0, \sqrt{\frac{2Q}{\pi}}\right]$ , where it reaches the minimal value  $Q\left(1 - \frac{2}{\pi}\right)$ , and then increases for  $a \geq \sqrt{\frac{2Q}{\pi}}$ . Note that this two-point strategy requires a power cost  $P \geq Q\left(1 - \frac{2}{\pi}\right)$  in order to be implemented. This strategy induces a binary interim state  $X_1 = a \cdot \text{sign}(X_0) \in \{-a, a\}$  for which the estimation cost outperforms, in some cases, the best estimation cost obtained via the linear scheme, see (Witsenhausen, 1968, Theorem 2).

### Dirty Paper Coding (DPC) based scheme for non-causal decoder

In Grover and Sahai (2010), the authors investigate a vector version of Witsenhausen counterexample in which the decoder is non-causal.

**Definition 37** *For  $n \in \mathbb{N}^*$ , a “control design” with non-causal encoder and non-causal decoder is a tuple of stochastic functions  $c = (f, g)$  defined by*

$$f : \mathcal{X}_0^n \longrightarrow \mathcal{U}_1^n, \quad g : \mathcal{Y}_1^n \longrightarrow \mathcal{U}_2^n. \quad (\text{V.84})$$

We denote by  $\mathcal{C}_{\text{nc}}(n)$  the set of control designs with non-causal encoder and non-causal decoder.

In (Grover and Sahai, 2010, App. D.1-D.7), the authors investigate a Dirty Paper Coding (DPC) based scheme by using a Gaussian channel input  $U_1 \sim \mathcal{N}(0, P)$ ,  $U_1 \perp X_0$  and the auxiliary random

variable  $W = U_1 + \alpha X_0$ . This leads to

$$I(W; Y_1) - I(W; X_0) = \frac{1}{2} \log_2 \left( \frac{P(P+Q+N)}{PQ(1-\alpha)^2 + N(P+\alpha^2Q)} \right), \quad (\text{V.85})$$

$$H(U_1 + X_0 | W, Y_1) = \frac{1}{2} \log_2 \left( (2\pi e) \cdot \frac{NPQ(1-\alpha)^2}{PQ(1-\alpha)^2 + N(P+\alpha^2Q)} \right). \quad (\text{V.86})$$

Since the random variables are jointly Gaussian, the optimization problem writes

$$\text{MMSE}_{\text{dpc}}(P) = \min_{\substack{\alpha \in \mathbb{R}, \\ P(P+Q+N) \geq PQ(1-\alpha)^2 + N(P+\alpha^2Q)}} \frac{NPQ(1-\alpha)^2}{PQ(1-\alpha)^2 + N(P+\alpha^2Q)}. \quad (\text{V.87})$$

**Proposition 9** *Let  $P^* \geq 0$ , the unique positive root of equation  $P^2(P+Q+N) = QN^2$ .*

- If  $P \leq P^*$ , the estimation cost for DPC is given by

$$\text{MMSE}_{\text{dpc}}(P) = \frac{N(N\sqrt{Q} - P\sqrt{P+Q+N})^2}{(P+N)^2(P+Q+N)}, \quad (\text{V.88})$$

which is achieved with  $\alpha^* = \frac{P(\sqrt{Q} + \sqrt{P+Q+N})}{\sqrt{Q}(P+N)}$ .

- If  $P > P^*$ , then  $\text{MMSE}_{\text{dpc}}(P) = 0$  which is achieved with  $\alpha^* = 1$ .

This result is proved in (Grover and Sahai, 2010, App. D.1-D.7), we recall the main proof arguments in (Le Treust and Oechtering, 2022, App. F, [W9]). In (Grover and Sahai, 2010, App. D.8), the authors additionally investigate a combination between the linear scheme and the DPC scheme. Given a parameter  $-\sqrt{\frac{P}{Q}} \leq \beta \leq \sqrt{\frac{P}{Q}}$ , the transmit power  $P$  is divided into a linear part  $U_{1,1} = -\beta X_0$  and a part  $U_{1,2}$  used to implement DPC against the state  $(1-\beta)X_0 \sim \mathcal{N}(0, (1-\beta)^2Q)$  with power constraint  $\mathbb{E}[U_{1,2}^2] \leq P - \beta^2Q$ . By using the change of variable  $\beta = -\rho\sqrt{\frac{P}{Q}}$ , we obtain the correlation matrix of (V.69) and the auxiliary state-dependent channel  $Y_1 = \tilde{X} + \tilde{S} + Z$  where  $\tilde{X} + \tilde{S} = X_0 + U_1$ , the channel state  $\tilde{S}$  is defined in (V.70), and  $\tilde{X} \perp (\tilde{S}, X_0)$  with  $\mathbb{E}[\tilde{X}^2] \leq P(1-\rho^2)$ . Therefore, we replace  $P$  and  $Q$  in (V.88), respectively by  $P(1-\rho^2)$  and  $(\sqrt{Q} + \rho\sqrt{P})^2$ , and we obtain

$$\text{MMSE}_{\text{lin+dpc}}(P) = \min_{\rho \in [-1,1]} \frac{N(P(1-\rho^2)\sqrt{P+Q+2\rho\sqrt{PQ}} + N - N(\sqrt{Q} + \rho\sqrt{P}))^2}{(P(1-\rho^2) + N)^2(P+Q+2\rho\sqrt{PQ} + N)}. \quad (\text{V.89})$$

In the next section, we will see that this estimation cost  $\text{MMSE}_{\text{lin+dpc}}(P)$  outperforms all other estimation costs.

## V.2.e Discussion

In Figure V.8, we compare the estimation cost proposed in Sec. V.2.c and Sec. V.2.c, with the estimation costs from the literature, for  $(Q, N) = (0.1, 0.01)$ .

- The blue curve corresponds to the estimation cost of the best linear scheme  $\text{MMSE}_\ell(P)$  defined in (V.62), see also (Witsenhausen, 1968, Lemma 11).
- The green curve depicts the estimation cost of Witsenhausen two point strategy  $(P_{\text{two}}, \text{MMSE}_{\text{two}})$  defined in (V.82) and (V.83), see also (Witsenhausen, 1968, Sec. 5).
- The red curve presents the estimation cost of the Grover and Sahai's combination of the linear scheme and DPC scheme  $\text{MMSE}_{\text{lin+dpc}}(P)$  defined in (V.89), when the decoder is non-causal, see also (Grover and Sahai, 2010, App. D.1-D.8).

The coordination coding scheme we propose in Sec. V.2.c is restricted to Gaussian random variables. The estimation cost  $\text{MMSE}_G(P)$  defined in (V.67) consists of the brown line when  $Q > 4N$  and  $P \in$

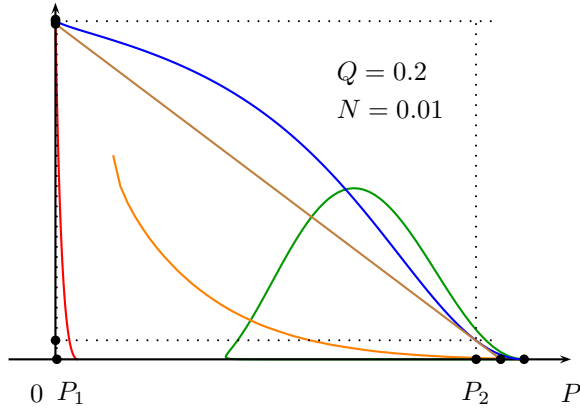


Figure V.8 – Blue: optimal linear scheme  $\text{MMSE}_\ell(P)$  in (V.62), Brown: time sharing  $\frac{N \cdot (Q - N - P)}{Q}$  in (V.67), Green: two-point strategy  $(P_w, \text{MMSE}_w)$  in (V.82) and (V.83) ; Red: combination DPC and linear scheme  $\text{MMSE}_{\text{lin+dpc}}(P)$  in (V.89) ; Orange: MMSE when  $W_2 = \text{sign}(X_1)$  and  $W_1$  is Costa’s auxiliary random variable  $\text{MMSE}_{\text{coord}}(P)$  in (V.78).

$[P_1, P_2]$ , and of the estimation cost  $\text{MMSE}_\ell(P)$  represented by the blue line, otherwise. Note that the function  $P \mapsto \text{MMSE}_G(P)$  is the convexification of the linear estimation cost function  $P \mapsto \text{MMSE}_\ell(P)$ .

We reduce the estimation cost by using the auxiliary random variable  $W_2 = \text{sign}(X_1)$  that encapsulates the sign of the interim state, see Sec. V.2.c. This strategy requires a certain power cost level for the first controller to transmit the sign of  $X_1$  to the second controller. The Fig. V.8, shows the existence of some weight parameter  $\kappa \in [0, 1]$ , such that our coordination coding scheme of Sec. V.2.c, outperforms Witsenhausen two point strategy

$$\kappa P + (1 - \kappa)\text{MMSE}_{\text{coord}}(P) \leq \kappa P_{\text{two}} + (1 - \kappa)\text{MMSE}_{\text{two}}. \quad (\text{V.90})$$

Note also that the power cost required to implement our coordination coding scheme of Sec. V.2.c, is strictly less than the minimal power cost  $Q(1 - \frac{2}{\pi})$  needed to implement the Witsenhausen two point strategy.

When the decoder is non-causal, the combination between the linear scheme and the DPC scheme  $\text{MMSE}_{\text{lin+dpc}}(P)$  proposed in Grover and Sahai (2010), Pareto-dominates all the other solutions.

### V.3 Strong coordination of signals and actions

A general information-theoretic framework to study coordination in networks was put forward in Cuff et al. (2010), related to earlier work on “Shannon’s reverse coding theorem” Bennett et al. (2002) and the compression of probability distribution sources and mixed quantum states Kramer and Savari (2007); Soljanin (2002); Winter (2002). This framework also relates to the game-theoretic perspective on coordination Gossner et al. (2006) with applications, for instance, to power control Larrousse et al. (2018). Recent extensions of the framework have included the possibility of coordination through interactive communication Gülcü and Barg (2016); Haddadpour et al. (2017); Yassaee et al. (2015). Two information-theoretic metrics have been proposed to measure the level of coordination: *empirical coordination*, introduced in Chap. IV, which requires the joint histogram of the devices’ actions to approach a target distribution, and *strong coordination*, which requires the joint distribution of sequences of actions to converge to an i.i.d. target distribution, e.g., in  $\ell_1$  distance Cuff (2013); Cuff et al. (2010). Empirical coordination captures an “average behavior” over multiple repeated actions of the devices, in contrast, strong coordination captures the behavior of sequences. A byproduct of strong coordination is that it enforces some level of “security,” in the sense of guaranteeing that sequence of actions will be *unpredictable* to an outside observer beyond what is known about the target joint distribution of sequences.

Strong coordination in networks was first studied over error free links in [Cuff et al. \(2010\)](#) and later extended to noisy communication links in [Haddadpour et al. \(2017\)](#). In the latter setting, the signals that are transmitted and received over the physical channel become a part of what can be observed, and one can therefore coordinate the actions of the devices with their communication signals [Cuff and Schieler \(2011\)](#), ([Le Treust, 2017](#), [J6]).

To illustrate the relevance of this joint coordination problem, we can consider the situation in which two stockbrokers who work for two competing companies decide to collude. One stockbroker, Agent 1, has access to a source of information on future market developments: he sees the symbol 0 when it is profitable to buy and the symbol 1 when it is profitable to sell. Agent 1 can accordingly act to buy or sell, which is represented by 0 or 1, respectively. The other stockbroker, Agent 2, takes similar actions but is less informed and only sees the actions of Agent 1. Both agents are assumed to benefit from helping each other but their cooperation would be illegal and any coordination should therefore be kept secret. Secrecy is achieved by keeping the signal distributions to be statistically indistinguishable from i.i.d., so that an outside observer working for a regulating entity would not be able to prove that an exchange of information takes place.

As the stockbrokers example above suggests, the joint strong coordination of signals and actions with a noisy link is particularly interesting if security is required: if for example we require the actions of the agents to appear independent of the communication, a malicious eavesdropper who observes the output of the channel cannot infer anything about the source and the reconstruction without having access to the source of common randomness. This “secure coordination” was investigated for error-free links in [Satpathy and Cuff \(2016\)](#).

In this section, we address the problem of strong coordination in a two-node network comprised of an information source and a noisy channel, in which both nodes have access to a common source of randomness. This scenario presents two conflicting goals: the encoder needs to convey a message to the decoder in order to coordinate the actions, while simultaneously coordinating the signals that encode the message.

Up to now, the coordination problem assumes that the source and the channel follow distributions which are fixed ahead of time and known by the agents. However, this constraint prevents us from modeling situations in which the agent reacts to an external stimulus, and in which the channel statistics depend on the environment. For instance, consider a situation where the actions of an agent might be constrained by obstacles that prevent it from making certain choices. In this case the probability distributions given by nature could change with time and would be partially if not completely unknown to some of the agents. To include such situations in the coordination framework, we introduce a random state capturing the effect of the environment, to model actions and channels that change with external factors, and we consider a general setting in which we have a state-dependent channel and a source with decoder’s side information as in ([Le Treust, 2014](#), [IC12]), ([Le Treust, 2015c](#), [IC13]), [Larrousse et al. \(2015a\)](#).

We derive an inner and an outer bound for the strong coordination region by developing a joint source-channel scheme in which an auxiliary codebook allows us to satisfy both coordination of signals and actions. Since the two bounds do not match, the optimality of our general achievability scheme remains an open question. However, we succeed in characterizing the strong coordination region in some special cases: i) when the channel is noiseless, ii) when the decoder is lossless, and iii) when the random variables of the channel are independent from the random variables of the source. In all these cases, the sets of achievable target distributions are the same as for empirical coordination ([Le Treust, 2015c](#), [IC13]), but we show that a positive rate of common randomness is required for strong coordination.

The remainder of this section is organized as follows. Section [V.3.a](#) describes a simple model in which there is no state and no side information and derives an inner and an outer bound for the strong coordination region. The information-theoretic modeling of coordination problems relevant to this work is best illustrated in this simplified scenario. Section [V.3.b](#) extends the inner and outer bounds to the general case of a noisy channel with state and side information at the decoder. Section [V.3.c](#) characterizes the strong coordination region for three special cases and shows that the separation principle does not hold for strong coordination.

### V.3.a Inner and outer bounds for the strong coordination region

Before we consider the general model with a state and side information, it is helpful to focus on the simpler model depicted in Figure V.9 to understand the nature of the problem. Two agents, the encoder and the decoder, wish to coordinate their behaviors: the stochastic actions of the agents should follow a known and fixed joint distribution.

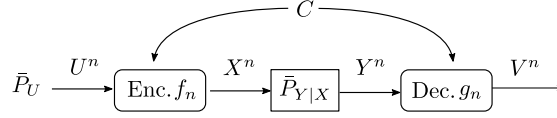


Figure V.9 – Coordination of signals and actions for a two-node network with a noisy channel with non-causal encoder and decoder.

Given a rate parameter  $R_0 \geq 0$ , we suppose that the encoder and the decoder have access to a shared source of uniform randomness  $C \in \llbracket 1, 2^{nR_0} \rrbracket$ . Let  $U^n \in \mathcal{U}^n$  be an i.i.d. source with distribution  $\bar{P}_U$ . The encoder observes the sequence  $U^n \in \mathcal{U}^n$  and selects a signal  $X^n = f_n(U^n, C)$ , where  $f_n : \mathcal{U}^n \times \llbracket 1, 2^{nR_0} \rrbracket \rightarrow \mathcal{X}^n$  is a stochastic map. The signal  $X^n$  is transmitted over a discrete memoryless channel parametrized by the conditional distribution  $\bar{P}_{Y|X}$ . Upon observing  $Y^n$  and common randomness  $C$ , the decoder selects an action  $V^n = g_n(Y^n, C)$ , where  $g_n : \mathcal{Y}^n \times \llbracket 1, 2^{nR_0} \rrbracket \rightarrow \mathcal{V}^n$  is a stochastic map. For block length  $n$ , the pair  $(f_n, g_n)$  constitutes a code.

We recall the definitions of empirical and strong coordination [Cuff \(2009\)](#); [Cuff et al. \(2010\)](#).

**Definition 38** A distribution  $\bar{P}_{UXYV}$  is achievable for empirical coordination if there exists a sequence  $\{(f_n, g_n)\}_{n \geq 1}$  of encoders-decoders such that

$$\forall \delta > 0, \quad \lim_{n \rightarrow +\infty} \min_{(f_n, g_n)_{n \geq 1}} \Pr \left( \left\| Q_{UXYV}^n - \bar{P}_{UXYV} \right\|_1 > \delta \right) = 0, \quad (\text{V.91})$$

where  $Q_{UXYV}^n$  is the empirical distribution of the random variables  $(U^n, X^n, Y^n, V^n)$  induced by the code. The empirical coordination region  $\mathcal{R}_e$  is the set of achievable distributions  $\bar{P}_{UXYV}$ .

**Definition 39** A pair  $(\bar{P}_{UXYV}, R_0)$  is achievable for strong coordination if there exists a sequence  $\{(f_n, g_n)\}_{n \geq 1}$  of encoders-decoders with rate of common randomness  $R_0$ , such that

$$\lim_{n \rightarrow \infty} \left\| \mathcal{P}_{U^n X^n Y^n V^n}^c - \bar{P}_{UXYV}^{\otimes n} \right\|_1 = 0$$

where  $\mathcal{P}_{U^n X^n Y^n V^n}^c \in \Delta(\mathcal{U}^n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{V}^n)$  is the joint distribution induced by the code, and  $\bar{P}_{UXYV}^{\otimes n}$  denotes the i.i.d. probability distribution. The strong coordination region  $\mathcal{R}$  is the closure of the set of achievable pairs  $(\bar{P}_{UXYV}, R_0)$ .

Our first result is an inner and outer bound for the strong coordination region  $\mathcal{R}$  ([Cervia et al., 2017b](#), [IC19]).

**Theorem 28** Let  $\bar{P}_U$  and  $\bar{P}_{Y|X}$  be the given source and channel parameters, then  $\mathcal{R}'_{in} \subseteq \mathcal{R} \subseteq \mathcal{R}'_{out}$  where:

$$\mathcal{R}'_{in} := \left\{ \begin{array}{l} (\bar{P}_{UXYV}, R_0) \text{ such that:} \\ \bar{P}_{UXYV} = \bar{P}_U \bar{P}_{X|U} \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXYWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W; U) \leq I(W; Y) \\ R_0 \geq I(W; UXV|Y) \end{array} \right\} \quad (\text{V.92})$$

$$\mathcal{R}'_{out} := \left\{ \begin{array}{l} (\bar{P}_{UXYV}, R_0) \text{ such that:} \\ \bar{P}_{UXYV} = \bar{P}_U \bar{P}_{X|U} \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXYWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W; U) \leq I(X; Y) \\ R_0 \geq I(W; UXV|Y) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 4 \end{array} \right\}. \quad (\text{V.93})$$

The proof of Theorem 28 is provided in (Cervia et al., 2020, [J9]).

**Remark 20** Observe that the decomposition of the joint distributions  $\bar{P}_{UXYV}$  and  $\bar{P}_{UWXYV}$  is equivalently characterized in terms of Markov chains:

$$Y - X - U, \quad \left\{ \begin{array}{l} Y - X - (U, W) \\ V - (Y, W) - (X, U) \end{array} \right. . \quad (\text{V.94})$$

**Remark 21** The empirical coordination region for the setting of Figure V.9 was investigated in Cuff and Schieler (2011), in which the authors derived an inner and outer bound. Note that the information constraint  $I(W; U) \leq I(W; Y)$  and the decomposition of the joint probability distribution  $\bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY}$  are the same for empirical coordination (Cuff and Schieler, 2011, Theorem 1). The main difference is that strong coordination requires a positive rate of common randomness  $R_0 > I(W; UXV|Y)$ .

### V.3.b Strong coordination region with state and side information

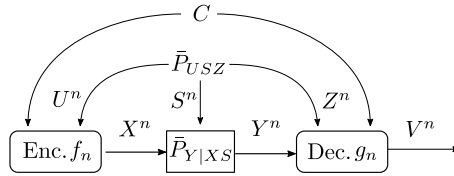


Figure V.10 – Coordination of signals and actions for a two-node network with a noisy channel with state and side information at the decoder.

The problem of source coding with side information at the receiver was notably considered in Wyner and Ziv (1976), while Gel'fand and Pinsker (1980) studies channel coding with non-causal state information available at the encoder. In practice, it is realistic to assume that the encoder and the decoder have access to some side information, without having full knowledge of it as in Cover and Chiang (2002); Merhav and Shamai (2003).

In this paper, we study a joint source-channel strong coordination system whose encoder has the channel state information available and whose decoder has access to side information that may be correlated to the source.

More precisely, in this section we consider the model depicted in Figure V.10, where the noisy channel depends on a state  $S^n$ , and the decoder has access to non-causal side information  $Z^n$ .

The encoder selects a signal  $X^n = f_n(U^n, C)$ , with  $f_n : \mathcal{U}^n \times \llbracket 1, 2^{nR_0} \rrbracket \rightarrow \mathcal{X}^n$  a stochastic map, and transmits it over the discrete memoryless channel  $\bar{P}_{Y|XS}$  where  $S$  represents the state. The decoder then selects an action  $V^n = g_n(Y^n, Z^n, C)$ , where  $g_n : \mathcal{Y}^n \times \mathcal{Z}^n \times \llbracket 1, 2^{nR_0} \rrbracket \rightarrow \mathcal{V}^n$  is a stochastic map and  $Z^n$  represents the side information available at the decoder.

**Remark 22** The channel state information and side information at the decoder are represented explicitly by the random variables  $S^n$  and  $Z^n$  respectively, and we make no assumptions on the correlation of  $(U^n, S^n, Z^n)$ . This includes scenarios where the encoder has access to partial, perfect or noisy channel

state information, since the variables  $U^n$  and  $S^n$  are possibly correlated. Moreover, the decoder side information  $Z^n$  can contain partial, perfect or noisy information on the channel state, on the source, or on both of them.

We recall the notions of empirical and strong coordination in this setting [Cuff \(2009\)](#); [Cuff et al. \(2010\)](#).

**Definition 40** A distribution  $\bar{P}_{USZXYV}$  is achievable for empirical coordination if there exists a sequence  $\{(f_n, g_n)\}_{n \geq 1}$  of encoders-decoders such that

$$\forall \delta > 0, \quad \lim_{n \rightarrow +\infty} \min_{\{(f_n, g_n)\}_{n \geq 1}} \Pr \left( \left\| Q_{USZXYV}^n - \bar{P}_{USZXYV} \right\|_1 > \delta \right) = 0, \quad (\text{V.95})$$

where  $Q_{USZXYV}^n$  is the empirical distribution of the random variables  $(U^n, X^n, Y^n, V^n)$  induced by the code. The empirical coordination region  $\mathcal{R}_e$  is the closure of the set of achievable distributions  $\bar{P}_{USZXYV}$ . A pair  $(\bar{P}_{USZXYV}, R_0)$  is achievable for strong coordination if there exists a sequence  $\{(f_n, g_n)\}_{n \geq 1}$  of encoders-decoders with rate of common randomness  $R_0$ , such that

$$\lim_{n \rightarrow \infty} \left\| P_{U^n S^n Z^n X^n Y^n V^n} - \bar{P}_{USZXYV}^{\otimes n} \right\|_1 = 0$$

where  $P_{U^n S^n Z^n X^n Y^n V^n}$  is the joint distribution induced by the code. The strong coordination region  $\mathcal{R}$  is the closure of the set of achievable pairs  $(\bar{P}_{USZXYV}, R_0)$ .

In the case of non-causal encoder and decoder, the problem of characterizing the strong coordination region for the system model in [Figure V.10](#) is still open, but we establish the following inner and outer bounds.

**Theorem 29** Let  $\bar{P}_{USZ}$  and  $\bar{P}_{Y|XS}$  be the given source and channel parameters, then  $\mathcal{R}_{in} \subseteq \mathcal{R} \subseteq \mathcal{R}_{out}$  where:

$$\mathcal{R}_{in} := \left\{ \begin{array}{l} (\bar{P}_{USZXYV}, R_0) \text{ such that:} \\ \bar{P}_{USZXYV} = \bar{P}_{USZ} \bar{P}_{X|U} \bar{P}_{Y|XS} \bar{P}_{V|UXYSZ} \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{USZWXYV} = \bar{P}_{USZ} \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|XS} \bar{P}_{V|WYZ} \\ I(W; U) \leq I(W; YZ) \\ R_0 \geq I(W; USXV|YZ) \end{array} \right\}, \quad (\text{V.96})$$

$$\mathcal{R}_{out} := \left\{ \begin{array}{l} (\bar{P}_{USZXYV}, R_0) \text{ such that:} \\ \bar{P}_{USZXYV} = \bar{P}_{USZ} \bar{P}_{X|U} \bar{P}_{Y|XS} \bar{P}_{V|UXYSZ} \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{USZWXYV} = \bar{P}_{USZ} \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|XS} \bar{P}_{V|WYZ} \\ I(W; U) \leq \min\{I(XUS; YZ), \\ I(XS; Y) + I(U; Z)\} \\ R_0 \geq I(W; USXV|YZ) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 5 \end{array} \right\}. \quad (\text{V.97})$$

The proof of [Theorem 29](#) is stated in [\(Cervia et al., 2020, \[J9\]\)](#).

**Remark 23** As in [Theorem 28](#), even if inner and outer bound do not match, they only differ on the upper bound on  $I(W; U)$ . Note that we cannot compare  $I(XUS; YZ)$  and  $I(XS; Y) + I(U; Z)$ . Hence, in  $\mathcal{R}_{out}$  the upper bound on the mutual information  $I(W; U)$  is the minimum of the two.

**Remark 24** Observe that the decomposition of the joint distributions  $\bar{P}_{USZXYV}$  and  $\bar{P}_{USZWXYV}$  is equivalently characterized in terms of Markov chains:

$$\left\{ \begin{array}{l} Z - (U, S) - (X, Y) \\ Y - (X, S) - U \end{array} \right. , \quad \left\{ \begin{array}{l} Z - (U, S) - (X, Y, W) \\ Y - (X, S) - (U, W) \\ V - (Y, Z, W) - (X, S, U) \end{array} \right. . \quad (\text{V.98})$$

### V.3.c Characterization of the strong coordination region for special cases

Although the inner and outer bounds in Theorem 29 do not match in general, we characterize the strong coordination region exactly in three special cases: perfect channel, lossless decoding and separation between the channel and the source.

The empirical coordination region for these three settings was derived in (Le Treust, 2015c, [IC13]), (Le Treust, 2015, [W1]). In this section we recover the same information constraints, but we show that for strong coordination a positive rate of common randomness is also necessary. This reinforces the conjecture, stated in Cuff et al. (2010), that with enough common randomness the strong coordination capacity region is the same as the empirical coordination capacity region for any specific network setting.

#### Perfect channel

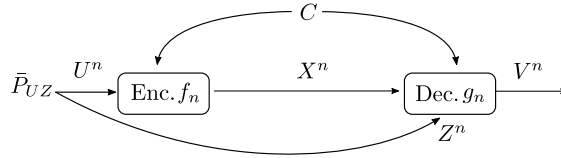


Figure V.11 – Coordination of signals and actions for a two-node network with a perfect channel.

Suppose we have a perfect channel as in Figure V.11:  $\bar{P}_{Y|XS}(\mathbf{y}|\mathbf{x}, \mathbf{s}) = \mathbb{1}_{X=Y}\{\mathbf{x} = \mathbf{y}\}$ . In this case  $X^n = Y^n$  and  $Z^n$  plays the role of side information at the decoder. We characterize the strong coordination region  $\mathcal{R}_{PC}$ .

**Theorem 30** In the setting of Theorem 29, suppose that  $\bar{P}_{Y|XS}(\mathbf{y}|\mathbf{x}, \mathbf{s}) = \mathbb{1}_{X=Y}\{\mathbf{x} = \mathbf{y}\}$ . Then the strong coordination region is

$$\mathcal{R}_{PC} := \left\{ \begin{array}{l} (\bar{P}_{UZXV}, R_0) \text{ such that:} \\ \bar{P}_{UZXV} = \bar{P}_{UZ}\bar{P}_{X|U}\bar{P}_{V|UXZ} \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UZXV} = \bar{P}_{UZ}\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{V|WXZ} \\ I(WX; U) \leq H(X) + I(W; Z|X) \\ R_0 \geq I(W; UV|XZ) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 4 \end{array} \right\} \quad (\text{V.99})$$

The proof of Theorem 30 is stated in (Cervia et al., 2020, [J9]).

**Remark 25** Observe that the decomposition of the joint distributions  $\bar{P}_{UZXV}$  and  $\bar{P}_{UZXV}$  is equivalently characterized in terms of Markov chains:

$$Z - U - X, \quad \left\{ \begin{array}{l} Z - U - (X, W) \\ V - (X, Z, W) - U \end{array} \right. . \quad (\text{V.100})$$



**Case  $H(Y|X) = 0$**  Suppose that, instead of assuming that there is a perfect channel between the encoder and the decoder, we consider the more general case  $H(Y|X) = 0$ . Then, we recover a very similar result to  $\mathcal{R}_{PC}$ .

**Proposition 10** Consider the setting of Theorem 29 and suppose that  $H(Y|X) = 0$ . Then the strong coordination region is

$$\mathcal{R}_{H(Y|X)} := \left\{ \begin{array}{l} (\bar{P}_{USZXYV}, R_0) \text{ such that:} \\ \bar{P}_{USZXYV} = \bar{P}_{USZ} \bar{P}_{X|U} \bar{P}_{Y|XS} \bar{P}_{V|USZXY} \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{USZWXYZV} = \bar{P}_{USZ} \bar{P}_{W|U} \bar{P}_{X|UW} \\ \quad \bar{P}_{Y|XS} \bar{P}_{V|WZY} \\ I(WX; U) \leq H(X) + I(W; Z|X) \\ R_0 \geq I(W; UV|XZ) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 4 \end{array} \right.$$

**Lossless decoding**

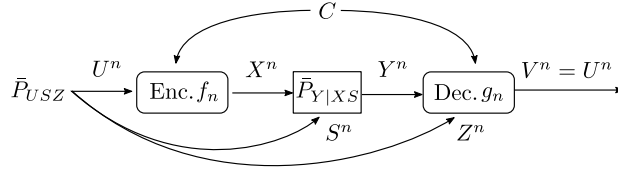


Figure V.12 – Coordination of signals and actions for a two-node network with a noisy channel and a lossless decoder.

Suppose that the decoder wants to reconstruct the source losslessly, i.e.,  $V = U$  as in Figure V.12. Then, we characterize the strong coordination region  $\mathcal{R}_{LD}$ .

**Theorem 31** Consider the setting of Theorem 29 and suppose that  $\bar{P}_{V|USXYZ}(\mathbf{v}|\mathbf{u}, \mathbf{s}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbb{1}_{V=U}\{\mathbf{u} = \mathbf{v}\}$ . Then the strong coordination region is

$$\mathcal{R}_{LD} := \left\{ \begin{array}{l} (\bar{P}_{USZXY}, R_0) \text{ such that:} \\ \bar{P}_{USZXYV} = \bar{P}_{USZ} \bar{P}_{X|U} \bar{P}_{Y|XS} \mathbb{1}_{V=U} \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{USZWXYZV} = \bar{P}_{USZ} \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|XS} \mathbb{1}_{V=U} \\ I(W; U) \leq I(W; YZ) \\ R_0 \geq I(W; USX|YZ) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y}| + 3 \end{array} \right. \quad (\text{V.101})$$

The proof of Theorem 31 is stated in (Cervia et al., 2020, [J9]).

**Remark 26** Observe that the decomposition of the joint distributions  $\bar{P}_{USZXYV}$  and  $\bar{P}_{USZWXYZV}$  is equivalently characterized in terms of Markov chains:

$$\left\{ \begin{array}{l} Z - (U, S) - (X, Y) \\ Y - (X, S) - U \end{array} \right\}, \quad \left\{ \begin{array}{l} Z - (U, S) - (X, Y, W) \\ Y - (X, S) - (U, W) \end{array} \right\}. \quad (\text{V.102})$$

**Remark 27** An equivalent characterization of the region is:

$$\mathcal{R}_{LD} := \left\{ \begin{array}{l} (\bar{P}_{USZXY}, R_0) \text{ such that:} \\ \bar{P}_{USZXY} = \bar{P}_{USZ} \bar{P}_{X|U} \bar{P}_{Y|XS} \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{USZWXY} = \bar{P}_{USZ} \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|XS} \\ H(U) \leq I(WU; YZ) \\ R_0 \geq I(W; USX|YZ) + H(U|WYZ) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y}| + 1 \end{array} \right\} \quad (\text{V.103})$$

The region in (V.103) is achievable since with the choice of the auxiliary random variable  $W'' = (W, U)$ , the constraints in (V.101) become

$$I(WU; U) = H(U) \leq I(WU; YZ) \quad (\text{V.104})$$

$$\begin{aligned} R_0 &\geq I(WU; USX|YZ) \\ &= I(W; USX|YZ) + I(U; USX|WYZ) \\ &= I(W; USX|YZ) + H(U|WYZ) - H(U|USXWYZ) \\ &= I(W; USX|YZ) + H(U|WYZ). \end{aligned} \quad (\text{V.105})$$

Moreover, the converse in the proof of Theorem 31 is still valid with the identification  $W = (C, U_{\sim T}, Y_{\sim T}, Z_{\sim T}, T)$ .

Note that (Le Treust, 2015c, [IC13], Section IV.B) gives a characterization of the empirical coordination region and the constraint for the mutual information is

$$0 \leq I(WU; YZ) - H(U) = I(WU; YZ) - H(U) - I(W; S|U)$$

which is the same as in (V.104) because of the Markov chain  $SZ - U - W$ .

### Independence between source and channel

Suppose that the channel state  $P_{S^n}$  is independent of the source and side information  $P_{U^n Z^n}$ , and that the target joint distribution is of the form  $\bar{P}_{UZV}^{\otimes n} \bar{P}_{SXY}^{\otimes n}$ . For simplicity, we will suppose that the encoder has perfect state information (see Figure V.13). Then we characterize the strong coordination region  $\mathcal{R}_{\text{IND}}$ .

Note that in this case the coordination requirements are three-fold: the random variables  $(U^n, Z^n, V^n)$  should be coordinated, the random variables  $(S^n, X^n, Y^n)$  should be coordinated and finally  $(U^n, Z^n, V^n)$  should be independent of  $(S^n, X^n, Y^n)$ . We introduce two auxiliary random variables  $W_1$  and  $W_2$ , where  $W_2$  is used to accomplish the coordination of  $(U^n, Z^n, V^n)$ , while  $W_1$  has the double role of ensuring the independence of source and state as well as coordinating  $(S^n, X^n, Y^n)$ .

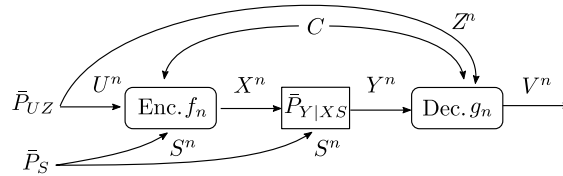


Figure V.13 – Coordination of signals and actions for a two-node network with a noisy channel where the source is separated from the channel.

**Theorem 32** Consider the setting of Theorem 29 and suppose that  $\bar{P}_{USXYZV} = \bar{P}_{UZV}\bar{P}_{SXY}$ . Then, the strong coordination region is

$$\mathcal{R}_{IND} := \left\{ \begin{array}{l} (\bar{P}_{USZXY}, R_0) \text{ such that:} \\ \bar{P}_{USZXYV} = \bar{P}_{UZ}\bar{P}_{V|UZ}\bar{P}_S\bar{P}_{X|S}\bar{P}_{Y|XS} \\ \exists (W_1, W_2) \text{ taking values in } \mathcal{W}_1 \times \mathcal{W}_2 \\ \bar{P}_{USZW_1W_2XYV} = \bar{P}_{UZ}\bar{P}_{W_2|U}\bar{P}_{V|ZW_2} \\ \quad \bar{P}_S\bar{P}_{X|S}\bar{P}_{W_1|SX}\bar{P}_{Y|XS} \\ I(W_1; S) + I(W_2; U) \leq I(W_1; Y) + I(W_2; Z) \\ R_0 \geq I(W_1; SX|Y) + I(W_2; UV|Z) \\ (|\mathcal{W}_1|, |\mathcal{W}_2|) \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 4. \end{array} \right. \quad (\text{V.106})$$

The proof of Theorem 32 is stated in (Cervia et al., 2020, [J9]).

**Remark 28** Observe that the decomposition of the joint distribution  $\bar{P}_{USZW_1W_2XYV}$  is equivalently characterized in terms of Markov chains:

$$\begin{cases} Z - U - W_2 \\ Y - (X, S) - W_1 \\ V - (Z, W_2) - U \end{cases} \quad (\text{V.107})$$

### Coordination under secrecy constraints

In this section we briefly discuss how in the separation setting of Section V.3.c, strong coordination offers additional security guarantees “for free”. In this context, the common randomness is not only useful to coordinate signals and actions of the nodes but plays the role of a secret key shared between the two legitimate users.

To simplify the notation, we do not consider channel state and side information at the decoder. Suppose there is an eavesdropper who observes the signals sent over the channel. We will show that not knowing the common randomness, the eavesdropper cannot infer any information about the actions.

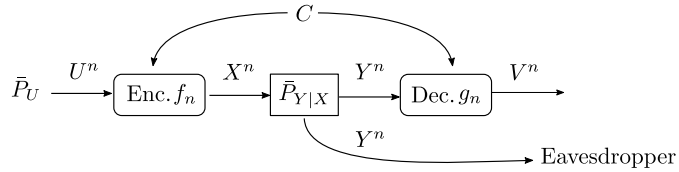


Figure V.14 – Wiretap channel: strong coordination implies secrecy.

**Lemma 4** In the setting of Theorem 32, without state and side information at the decoder, suppose that there is an eavesdropper that receives the same sequence  $Y^n$  as the decoder but has no knowledge of the common randomness. There exists a sequence  $(f_n, g_n)$  of strong coordination codes achieving the pair  $(\bar{P}_{UV}\bar{P}_{XY}, R_0) \in \mathcal{R}_{IND}$  such that the induced joint distribution  $P_{U^nV^nX^nY^n}^{RC}$  satisfies the strong secrecy condition Bloch and Barros (2011):

$$\lim_{n \rightarrow \infty} \mathbb{D}(P_{U^nV^nY^n}^{RC} \| P_{U^nV^n}^{RC} P_{Y^n}^{RC}) = \lim_{n \rightarrow \infty} I(U^nV^n; Y^n) = 0. \quad (\text{V.108})$$

### V.3.d Is separation optimal?

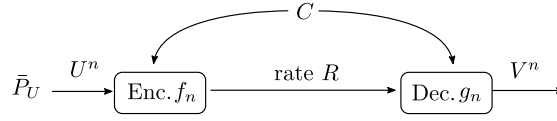


Figure V.15 – Coordination of the actions  $U^n$  and  $V^n$  for a two-node network with an error-free link of rate  $R$ .

Strong coordination over error-free channels was investigated in Cuff (2008); Cuff et al. (2010). When extending this analysis to noisy channels, it is natural to ask whether some form of separation theorem holds between coding for coordination and channel coding. In this section, we show that unlike the case of empirical coordination, separation does not hold for strong coordination.

If the separation principle is still valid for strong coordination, by concatenating the strong coordination of the source and its reconstruction with the strong coordination of the input and output of the channel we should retrieve the same mutual information and rate constraints. In order to prove that separation does not hold, first we consider the optimal result for coordination of actions in Cuff (2008); Cuff et al. (2010) and then we compare it with our result on joint coordination of signals and actions. In particular, since we want to compare the result in Cuff (2008); Cuff et al. (2010) with an exact region, we consider the case in which the channel is perfect and the target joint distribution is of the form  $\bar{P}_{UV}^{\otimes n} \bar{P}_X^{\otimes n}$ . The choice of a perfect channel might appear counterintuitive but it is motivated by the fact that we are trying to find a counterexample. As a matter of fact, if the separation principle holds for any noisy link, it should in particular hold for a perfect one.

We start by considering the two-node network with fixed source  $\bar{P}_U$  and an error-free link of rate  $R$  (Figure V.15). For this setting, Cuff (2008); Cuff et al. (2010) characterize the strong coordination region as

$$\mathcal{R}_{\text{Cuff}} := \left\{ \begin{array}{l} (\bar{P}_{UV}, R, R_0) \text{ such that} \\ \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \\ R \geq I(U; W) \\ R + R_0 \geq I(UV; W) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \quad (\text{V.109})$$

The result in Cuff (2008); Cuff et al. (2010) characterizes the trade-off between the rate  $R_0$  of available common randomness and the required description rate  $R$  for simulating a discrete memoryless channel for a fixed input distribution. We compare this region to our results when the requirement to coordinate the signals  $X^n$  and  $Y^n$  in addition to the actions  $U^n$  and  $V^n$  is relaxed. We consider, in the simpler scenario with no state and no side information, the intersection  $\mathcal{R}_{UV \otimes X} := \mathcal{R}_{\text{PC}} \cap \mathcal{R}_{\text{IND}}$ . The following result characterizes the strong coordination region.

**Proposition 11** Consider the setting of Theorem 28 and suppose that  $\bar{P}_{Y|X}(\mathbf{y}|\mathbf{x}) = \mathbb{1}_{X=Y}\{\mathbf{x} = \mathbf{y}\}$  and  $\bar{P}_{UXV} = \bar{P}_{UV} \bar{P}_X$ . Then, the strong coordination region is

$$\mathcal{R}_{UV \otimes X} := \left\{ \begin{array}{l} (\bar{P}_{UXV}, R_0) \text{ such that:} \\ \bar{P}_{UXV} = \bar{P}_U \bar{P}_{V|U} \bar{P}_X \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \bar{P}_X \\ I(W; U) \leq H(X) \\ R_0 \geq I(UV; W) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \quad (\text{V.110})$$

To compare  $\mathcal{R}_{\text{Cuff}}$  and  $\mathcal{R}_{UV \otimes X}$ , suppose that in the setting of Figure V.15 we use a codebook to send a message to coordinate  $U^n$  and  $V^n$ . In order to do so we introduce an i.i.d. source  $X^n$  with distribution

$P_X$  over  $\mathcal{X}$  in the model and we use the typical sequences of  $X^n$  as a codebook  $\mathcal{C}$ . Note that the codebook  $\mathcal{C}$  can be seen as an optimal channel code for the perfect channel. Hence, asymptotically  $R = H(X)$  and we rewrite the information constraints in (V.109) as

$$\begin{aligned} H(X) &\geq I(U; W), \\ R_0 &\geq I(UV; W) - H(X). \end{aligned}$$

This defines a new region:

$$\mathcal{R}_{\text{Cuff}, H(X)} := \left\{ \begin{array}{l} (\bar{P}_{UXV}, R_0) \text{ such that:} \\ \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ X \text{ generated according to } \bar{P}_X \\ \exists W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \bar{P}_X \\ I(W; U) \leq H(X) \\ R_0 \geq I(UV; W) - H(X) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \quad (\text{V.111})$$

The distributions  $\bar{P}_{UXV}$  in  $\mathcal{R}_{\text{Cuff}, H(X)}$  coordinate separately  $X^n$  and  $(U^n, V^n)$ . In Cuff (2008) the joint distribution  $P_{U^n V^n}$  should be  $\varepsilon$ -close to the i.i.d. distribution  $\bar{P}_{UV}^{\otimes n}$ , and we have imposed that  $X^n$  is generated according to the uniform distribution.

Observe that, while the information constraint is the same in the two regions (V.110) and (V.111), the rate of common randomness  $R_0$  required for the strong coordination region in (V.110) is larger than the rate of common randomness in (V.111). In fact, in the setting of Figure V.15 both  $X^n$  and the pair  $(U^n, V^n)$  achieve coordination separately (i.e.  $P_X^n$  is close to  $\bar{P}_X^{\otimes n}$  and  $P_{U^n V^n}$  is close to  $\bar{P}_{UV}^{\otimes n}$ ), but there is no extra constraint on the joint distribution  $P_{U^n X^n V^n}$ . On the other hand, the structure of our setting in (V.110) is different and requires the control of the joint distribution  $P_{U^n X^n V^n}$  which has to be  $\varepsilon$ -close to the i.i.d. distribution  $\bar{P}_{UV}^{\otimes n} \bar{P}_X^{\otimes n}$ . Since we are imposing a more stringent constraint, it requires more common randomness.

**Remark 29** We found  $\mathcal{R}_{UV \otimes X}$  as the intersection of two regions, but we can give it the following interpretation starting from  $\mathcal{R}_{\text{Cuff}}$ . By identifying  $R = H(X)$  in  $\mathcal{R}_{\text{Cuff}}$ , we find that the rate of common randomness has to be greater than  $I(UV; W) - H(X)$ . But this is not enough to ensure that  $X^n$  is independent of  $(U^n, V^n)$ . In order to guarantee that, we apply a one-time pad on  $X^n$  (which requires an amount of fresh randomness equal to  $H(X)$ ) and we have

$$R_0 \geq I(UV; W) - H(X) + H(X) = I(UV; W)$$

which is the condition on the rate of common randomness in (V.110).

The following example shows that, unlike the case of empirical coordination (Le Treust, 2017, [J6]), separation does not hold for strong coordination.

**Example 3** The difference in terms of rate of common randomness  $R_0$  is better shown in an example: when separately coordinating the two blocks  $X^n$  and  $(U^n, V^n)$  without imposing a joint behavior  $P_{U^n V^n X^n}$ , the same bits of common randomness can be reused for both purposes, and the required rate  $R_0$  is lower. We consider the case, already analyzed in Cuff (2008, 2013), of a Bernoulli-half source  $U$ , and  $V$  which is an erasure with probability  $p_e$  and is equal to  $U$  otherwise. In Cuff (2013) the authors prove that the optimal choice for the joint distributed  $P_{UWV}$  is the concatenation of two erasure channels  $\bar{P}_{W|U}$  and  $\bar{P}_{V|W}$  with erasure probability  $p_1$  and  $p_2$  respectively. Then we have

$$p_2 \in [0, \min\{1/2; p_e\}], \quad p_1 = 1 - \frac{1 - p_e}{1 - p_2},$$

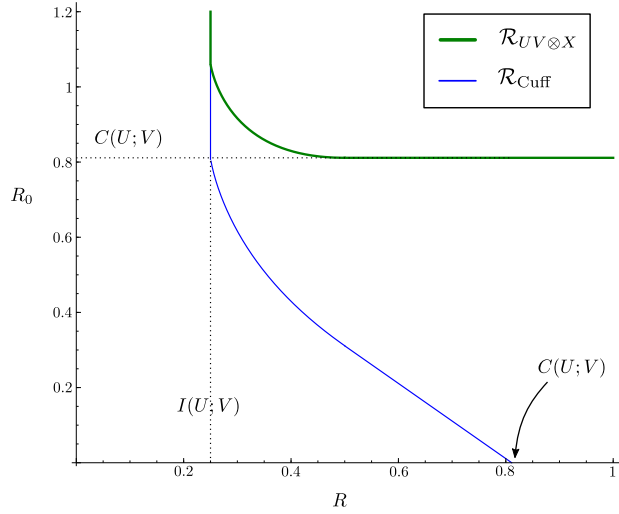


Figure V.16 – Comparison of the joint coordination region  $\mathcal{R}_{UV \otimes X}$  with  $\mathcal{R}_{\text{Cuff}}$  Cuff (2008, 2013): boundaries of the regions for a binary erasure channel with erasure probability  $p_e = 0.75$  and a Bernoulli-half input.

and therefore we obtain

$$\begin{aligned} I(U; W) &= 1 - p_1, \\ I(UV; W) &= h(p_e) + (1 - p_1)(1 - h(p_2)), \end{aligned}$$

where  $h$  is the binary entropy function. Figure V.16 shows the boundaries of the regions (V.109) (red) and (V.110) (red) for  $p_e = 0.75$  and a Bernoulli-half input. The dotted bound  $R \geq I(U; V)$  comes directly from combining  $R \geq I(U; W)$  with the Markov chain  $U - W - V$ . At the other extreme, if  $R_0 = 0$  in (V.109),  $R + R_0 \geq I(UV; W) \geq C(U; V)$  where  $C(U; V) := \min_{U-W-V} I(U, V; W)$  is Wyner’s common information Cuff (2008). On the other hand, in our setting (V.110),  $R_0 \geq I(UV; W) \geq C(U; V)$  for any value of  $R = H(X)$ .

Moreover, note that as  $R = H(X)$  tends to infinity, there is no constraint on the auxiliary random variable  $W$  (aside from the Markov chain  $U - V - W$ ) and similarly to Lapidath and Wigger (2016) the minimum rate of common randomness  $R_0$  needed for strong coordination is Wyner’s common information  $C(U; V)$ . In particular to achieve joint strong coordination of  $(U, X, V)$  a positive rate of common randomness is required. The boundaries of the rate regions only coincide on one extreme, and  $\mathcal{R}_{UV \otimes X}$  is strictly contained in  $\mathcal{R}_{\text{Cuff}}$ .

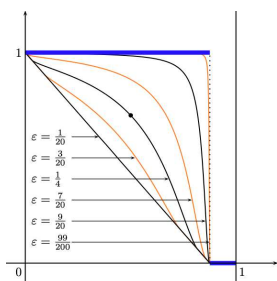
## V.4 Chapter summary

In this chapter, we present three research directions that are based on the coordination results of Chap. IV. First, we revisit the problems of state masking of Merhav and Shamai (2007) and state amplification Kim et al. (2008) through the lens of empirical coordination. Second, we extend the coordination result for finite alphabet to the Gaussian case, and we provide fundamental bounds for a long-standing open problem in the control literature, the counterexample of Witsenhausen (1968). Third, we study a finer notion of coordination, namely the strong coordination introduced in Cuff et al. (2010), which builds a bridge with Physical Layer Security problems, see Bloch and Barros (2011).

In the next chapter, we will consider that the encoder and decoder select their coding strategies in a strategic manner, i.e. they optimize their own cost function.

# Chapter VI

## Bayesian persuasion with restricted communication



We consider a game of information transmission in which the encoder and the decoder optimize their respective cost functions. As in the Bayesian persuasion model of [Kamenica and Gentzkow \(2011\)](#), we assume that the encoder commits to a strategy and we study the incentive constraints of the decoder. The natural solution concept is the Stackelberg equilibrium in which the encoder is the leader and the decoder is the follower. In the first section, we study the restrictions imposed by the channel on this strategic communication, see also ([Le Treust and Tomala, 2019, \[J8\]](#)), ([Le Treust and Tomala,](#)

[2016, \[IC16\]](#)), ([Le Treust and Tomala, 2020, \[IC26\]](#)), ([Le Treust and Tomala, 2017, \[NC8\]](#)). The strategic nature of the problem changes when the decoder privately observes a random variable correlated with the source. This aspect is treated in the second section, that corresponds to ([Le Treust and Tomala, 2018b, \[IC20\]](#)), ([Le Treust and Tomala, 2021, \[IC28\]](#)), ([Le Treust and Tomala, 2018a, \[W2\]](#)), ([Le Treust and Tomala, 2019, \[W3\]](#)). The third section presents Rony Bou Rouchael's PhD contributions on persuasion problems with two decoders via the [Gray and Wyner \(1974\)](#) communication network. These contributions are presented in ([Bou Rouchael and Le Treust, 2020, \[IC24\]](#)), ([Bou Rouchael and Le Treust, 2021, \[IC28\]](#)), ([Bou Rouchael and Le Treust, 2022c, \[IC31\]](#)), ([Bou Rouchael and Le Treust, 2022b, \[W7\]](#)), ([Bou Rouchael and Le Treust, 2022a, \[NC5\]](#)), ([Bou Rouchael and Le Treust, 2022d, \[W5\]](#)).

### Contents

---

VI.1 Persuasion with limited communication capacity . . . . .	<b>83</b>
VI.1.a Bayesian persuasion model . . . . .	86
VI.1.b Persuasion model with restricted communication . . . . .	87
VI.1.c Results . . . . .	90
VI.1.d Convexification with information constraint . . . . .	93
VI.1.e Illustrating example . . . . .	94
VI.1.f Beyond identical problems . . . . .	100
VI.2 Persuasion game with side information at the decoder . . . . .	<b>101</b>
VI.2.a System model . . . . .	102
VI.2.b Convex closure formulation . . . . .	104
VI.2.c Doubly symmetric binary source . . . . .	106
VI.3 Bayesian persuasion with several decoders . . . . .	<b>111</b>
VI.3.a System model . . . . .	111
VI.3.b Main result . . . . .	112
VI.4 Chapter summary . . . . .	<b>113</b>

---

## VI.1 Persuasion with limited communication capacity

In modern internet societies, pieces of information are repeatedly and continuously disclosed to decision makers by informed agents. Information transmission is affected by at least two sources of friction. First, the encoder and the decoder of a given signal may have nonaligned incentives, in which case the encoder might be unwilling to transmit truthful information. Second, communication between agents is often imperfect. The encoder and the decoder may have time constraints to write or read signals, forcing the encoder to summarize his arguments and making him unable to convey all the details. Further, there might be discrepancies between the informational content of a signal that is intended by the encoder and the one understood by the decoder. For instance, if the mother tongue of the encoder and of the decoder are different, there are possible translation errors (See [Blume et al., 2007](#)). Additionally, signals travelling in a network of computers might be subject to random shocks, internal errors or protocol failures. Studying the effect of noise in communication channels is the starting point of Information Theory ([Shannon, 1948](#)).

How does imperfect communication reduce the possibilities of persuasion in a sender-receiver game? When the encoder communicates many pieces of information, to what extent does tying the pieces together help in overcoming the communication limitations?

We consider an encoder and a decoder who communicate over an imperfect channel and are engaged in a series of  $n \in \mathbb{N}^*$  persuasion problems. The encoder observes  $n$  independent and identically distributed pieces of information and sends  $k \in \mathbb{N}^*$  signals to the decoder. Messages are sent through a channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$  that consists of two finite sets  $\mathcal{X}$ ,  $\mathcal{Y}$  of respectively inputs and outputs signals and of a transition probability  $\mathcal{T}_{Y|X} \in \Delta(\mathcal{Y})^{|\mathcal{X}|}$  such that when the encoder chooses input symbol  $x$ , the decoder receives output symbol  $y$  with probability  $\mathcal{T}_{Y|X}(y|x) \in [0, 1]$ . Upon receiving  $k$  output signals from the channel, the decoder chooses  $n$  actions, one for each problem. Cost functions are additively separable across persuasion problems. We assume that the encoder is able to commit to a disclosure strategy that maps sequences of pieces of information to distributions of sequences of input signals.

We study the optimal average cost secured by the encoder by committing to a strategy. We give an upper bound on this optimal cost and show that this bound is achieved asymptotically when the numbers  $n$  and  $k$  grow large. To prove this latter statement, we borrow techniques from Information Theory, namely, the coding and decoding schemes of [Shannon \(1948, 1959\)](#). This machinery allows to transmit a sequence of signals over a noisy channel with the property that the decoder recovers almost all signals correctly. The information theoretic literature typically considers an obedient decoder who calculates the decoded signals and takes them at face value. In the persuasion game framework, the decoder is strategic and may not follow any prescribed scheme. Rather, the decoder takes into account the strategy of the encoder and the received outputs, calculates its Bayesian belief about the sequence of states, and chooses a sequence of actions that minimizes its cost. Our technical contribution is to construct a strategy of the encoder for which we are able to estimate and to control those Bayesian beliefs in order to ensure that the strategic decoder chooses a desired sequence of actions.

Our upper bound is the value of an *optimal splitting problem with information constraint*, which represents the best cost that the encoder can achieve by sending a signal, subject to the constraint that the mutual information between the state and the signal is no more than the capacity of the channel. We show that this value is given by the convex closure of the cost function of the encoder, subject to a constraint on the entropy of posterior beliefs. This is also given by the convex closure of a modified cost function, where the encoder pays a cost proportional to the mutual information between the state and the signal.

### Motivating example.

There are relevant situations where an encoder discloses information about a large number of independent state parameters. For instance, one can think of testing product quality: a firm has many items to sell, which are ex-ante identical, and the authorities (e.g., the FDA for drugs) design quality tests<sup>1</sup>. One can also think about designing and grading exams to assess the quality of a large number of students<sup>2</sup>.

<sup>1</sup>See e.g., [Perez-Richet and Skreta, 2022](#).

<sup>2</sup>See [Boleslavsky and Cotton, 2015](#) for a model of grading standards through Bayesian persuasion.



As an example, consider an innovating firm that has several projects to be financed by investors. The board of investors audits the firm, which is given a limited amount of time to present all the projects. How to best structure arguments in order to get the maximum number of projects approved?

To be specific, let us assume that all projects are ex-ante identical and equally likely to be of good or bad quality. When a project is approved, it yields a cost of  $-1$  to the investors if it is good, and a cost of  $7$  if it is bad; rejecting a project yields a cost of  $0$ . The objective of the firm is to get a minimum number of projects approved.

Suppose that the firm commits to an information disclosure mechanism, i.e., distributions of signals conditional on states (as in [Kamenica and Gentzkow, 2011](#)) and faces no restriction on the number of signals. To invest, the board of investors must be persuaded that the project is good with probability at least  $7/8$ . Thus, for each project, the firm would optimally draw a good signal  $g$  or a bad signal  $b$  with the following probabilities:

$$\Pr(g \mid \text{project is good}) = 1, \quad \Pr(g \mid \text{project is bad}) = 1/7.$$

This way, the belief that the project is good upon receiving the good signal is as follows:

$$\Pr(\text{project is good} \mid g) = 7/8,$$

and the project is accepted with probability  $4/7$  (see Section [VI.1.e](#)).

Now, suppose that the auditing board gives the firm only half the time it would require to talk about all projects. Namely, there is an even number  $n$  of projects, but the firm has only  $n/2$  signals available.

A simple strategy the firm can adopt would be to select half of the projects, focus on them, and communicate optimally for each of them. With this strategy, half of the projects are accepted with probability  $4/7$  each, so in expectation, the average number of accepted projects is  $2/7$ . This is not optimal, and a better strategy would be to pair projects by two and to draw one signal  $g, b$  for each pair in the following way:

$$\Pr(g \mid \text{both projects are good}) = 1, \quad \Pr(g \mid \text{both projects are bad}) = 0,$$

$$\Pr(g \mid \text{only one project is good}) = 1/6.$$

The total probability of  $g$  is  $1/3$  and upon observing this signal, the beliefs about quality are as follows:

$$\Pr(\text{both projects are good} \mid g) = 6/8,$$

$$\Pr(\text{only project 1 is good} \mid g) = \Pr(\text{only project 2 is good} \mid g) = 1/8.$$

Therefore, each project is believed to be good with probability  $7/8$  and both projects are accepted when  $g$  is received. Thus, the expected average number of accepted projects is  $1/3 > 2/7$ .

We thus see that tying projects together improves upon communication about each project separately. Suppose that the number of projects is large. Is it possible to find a more complex strategy that further lowers the cost?

Our main result, [Theorem 33](#), gives an upper bound on the expected average number of accepted projects when the number of signals is half the number of projects. The upper bound is tight: the optimal value approaches it as the number of project increases. In this example, the upper bound is  $\lambda^*$  where  $(\lambda^*, p^*)$  is the unique solution in  $[0, 1] \times [0, \frac{1}{2}]$  of the system of equations:

$$\frac{1}{2} = \lambda^* \frac{7}{8} + (1 - \lambda^*) p^*, \quad \frac{1}{2} = \lambda^* H\left(\frac{7}{8}\right) + (1 - \lambda^*) H(p^*),$$

where  $H(p) = -p \log(p) - (1 - p) \log(1 - p)$  is the binary entropy function. The first equation is Bayes plausibility ([Kamenica and Gentzkow, 2011](#)) coming from Bayes' rule, saying that the expected posterior belief is the prior belief. The second equation requires the expected entropy of the posterior to be  $\frac{1}{2}$ , which means that the *mutual information* between the quality of the project and the signal sent to the decoder is equal to the number of signals per project that the firm is able to transmit.

Numerically  $\lambda^* \approx 0.519 < \frac{4}{7} \approx 0.571$ . Thus, for large  $n$ , the expected number of projects is higher than  $1/3$  but bounded away from the value obtained with unrestricted communication.

## Related literature

We now describe the relationships between our contribution and the literature. This work is at the junction of Bayesian persuasion and Information Theory.

The traditional game theoretic approach to strategic information disclosure assumes perfect communication and analyzes in isolation the problem of sending a single signal. These are the well-known sender-receiver games where an informed player, the encoder, communicates once with a decoder who takes an action. In the *cheap talk* version of this game, the signal sent by the encoder is costless and unverifiable; see for instance the seminal paper of Crawford and Sobel (1982). In the *Bayesian persuasion* game (Kamenica and Gentzkow, 2011), the encoder chooses verifiably an information disclosure device prior to learning his information. That is, the encoder is an *information designer* (Bergemann and Morris, 2016, 2017; Taneva, 2019) who chooses, without knowledge of the state, the information or signaling structure which releases information to the decision maker.

In parallel, Information Theory considers agents with perfectly aligned interests and analyzes the *rate* of information transmission. the encoder observes an information *flow*, which is a stochastic process, and sends signals to the decoder over an imperfect channel represented by a transition probability from input to output signals. Truthful information transmission is the common goal of the encoder and the decoder. The *rate of information transmission* is the average number of correct guesses made by the decoder. Shannon's theory (Shannon, 1948, 1959) determines whether a source of information can be transmitted over the channel with arbitrarily small probability of error and shows that the rate of the source of information has to be smaller than the *capacity of the channel* defined as the maximal *mutual information* between input and output signals.

Our model of persuasion has two essential features. the encoder and the decoder are engaged in a large number of identical copies of the same game and communication is restricted to an imperfect channel. As Kamenica and Gentzkow (2011), we consider the cost obtained by the encoder as a function of the belief of the decoder, when the decoder takes optimal actions. With unrestricted communication, that is on a perfect channel with large set of inputs, the optimal cost for the encoder is given by the convex closure of this function. Then, solving any number of identical games amounts to solving each copy separately. With a single copy, the game of persuasion with a noisy channel is studied by Tsakas and Tsakas (2021) who prove the existence of optimal solutions and show monotonicity of the encoder's cost with respect to the noise of the channel. Considering many copies of the base game *and* restricted communication, we show that *linking independent problems together* yields a better cost to the encoder: the optimal strategy correlates all signals with the state parameters of all problems. In this respect, our work bears some similarity with Jackson and Sonnenschein (2007), who showed that a mechanism designer can achieve more outcomes in an incentive compatible manner by linking many identical problems together.

The optimal cost that we characterize is related to models where the cost of information is measured by mutual information. Such information costs have been introduced in the literature on rational inattention by Sims (2003), (See also Martin, 2017; Matejka and McKay, 2015; Steiner et al., 2017). The use of mutual information has been axiomatized in Morris and Strack (2019) and Hebert and Woodford (2018). In the context of persuasion, Gentzkow and Kamenica (2014) consider a model where the encoder gets his cost from the game, minus a cost that is proportional to the mutual information between the state and the signal; see also Matyskova (2018). With Lagrangian methods, we find that the value of our optimal splitting problem with information constraint is the convex closure of the cost function, net of such an information cost, a similar convexification problem is found in Caplin and Dean (2013)

Different from those papers, the mutual information is not a primitive of our model. Our finding is that the noise and limitations in communication induce a *shadow cost* measured by the mutual information. Entropy and mutual information appear endogenously in several papers on repeated games where players have bounded rationality (Neyman and Okada, 1999, 2000), are not able to freely randomize their actions (Gossner and Vieille, 2002), or observe actions imperfectly (Gossner and Tomala, 2006, 2007). A related paper is Gossner et al. (2006), henceforth GHN, who also consider a sender-receiver game. In GHN, the encoder and the decoder play an infinitely repeated game with common interests: both the encoder and the decoder want to choose the action that matches the state. the encoder knows the infinite sequence of states and can communicate with the decoder only through his actions. GHN characterize the best average cost that the encoder (and the decoder) can achieve. Their solution resembles ours: the optimal value is the cost obtained when the encoder can send a direct signal to the decoder, subject to an

information constraint.

### VI.1.a Bayesian persuasion model

In this section, we briefly present the Bayesian persuasion model of [Kamenica and Gentzkow \(2011\)](#), depicted on Fig. VI.1. The encoder and the decoder are endowed with distinct costs functions  $c_e : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}$  and  $c_d : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}$ . The main difference with the Information-Theoretic models of Chap. III is that:

1. the perfect channel has a large cardinality  $|\mathcal{X}| \geq |\mathcal{U}|$ ,
2. the sequences of symbols have length  $n = 1$ .

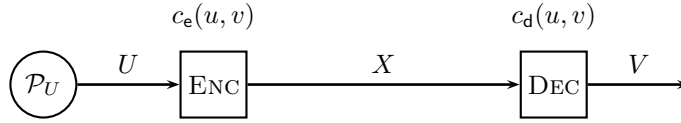


Figure VI.1 – Bayesian persuasion model

The encoding and decoding functions are defined by

$$\sigma : \mathcal{U} \rightarrow \Delta(\mathcal{X}), \quad \tau : \mathcal{X} \rightarrow \Delta(\mathcal{V}), \quad (\text{VI.1})$$

and induce a probability distribution given by

$$\Pr((U, X, V) = (u, x, v)) = \mathcal{P}_U(u)\sigma(x|u)\tau(v|x), \quad \forall (u, x, v). \quad (\text{VI.2})$$

Given an encoder strategy  $\sigma$ , we define the set of decoder best responses by

$$\text{BR}(\sigma) = \underset{\tau}{\operatorname{argmin}} \mathbb{E}[c_d(U, V)]. \quad (\text{VI.3})$$

If the set  $\text{BR}(\sigma)$  contains several strategies, we assume that the decoder selects the worst one for the encoder cost function. Note that in case of indifference in the model of [Kamenica and Gentzkow \(2011\)](#), the decoder selects the best response strategy  $\tau \in \text{BR}(\sigma)$  that also minimizes the encoder cost function.

**Definition 41** *The robust Bayesian persuasion problem writes*

$$C_e^* = \inf_{\sigma} \max_{\tau \in \text{BR}(\sigma)} \mathbb{E}[c_e(U, V)]. \quad (\text{VI.4})$$

In this strategic communication scenario, the encoder commits to implementing the strategy  $\sigma$ , by anticipating that the decoder selects a best response  $\tau \in \text{BR}(\sigma)$ .

**Definition 42** *Given the source probability distribution  $p \in \Delta(\mathcal{U})$ , the decoder selects a best response action  $v$  in*

$$\mathcal{V}^*(p) = \underset{v \in \mathcal{V}}{\operatorname{argmin}} \mathbb{E}_p[c_d(U, v)]. \quad (\text{VI.5})$$

Given  $p \in \Delta(\mathcal{U})$ , we define the robust cost function by

$$\psi_e(p) = \max_{v \in \mathcal{V}^*(p)} \mathbb{E}_p[c_e(U, v)]. \quad (\text{VI.6})$$

The next results is a reformulation of ([Kamenica and Gentzkow, 2011](#), Prop. 1), where  $\operatorname{vex} f : \Delta(\mathcal{U}) \rightarrow \mathbb{R} \cup \{-\infty\}$  denotes the convex closure of a function  $f$ , i.e. the largest convex function everywhere smaller than  $f$  on  $\Delta(\mathcal{U})$ .

**Proposition 12** *If  $|\mathcal{X}| \geq |\mathcal{U}|$ , we have*

$$C_e^* = \text{vex } \psi_e(\mathcal{P}_U). \quad (\text{VI.7})$$

*Proof.* [Prop. 12] The main argument comes from the splitting lemma of [Aumann and Maschler \(1995\)](#), also referred to as Bayes plausibility in [Kamenica and Gentzkow \(2011\)](#). We recall the definition of convex closure,

$$\text{vex } \psi_e(\mathcal{P}_U) = \inf \left\{ \sum_{x \in \mathcal{X}} \lambda_x \psi_e(p_x), \quad \sum_{x \in \mathcal{X}} \lambda_x p_x = \mathcal{P}_U \right\}, \quad (\text{VI.8})$$

where the infimum is taken over the parameters  $(\lambda_x, p_x)_{x \in \mathcal{X}}$  with  $|\mathcal{X}| = |\mathcal{U}|$ , such that for each  $x \in \mathcal{X}$ ,  $p_x \in \Delta(\mathcal{U})$ ,  $\lambda_x \in [0, 1]$ , and  $\sum_{x \in \mathcal{X}} \lambda_x = 1$ .

Given an encoder strategy  $\sigma$ , we define for all  $x \in \mathcal{X}$ ,

$$\lambda_x = \sum_{u \in \mathcal{U}} \mathcal{P}_U(u) \sigma(x|u), \quad (\text{VI.9})$$

$$p_x(u) = \frac{\mathcal{P}_U(u) \sigma(x|u)}{\sum_{u' \in \mathcal{U}} \mathcal{P}_U(u') \sigma(x|u')}, \quad \forall u \in \mathcal{U}. \quad (\text{VI.10})$$

This implies that  $\lambda_x p_x(u) = \mathcal{P}_U(u) \sigma(x|u)$  for all  $(u, x) \in \mathcal{U} \times \mathcal{X}$ , and thus

$$\min_{\tau} \mathbb{E}[c_d(U, V)] = \min_{\tau} \sum_{u, x, v} \mathcal{P}_U(u) \sigma(x|u) \tau(v|x) c_e(u, v) \quad (\text{VI.11})$$

$$= \min_{\tau} \sum_{u, x, v} \lambda_x p_x(u) \tau(v|x) c_e(u, v) \quad (\text{VI.12})$$

$$= \sum_x \lambda_x \min_v \sum_u p_x(u) c_e(u, v) \quad (\text{VI.13})$$

$$= \sum_x \lambda_x \min_v \mathbb{E}_{p_x} [c_e(U, v)]. \quad (\text{VI.14})$$

BY using similar arguments, we have

$$\max_{\tau \in \text{BR}(\sigma)} \mathbb{E}[c_e(U, V)] = \sum_x \lambda_x \max_{v \in \arg\min \mathbb{E}_{p_x} [c_d(U, v)]} \mathbb{E}_{p_x} [c_e(U, v)] = \sum_{x \in \mathcal{X}} \lambda_x \psi_e(p_x), \quad (\text{VI.15})$$

and thus  $C_e^* \geq \text{vex } \psi_e(\mathcal{P}_U)$ .

Given the parameters  $(\lambda_x, p_x)_{x \in \mathcal{X}}$ , we define the encoder strategy by

$$\sigma(x|u) = \frac{\lambda_x p_x(u)}{\sum_{u'} \lambda_x p_x(u')}, \quad (\text{VI.16})$$

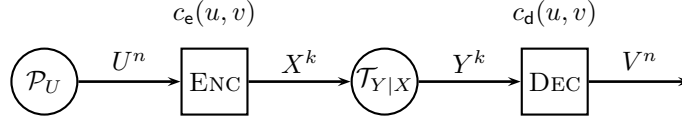
for which (VI.15) is also satisfied. Then, we obtain  $C_e^* \leq \text{vex } \psi_e(\mathcal{P}_U)$ .  $\square$

### VI.1.b Persuasion model with restricted communication

We consider an encoder and a decoder engaged in a series of identical persuasion problems and where the communication technology is fixed exogenously, depicted in Fig. VI.2. More specifically, we consider a fixed communication channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ , where  $\mathcal{X}, \mathcal{Y}$  are finite sets of signals and  $\mathcal{T}_{Y|X} \in \Delta(\mathcal{Y})^{|\mathcal{X}|}$  is a transition probability from  $\mathcal{X}$  to  $\mathcal{Y}$ .

Given two integers  $n, k$ , we define a repeated persuasion problem where the uncertainty is about a sequence  $u^n = (u_1, \dots, u_n)$  drawn i.i.d. from  $\mathcal{P}_U \in \Delta(\mathcal{U})$ . The decoder chooses a sequence of actions  $v^n = (v_1, \dots, v_n)$  and the cost for player  $i = e, d$  is as follows:

$$\bar{c}_i(u^n, v^n) = \frac{1}{n} \sum_{t=1}^n c_i(u_t, v_t).$$


 Figure VI.2 – Repeated persuasion problem via a noisy channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ .

To disclose information, the encoder can use the channel  $k$  times by choosing a sequence of input signals  $x^k = (x_1, \dots, x_k)$ . The channel then draws a sequence of output signals  $y^k$  with probability  $\mathcal{T}_{Y|X}^k(y^k|x^k) = \prod_{t=1}^k \mathcal{T}_{Y|X}(y_t|x_t)$  and sends it to the decoder. This defines the following persuasion game  $\Gamma(n, k)$ :

1. The encoder chooses a strategy  $\sigma : \mathcal{U}^n \rightarrow \Delta(X^k)$  which is announced to the decoder.
2. A sequence of states  $u^n$  is drawn i.i.d. from the prior  $\mathcal{P}_U$ , a sequence of input signals  $x^k$  is drawn with probability  $\sigma(x^k|u^n)$ , a sequence of output signals  $y^k$  is drawn with probability  $\mathcal{T}_{Y|X}^k(y^k|x^k)$  and is observed by the decoder.
3. the decoder chooses a sequence of actions  $v^n$ .

Then, player  $i = e, d$  gets the average cost  $\bar{c}_i(u^n, v^n)$ .

Notice that for  $n = k = 1$ , this is the model of [Tsakas and Tsakas \(2021\)](#) of a single persuasion problem with noisy communication. An interesting particular case is given by *perfect* channels where  $\mathcal{X} = \mathcal{Y}$  and  $\mathcal{T}_{Y|X}(y|x) = \mathbb{1}_{\{y=x\}}$  for all  $(x, y)$ . In such a case, the only limitation is given by the number of signals. If we let  $n = k = 1$  and choose a perfect channel with sufficiently many signals  $|\mathcal{X}| = |\mathcal{Y}| \geq |\mathcal{U}|$ , the model encompasses the standard persuasion game of [Kamenica and Gentzkow \(2011\)](#).

### Optimal robust cost

As a solution concept, we study the best cost the encoder can secure, regardless of which best reply is chosen by the decoder. A strategy of the decoder is a mapping  $\tau : \mathcal{Y}^k \rightarrow \mathcal{V}^n$ . Knowing  $\sigma$ , the decoder chooses a best reply  $\tau$ , which minimizes the expected cost. That is, for each  $y^k$ :

$$\tau(y^k) \in \operatorname{argmin}_{v^n \in \mathcal{V}^n} \sum_{u^n, x^k} \mathcal{P}_U^n(u^n) \sigma(x^k|u^n) \mathcal{T}_{Y|X}^k(y^k|x^k) \bar{c}_d(u^n, v^n).$$

Denote  $BR_d(\sigma)$  the set of best replies of the decoder to the strategy  $\sigma$ .

**Definition 43** *The optimal robust cost of the encoder in this problem is as follows:*

$$C_e^{n,k}(\mathcal{P}_U, \mathcal{T}_{Y|X}) = \inf_{\sigma} \max_{\tau \in BR_d(\sigma)} \sum_{u^n, x^k, y^k} \mathcal{P}_U^n(u^n) \sigma(x^k|u^n) \mathcal{T}_{Y|X}^k(y^k|x^k) \bar{c}_e(u^n, \tau(y^k)).$$

This definition differs from the conventional solution to Bayesian persuasion of [Kamenica and Gentzkow \(2011\)](#) where the decoder takes the best reply which is preferred by the encoder. Our choice is motivated by robustness; we ask the solution to be robust to the way the decoder breaks ties<sup>3</sup>. We stress that this choice does not matter for generic problems. Indeed, with slight perturbations of the cost function of the decoder, we can make sure that indifferences occur only at interior beliefs. When this is the case, the encoder can slightly change his strategy in order to avoid the indifference region.

The goal is to give an upper bound for the optimal robust cost and to characterize its limit when  $n$  and  $k$  tend to infinity.

<sup>3</sup>A similar approach is followed by [Inostroza and Pavan \(2021\)](#) and [Mathevet et al. \(2020\)](#).

### Optimal splitting problem with information constraint

To state our main results, we introduce some definitions.

**Definition 44** A splitting of  $\mathcal{P}_U \in \Delta(\mathcal{U})$  is a finite family  $(\lambda_w, \nu_w)_{w \in \mathcal{W}}$ , where for each  $w \in \mathcal{W}$ ,  $\nu_w \in \Delta(\mathcal{U})$ ,  $\lambda_w \in [0, 1]$ ,  $\sum_w \lambda_w = 1$  such that:

$$\mathcal{P}_U = \sum_w \lambda_w \nu_w. \quad (\text{VI.17})$$

A splitting of  $\mathcal{P}_U$  is a distribution of posterior beliefs whose average equals the prior. An “information structure” which draws a signal  $w$  with probability  $\Pr(w|u)$  in state  $u$ , induces a splitting  $(\lambda_w, \nu_w)_w$  with  $\lambda_w = \sum_{u'} \mathcal{P}_U(u') \Pr(w|u')$  and  $\nu_w(u) = \frac{\mathcal{P}_U(u) \Pr(w|u)}{\sum_{u'} \mathcal{P}_U(u') \Pr(w|u')}$ . From the splitting lemma (Aumann and Maschler, 1995) or Bayes plausibility (Kamenica and Gentzkow, 2011), for each decomposition of the prior belief into a convex combination of posterior  $\mathcal{P}_U = \sum_w \lambda_w \nu_w$ , the splitting  $(\lambda_w, \nu_w)_w$  is induced by some information structure, for example,  $\Pr(w|u) = \lambda_w \nu_w(u) / \mathcal{P}_U(u)$  for all  $(u, w) \in \mathcal{U} \times \mathcal{W}$ .

For each posterior belief  $\nu \in \Delta(\mathcal{U})$ , let the set of optimal actions of the decoder be:

$$\mathcal{V}^*(\nu) = \operatorname{argmin}_{v \in \mathcal{V}} \sum_u \nu(u) c_d(u, v).$$

We denote by  $c_e^*(\nu) = \min_{v \in \mathcal{V}^*(\nu)} \sum_u \nu(u) c_e(u, v)$  the *robust cost* of the encoder at the belief  $\nu \in \Delta(\mathcal{U})$ , i.e., the cost of the encoder when the decoder chooses the optimal action, which is *worst* for the encoder.

We now introduce tools borrowed from Information Theory; the reader is referred to Cover and Thomas (2006).

**Definition 45** 1. The (Shannon) entropy of a probability distribution  $q \in \Delta(\mathcal{S})$  over a finite set  $\mathcal{S}$  is as follows:

$$H(q) = - \sum_{s \in \mathcal{S}} q(s) \log q(s),$$

where the logarithm has basis 2 and  $0 \log 0 = 0$ .

2. Given a distribution  $p_X \in \Delta(\mathcal{X})$ , the mutual information between two random variables  $(X, Y)$ , drawn from the joint probability distribution  $p_X \mathcal{T}_{Y|X} \in \Delta(\mathcal{X} \times \mathcal{Y})$  is as follows:

$$I(X; Y) = H\left(\sum_x p_X(x) \mathcal{T}_{Y|X}(\cdot|x)\right) - \sum_x p_X(x) H(\mathcal{T}_{Y|X}(\cdot|x)) = H(Y) - H(Y|X).$$

3. The capacity of the channel  $(X, Y, \mathcal{T}_{Y|X})$  is defined by:

$$\mathbb{C}(\mathcal{T}_{Y|X}) = \max_{p_X \in \Delta(\mathcal{X})} I(X; Y).$$

The channel capacity  $\mathbb{C}(\mathcal{T}_{Y|X})$  is the minimal mutual information between two random variables  $(X, Y)$ , respectively the input and output of the channel, drawn from the joint probability distribution  $p_X(x) \mathcal{T}_{Y|X}(y|x)$ , where the minimum is over the marginal distribution  $p_X(x)$ . Intuitively, this is the minimal number of bits of information that can be transmitted reliably through the channel (see Cover and Thomas, 2006).

Equipped with these tools, our main definition is the following.

**Definition 46** For any  $C \geq 0$ , the optimal splitting problem with information constraint is:

$$\begin{aligned} C_e^*(\mathcal{P}_U, C) &= \inf \sum_w \lambda_w c_e^*(\nu_w) \\ \text{s.t.} \quad &\sum_w \lambda_w \nu_w = \mathcal{P}_U, \\ \text{and} \quad &H(U) - \sum_w \lambda_w H(\nu_w) \leq C. \end{aligned}$$

This is the best cost that the encoder can secure by choosing a splitting of the prior belief (i.e., an information structure) under the constraint that the expected reduction of entropy does not exceed the capacity  $C$  of the channel. The entropy reduction  $H(U) - \sum_w \lambda_w H(\nu_w) = H(U) - H(U|W)$  is nonnegative. It is equal to the mutual information  $I(U; W)$  between a random state  $U$  and a random signal  $W$ , drawn from the joint distribution  $(\lambda_w \nu_w(u))_{(u,w)}$ , see Prop. 2. The interpretation is thus that the encoder optimizes over a set of information structures that convey bounded information about the state.

Notice that  $C_e^*(\mathcal{P}_U, C)$  is greater than or equal to the convex closure (or convexification) of  $c_e^*$  at  $\mathcal{P}_U$  which is the unconstrained infimum  $\text{vex } c_e^*(\mathcal{P}_U) := \inf \{ \sum_w \lambda_w c_e^*(\nu_w) : \sum_w \lambda_w \nu_w = \mathcal{P}_U \}$ .

### VI.1.c Results

The main result shows that the value of the optimal splitting problem with information constraint provides an upper bound to the optimal robust cost and that the bound is achieved asymptotically.

**Theorem 33** 1. *The optimal robust cost of the encoder is no more than the value of the optimal splitting problem with information constraint. For each pair of integers  $n, k$ :*

$$C_e^{n,k}(\mathcal{P}_U, \mathcal{T}_{Y|X}) \leq C_e^*\left(\mathcal{P}_U, \frac{k}{n}C(\mathcal{T}_{Y|X})\right).$$

2. *The optimal robust cost of the encoder converges to the value of the optimal splitting problem with information constraint in the following sense. For each  $r \in [0, +\infty]$ , for each pair of sequences of integers  $(k_j, n_j)_{j \in \mathbb{N}^*}$  such that  $\lim_{j \rightarrow \infty} \min(n_j, k_j) = \infty$  and  $\lim_{j \rightarrow \infty} \frac{k_j}{n_j} = r$ , we have:*

$$\lim_{j \rightarrow \infty} C_e^{n_j, k_j}(\mathcal{P}_U, \mathcal{T}_{Y|X}) = C_e^*\left(\mathcal{P}_U, rC(\mathcal{T}_{Y|X})\right).$$

The proof of Theorem 33 is stated in (Le Treust and Tomala, 2019, [J8]). On the one hand, this result shows communication restrictions limits the cost that can be achieved through Bayesian persuasion. On the other hand, it quantifies the extent to which repeating the same problem and linking the copies together helps in overcoming those restrictions.

#### Sketch of proof

We give an intuition for the main arguments of the proof; the technical details are in the appendix.

**First point, upper bound.** The argument is that regardless of which strategies are used, the mutual information between the states and the signals to the decoder cannot exceed the capacity of the channel.

For simplicity, consider the case  $n = k = 1$  where the result says  $C_e^{1,1}(\mathcal{P}_U, \mathcal{T}_{Y|X}) \leq C_e^*(\mathcal{P}_U, C(\mathcal{T}_{Y|X}))$ . Take any strategy  $\sigma$  of the encoder. This induces the splitting  $\mathcal{P}_U = \sum_y \text{Pr}_\sigma(y) \nu_y$  where  $\text{Pr}_\sigma(y) = \sum_{u,x} \mathcal{P}_U(u) \sigma(x|u) \mathcal{T}_{Y|X}(y|x)$  is the probability of the signal  $y$  and

$$\nu_y(u) = \text{Pr}_\sigma(u|y) = \frac{\sum_x \mathcal{P}_U(u) \sigma(x|u) \mathcal{T}_{Y|X}(y|x)}{\text{Pr}_\sigma(y)}$$

is the posterior belief conditional on  $y$ . The mutual information of this splitting is:

$$H(U) - \sum_y \text{Pr}_\sigma(y) H(\nu_y) := I(U; Y)$$

where  $(U, X, Y)$  denotes a random triple of state, input and output signals drawn from the joint distribution  $\mathcal{P}_U(u) \sigma(x|u) \mathcal{T}_{Y|X}(y|x)$ . With an abuse of notation, we denote  $I(U; Y)$  the mutual information between  $U$  and  $Y$  without explicit reference to the distribution.

Since  $X$  is a sufficient statistic for  $Y$ ,  $X$  is more informative<sup>4</sup> about  $Y$  than  $U$ , that is  $I(U; Y) \leq I(X; Y)$ . Then, the mutual information between the input and the output is no more than  $C(\mathcal{T}_{Y|X})$  from the definition of the channel capacity.

<sup>4</sup>See Cover and Thomas, 2006, Theorem 2.8.1, p. 34.

The proof for general  $n$  and  $k$  is an elaboration of this argument. Since states are i.i.d., we can prove that for any strategy, the average cost is the one induced by some splitting whose mutual information is no more than  $\frac{k}{n}C(\mathcal{T}_{Y|X})$ . The trick is to introduce an auxiliary random variable  $T$  uniformly distributed over  $\{1, \dots, n\}$  and known by the decoder. Then, we regard the average cost over stages  $1, \dots, n$  as the expected cost for the randomly selected stage.

**Second point, asymptotic construction.** To make the intuition simple, let us consider a sequence of pairs of integers  $(k_j, n_j)_{j \in \mathbb{N}^*}$  such that  $k_j = n_j$  and let  $k = n$  be a large term of this sequence. Take a splitting  $(\lambda_w, \nu_w)_w$  of the prior  $\mathcal{P}_U$  which satisfies the information constraint. We want to show that for large  $n$ , there is a strategy  $\sigma$  of the encoder such that for any best reply  $\tau \in BR_d(\sigma)$  of the decoder, the cost of the encoder is at least about  $\sum_w \lambda_w c_e^*(\nu_w)$ . Let also  $v_w^* \in \mathcal{V}^*(\nu_w)$  such that  $c_e^*(\nu_w) = \sum_u \nu_w(u) c_e(u, v_w^*)$ .

A first intuition for the construction is as follows. From Shannon's coding Theorem<sup>5</sup>, if  $I(U; W) < C(\mathcal{T}_{Y|X})$ , then for large  $n$ , there exists functions  $f_1 : \mathcal{U}^n \rightarrow \mathcal{W}^n$ ,  $f_2 : \mathcal{W}^n \rightarrow \mathcal{X}^n$  and  $g : \mathcal{Y}^n \rightarrow \mathcal{W}^n$ , altogether a coding/decoding scheme, with the following properties. Given a sequence of states  $u^n$ , the encoder calculates a sequence of signals  $w^n = f_1(u^n)$  such that with probability close to one, the empirical frequency of the  $(u_t, m_t)$ 's is approximately the theoretical one  $\lambda_w \nu_w(u)$ . The encoder then calculates a sequence of inputs  $x^n = f_2(m^n)$  and sends them into the channel. If the decoder calculates  $\hat{m}^n = g(y^n)$ , then the signals are recovered with probability close to one:  $\Pr(m^n = \hat{m}^n) \approx 1$ .

This argument is standard in Information Theory but is not sufficient for proving our result. The proof is actually more complicated because the strategic decoder actually calculates the Bayesian posterior  $\Pr(u^n | y^n)$  and chooses at stage  $t$  an action  $v_t \in \mathcal{V}^*(\Pr(u_t | y^n))$ . Thus, the main task is to refine the construction in such a way that for any best reply of the decoder, with probability close to one, the optimal action  $v_t \in \mathcal{V}^*(\Pr(u_t | y^n))$  is equal to the recommended action  $v_{\hat{m}_t}^*$  at most stages, that is, for a set of stages whose proportion is close to one. This implies that the cost is approximately the target one.

The proof consists of three main steps. In the first step, we show that for each  $\varepsilon > 0$ , we can find a splitting  $\varepsilon$ -optimal for  $C_e^*(\mathcal{P}_U, C)$ , which satisfies the information constraint with strict inequality and such that for each posterior  $\nu_w$ , the action  $v_w^*$  which minimizes the encoder cost over  $\mathcal{V}^*(\nu_w)$  is *unique* in a neighborhood of  $\nu_w$ . This latter property ensures that the decoder plays  $v_w^*$  whenever its belief is close to  $\nu_w$ . We deduce that the difference between the realized cost and the target cost is bounded by the number of times  $t$  where the Bayesian posterior  $\Pr(u_t | y^n)$  is far away from  $\nu_{\hat{m}_t}$ . The goal is then to show that this number is small with probability close to one.

The second step consists in defining Shannon's strategy for this splitting. There, we adapt known construction from Information Theory to our setting.

At the third step, we prove that, under our construction, with probability close to one, the Bayesian posteriors  $\Pr(u_t | y^n)$  are close enough to the target posteriors  $\nu_{m_t}$  at most stages. This allows us to conclude that with probability close to one, the decoder plays the recommended actions at most stages and that the expected cost is close to the target one. This step, where we estimate the realized Bayesian beliefs, is new compared to the information theoretic literature, which typically focuses on the average number of mistakes in decoding. Summing up, our construction is similar to the ones found in this literature but is adapted to the context where the decoder is minimizing its cost.  $\square$

We now provide some direct implications of the theorem.

**Large capacity.** Reordering the information constraint as  $\sum_w \lambda_w H(\nu_w) \geq H(U) - C$ , we see that if  $C \geq H(U)$ , the constraint is satisfied by all splittings. The value of the problem is thus the unconstrained convexification of  $c_e^*$ :

$$C \geq H(U) \implies C_e^*(\mathcal{P}_U, C) = \text{vex } c_e^*(\mathcal{P}_U).$$

As a consequence, if we fix  $n$  and  $\mathcal{T}_{Y|X}$  and choose  $k$  large enough such that  $\frac{k}{n}C(\mathcal{T}_{Y|X}) \geq H(U)$ , then the encoder can achieve approximately the unconstrained minimum  $\text{vex } c_e^*(\mathcal{P}_U)$ .

The intuition is simple: for fixed size of the state space  $\mathcal{U}$ , if the imperfect channel can be used a large number of times, then the encoder is able to convey the state  $u \in \mathcal{U}$  with arbitrarily high probability.

<sup>5</sup>See Cover and Thomas, 2006, Theorem 10.4.1, p. 318.



More precisely, suppose  $C(\mathcal{T}_{Y|X}) > 0$  that is to say,  $\mathcal{T}_{Y|X}(\cdot|x)$  is not constant with respect to  $x$ . There exist distributions of inputs  $p_X^u \in \Delta(X)$  for  $u \in \mathcal{U}$ , that statistically identify the state:

$$u \neq u' \implies \sum_x p_X^u(x) \mathcal{T}_{Y|X}(\cdot|x) \neq \sum_x p_X^{u'}(x) \mathcal{T}_{Y|X}(\cdot|x).$$

For each state  $u \in \mathcal{U}$ , the encoder can draw an i.i.d. sequence of signals  $x_1, \dots, x_k$  from  $p_X^u$  and sends them through the channel. The posterior belief of the decoder conditional on  $y_1, \dots, y_k$  then converges to the truth (the Dirac mass on  $u$ ). Thus, asymptotically, the distributions of actions of the decoder will be close to the one under perfect communication.

**Small capacity.** When  $C$  is close to 0, the information constraint  $H(U) - \sum_w \lambda_w H(\nu_w) \leq C$  implies that the splitting is almost nonrevealing since:<sup>6</sup>

$$\sum_w \lambda_w \|\nu_w - \mathcal{P}_U\|_1 \leq \sqrt{2 \ln 2 (H(U) - \sum_w \lambda_w H(\nu_w))}.$$

It follows that  $C_e^*(\mathcal{P}_U, C)$  is approximately  $c_e^*(\mathcal{P}_U)$ , the cost obtained without any information transmission.

As a consequence, if we fix  $\mathcal{T}_{Y|X}$  and  $k$ , then for large  $n$ , the encoder cannot get substantially more than  $c_e^*(\mathcal{P}_U)$ .

**Perfect channels.** Our result applies to communication channels without noise. A communication channel has two sources of imperfection: the noise and the number of available signals, which is given exogenously. One insight of our work is that all that matters for the analysis is the capacity of the channel.

A channel  $(X, Y, \mathcal{T}_{Y|X})$  is called *perfect* if  $\mathcal{X} = \mathcal{Y}$  and  $\mathcal{T}_{Y|X}(y|x) = \mathbb{1}_{\{x=y\}}$ . For each integer  $|\mathcal{X}| \geq 2$ , we denote  $\mathcal{T}_{|\mathcal{X}|}^*$  the perfect communication channel with  $|\mathcal{X}|$  signals where  $|\mathcal{X}| = |\mathcal{Y}|$ . Its capacity is<sup>7</sup>  $C(\mathcal{T}_{|\mathcal{X}|}^*) = \log |\mathcal{X}|$ . We apply our results to the optimal robust cost  $C_e^{n,1}(\mathcal{P}_U, \mathcal{T}_{|\mathcal{X}|}^*)$  of the game where the persuasion problem is repeated  $n$  times and where the encoder can send *one* signal from a set with cardinality  $|\mathcal{X}|$ . Our method applies since for large  $|\mathcal{X}|$ , the channel  $\mathcal{T}_{|\mathcal{X}|}^*$  can be seen as having the use of a binary perfect channel  $k$  times, with  $k = \log_2 |\mathcal{X}|$ .

There are two simple extreme cases. First, if  $|\mathcal{X}| = 1$ , the capacity of the channel is 0 and the encoder cannot convey any information. Thus,  $C_e^{n,1}(\mathcal{P}_U, \mathcal{T}_{|\mathcal{X}|}^*) = C_e^*(\mathcal{P}_U, 0) = c_e^*(\mathcal{P}_U)$ . Second, if  $|\mathcal{X}| \geq |\mathcal{U}|^n$ , then the encoder can secure the unconstrained persuasion cost  $C_e^{n,1}(\mathcal{P}_U, \mathcal{T}_{|\mathcal{X}|}^*) = C_e^*(\mathcal{P}_U, \log |\mathcal{U}|) = \text{vex } c_e^*(\mathcal{P}_U)$  by treating each of the  $n$  problems separately and getting the cost  $\text{vex } c_e^*(\mathcal{P}_U)$  for each instance. The first point of Theorem 33 shows that this is the best possible cost.

More generally, Theorem 33 implies the following.

**Corollary 5** *Consider a persuasion problem repeated  $n$  times, where the encoder sends one signal from a set of cardinality  $|\mathcal{X}|$ . Then:*

1.  $C_e^{n,1}(\mathcal{P}_U, \mathcal{T}_{|\mathcal{X}|}^*) \leq C_e^*(\mathcal{P}_U, \frac{\log |\mathcal{X}|}{n})$ , for all  $n \in \mathbb{N}^*$ .
2. For any pair of sequences of integers  $(|\mathcal{X}_j|, n_j)_{j \in \mathbb{N}^*}$  such that  $\lim_{j \rightarrow \infty} \min(|\mathcal{X}_j|, n_j) = \infty$  and  $\lim_{j \rightarrow \infty} \frac{\log |\mathcal{X}_j|}{n_j} = C$ , we have  $\lim_{j \rightarrow \infty} C_e^{n_j,1}(\mathcal{P}_U, \mathcal{T}_{|\mathcal{X}_j|}^*) = C_e^*(\mathcal{P}_U, C)$ .

*Proof.* The first point follows directly from Theorem 33. To see the second point, it is enough to remark that a perfect channel  $\mathcal{T}_{|\mathcal{X}|}^*$  is “close” to  $k$  copies of a perfect binary channel  $\mathcal{T}_2^*$  with  $k$  such that  $2^k \leq |\mathcal{X}| < 2^{k+1}$ , that is  $k = \lfloor \log |\mathcal{X}| \rfloor$ . Having more signals at disposal is beneficial for the encoder and thus  $C_e^{n,1}(\mathcal{P}_U, \mathcal{T}_{|\mathcal{X}|}^*)$  is weakly increasing with  $|\mathcal{X}|$ . It follows that:

$$C_e^{n,k}(\mathcal{P}_U, \mathcal{T}_2^*) \leq C_e^{n,1}(\mathcal{P}_U, \mathcal{T}_{|\mathcal{X}|}^*) \leq C_e^{n,k+1}(\mathcal{P}_U, \mathcal{T}_2^*).$$

<sup>6</sup>See Cover and Thomas, 2006, Lemma 11.6.1, p. 370.

<sup>7</sup>See Cover and Thomas, 2006, p. 184.

Take a sequence  $(m_j, n_j)_{j \in \mathbb{N}^*}$  such that  $\lim_{j \rightarrow \infty} \min(m_j, n_j) = \infty$  and  $\lim_{j \rightarrow \infty} \frac{\log |\mathcal{X}_j|}{n_j} = C$ , and define  $k_j = \lfloor \log |\mathcal{X}_j| \rfloor$ . We have  $\lim_{j \rightarrow \infty} \min(k_j, n_j) = \infty$ ,  $\lim_{j \rightarrow \infty} \frac{k_j}{n_j} = C$  and the conclusion follows from Theorem 33.  $\square$

### VI.1.d Convexification with information constraint

In this section, we give some properties of the optimal splitting problem under information constraint. The motivation for this part of the results is two-fold. First, it is known that in a convexification problem, the number of posteriors (or of signals) can be chosen less than or equal to the number of states. One might wonder whether this remains true when there is a constraint on the feasible splittings. Second, models with costly information often use the mutual information as information cost (see e.g. Sims, 2003). We will see that in our case, this is derived by writing a Lagrangian for  $C_e^*(\mathcal{P}_U, C)$ .

Consider the optimal splitting under information constraint:

$$\inf \left\{ \sum_w \lambda_w c_e^*(\nu_w) : \sum_w \lambda_w \nu_w = \mathcal{P}_U, \sum_w \lambda_w H(\nu_w) \geq H(U) - C \right\}.$$

This is a special instance of the following optimization problem. Let  $f, g : \mathcal{S} \rightarrow \mathbb{R} \cup \{-\infty\}$  be two functions defined on a convex set  $\mathcal{S} \subseteq \mathbb{R}^d$ , where  $\mathcal{S}$  represents an abstract set of posteriors,  $f$  is a cost function and  $g$  is a constraint capturing the feasible splittings. For  $s \in \mathcal{S}$  and  $\gamma \in \mathbb{R}$  consider the problem:

$$F^g(s, \gamma) := \inf \left\{ \sum_w \lambda_w f(s_w) : \sum_w \lambda_w s_w = s, \sum_w \lambda_w g(s_w) \geq \gamma \right\}.$$

Let  $f^g : \mathcal{S} \times \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  defined by:

$$f^g(s, \gamma) = \begin{cases} f(s) & \text{if } \gamma \leq g(s), \\ +\infty & \text{otherwise.} \end{cases}$$

**Theorem 34** *Then, for each  $(s, \gamma) \in \mathcal{S} \times \mathbb{R}$ ,*

1.  $F^g(s, \gamma) = \text{vex } f^g(s, \gamma)$ .
2.  $F^g(s, \gamma) = \sup_{t \geq 0} \left\{ \text{vex}(f + tg)(s) - t\gamma \right\}$ .

The proof of Theorem 34 is stated in (Le Treust and Tomala, 2019, [J8]). Applying this result to the optimal splitting under information constraint, we get:

**Corollary 6** *For each  $\mathcal{P}_U \in \Delta(\mathcal{U})$  and  $C \geq 0$ ,*

1.  $C_e^*(\mathcal{P}_U, C)$  is the convexification of the function  $c_e^H : \Delta(\mathcal{U}) \times \mathbb{R} \rightarrow \mathbb{R}$  defined as:

$$c_e^H(\nu, \eta) = \begin{cases} c_e^*(\nu) & \text{if } \eta \leq H(\nu), \\ +\infty & \text{otherwise,} \end{cases}$$

calculated at  $(\nu, \eta) = (\mathcal{P}_U, H(U) - C)$ .

2.  $C_e^*(\mathcal{P}_U, C) = \sup_{t \geq 0} \left\{ \text{vex}(c_e^* + tH)(\mathcal{P}_U) - t(H(U) - C) \right\}$ .

Since it might be useful in other contexts, Theorem 34 is stated for general functions rather than specifically for the entropy function. This result has recently been generalized by Doval and Skreta (2018) to splitting problems with several constraints. The first point of the theorem states that the convexification with constraint, is the convexification of a bivariate function where an additional variable is added for the constraint (many variables when there are many constraints, see Doval and Skreta, 2018). The second point states that a Lagrangian function can be introduced and that the convexification under constraint is the convexification of the Lagrangian for some multiplier. The proof is in (Le Treust and Tomala, 2019, [J8], Appendix).

A direct implication of the second point of Corollary 6 is that there exists  $t^* = t^*(\mathcal{P}_U, \mathbf{C})$  such that:

$$C_e^*(\mathcal{P}_U, \mathbf{C}) = \text{vex}(c_e^* + t^*H)(\mathcal{P}_U) - t^*(H(U) - \mathbf{C}).$$

To see the existence of  $t^*$ , notice that  $\text{vex}(c_e^* + tH)(\mathcal{P}_U) - t(H(U) - \mathbf{C}) \geq (c_e^* + tH)(\mathcal{P}_U) - t(H(U) - \mathbf{C}) = c_e^*(\mathcal{P}_U) + t\mathbf{C}$ , which tends to  $+\infty$  as  $t \rightarrow +\infty$ . Therefore,  $t \mapsto \text{vex}(c_e^* + tH)(\mathcal{P}_U) - t(H(U) - \mathbf{C})$  reaches a minimum at some  $t^*$ .

If  $(\lambda_w^*, \nu_w^*)_w$  is an optimal splitting, let  $\mathcal{I}^* = H(U) - \sum_w \lambda_w^* H(\nu_w^*)$  be its mutual information. We have the following:

$$C_e^*(\mathcal{P}_U, \mathbf{C}) = \sum_w \lambda_w^* c_e^*(\nu_w^*) - t^*(\mathcal{I}^* - \mathbf{C}). \quad (\text{VI.18})$$

We then find the usual Kuhn-Tucker slackness conditions. If  $\mathcal{I}^* < \mathbf{C}$ , then  $t^* = 0$  and the unconstrained optimum is feasible. If  $t^* > 0$ , the constraint is binding. The Lagrange multiplier  $t^*$  can be interpreted as the *shadow price of capacity*, that is, the marginal value of an extra unit of communication capacity.

This characterization can be related with the *cost of information* considered in the literature on rational inattention (See [Sims, 2003](#)) where the agent pays a cost proportional to the mutual information between the state and the signal he observes. In particular, [Caplin and Dean \(2013\)](#) consider the convexification of a utility function net of such an information cost. For persuasion games, [Gentzkow and Kamenica \(2014\)](#) assume that the encoder pays a cost for choosing a disclosure strategy which is also related to the mutual information and also take the convexification of the net utility function.

Equation (VI.18) can be seen as a microfoundation of the use of mutual information as the information cost: the limit optimal value of persuasion for a large number of copies of problems with communication over an imperfect channel, has the same value as a problem of persuasion with an information cost. There are some differences, however. First, the information cost is not the mutual information, but the difference between the mutual information and the capacity of the channel. That is, a cost reduces the cost only when the encoder would like to send more information bits than the capacity. Second, the unit price of capacity is endogenous and given by the Lagrange multiplier of the information constraint.

A direct implication is an upper bound of the number of posteriors needed to achieve the convexification.

**Corollary 7** *In the optimization problem,*

$$C_e^*(\mathcal{P}_U, \mathbf{C}) = \inf \left\{ \sum_w \lambda_w c_e^*(\nu_w) : \sum_w \lambda_w \nu_w = \mathcal{P}_U, \sum_w \lambda_w H(\nu_w) \geq H(U) - \mathbf{C} \right\},$$

*the number of posteriors can be chosen to be at most  $\min\{|\mathcal{V}|, |\mathcal{U}| + 1\}$ .*

The proof of Corollary 7 is stated in ([Le Treust and Tomala, 2019](#), [J8]). Without the information constraint, the usual bound is  $\min\{|\mathcal{V}|, |\mathcal{U}|\}$ , i.e. the number of posteriors or of signals can be upper bounded by the number of actions and the number of states. For the number of actions, the argument is that two signals for which the decoder chooses the same action can be merged into one and the corresponding two posteriors replaced by the average. The argument still holds due to the concavity of the entropy function: replacing two posteriors by their average increases the expected entropy and thus helps in satisfying the information constraint.

For the bound given by the number of states, the usual technical argument is that any point in the convex hull of the epigraph of a function on  $\Delta(\mathcal{U})$  is a convex combination involving  $|\mathcal{U}|$  points. From Corollary 6, we consider the convexification of a function defined on  $\Delta(\mathcal{U}) \times \mathbb{R}$  a domain with one extra dimension; thus, an extra posterior might be needed. A similar observation is made in [Boleslavsky and Kim \(2018\)](#), where due to an incentive constraint, an extra posterior is needed. In Section VI.1.e, we provide an example where  $|\mathcal{U}| + 1$  posteriors are used at the optimum.

## VI.1.e Illustrating example

### Unrestricted communication

In this example, the encoder is a firm that persuades the decoder to invest in a risky project. If the decoder does not invest (action  $v_0$ ), the cost is 0 for both players. If the decoder invests (action  $v_1$ ), the

project has a cost 7 in the bad state  $u_0$  and  $-1$  in the good state  $u_1$ . Both states are equally likely. the encoder obtains the cost  $-1$  only if the decoder invests. The cost table is as follows, the entries are pairs of costs for the players  $i = e, d$  depending on the state and action.

	$v_0$	$v_1$	$\mathcal{P}_U$
$u_0$	0, 0	-1, 7	$\frac{1}{2}$
$u_1$	0, 0	-1, -1	$\frac{1}{2}$

The decoder invests for sure only when he holds a belief  $\nu$  such that  $\nu(u_1) > 7/8$ . If  $\nu(u_1) = 7/8$  he is indifferent. Assuming that in case of indifference he does not invest, the robust cost of the encoder  $c_e^*(\nu)$  is 0 if  $\nu(u_1) > 7/8$  and  $-1$  otherwise.

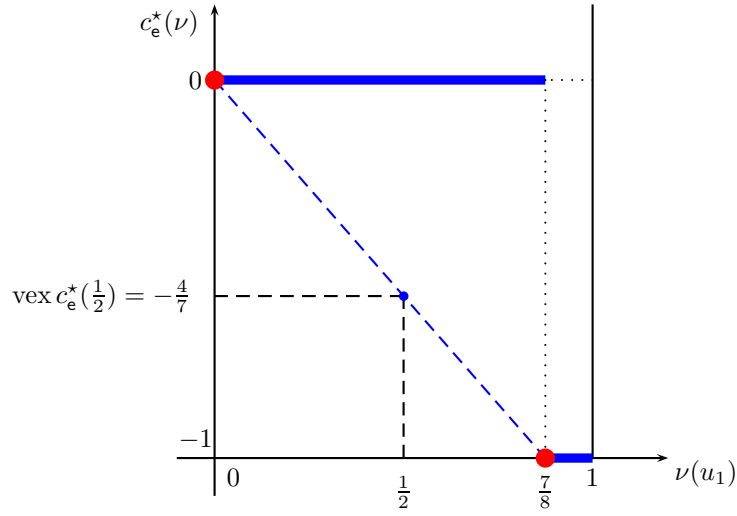


Figure VI.3 – Convexification.

The convexification function  $\text{vex } c_e^*(\nu)$  is continuous and equal to  $-\frac{8}{7}\nu(u_1)$  for  $\nu(u_1) \leq \frac{7}{8}$  and 0 otherwise. It is easy to see that it does not depend on the action chosen by the decoder at  $\nu(u_1) = \frac{7}{8}$ , see Fig. VI.3. If the decoder were to choose  $v_1$  at the point of indifference, then the optimal splitting for the encoder would be as follows:

$$\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{3}{7}(1, 0) + \frac{4}{7}\left(\frac{1}{8}, \frac{7}{8}\right),$$

where a belief is denoted  $\nu = (\nu(u_0), \nu(u_1))$ . This yields a cost of  $\frac{3}{7}$  which is the lowest that the encoder can achieve given the uniform prior. For any small  $\varepsilon > 0$ , we can perturb the previous splitting and get the following:

$$\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{3+8\varepsilon}{7+8\varepsilon}(1, 0) + \frac{4}{7+8\varepsilon}\left(\frac{1}{8}-\varepsilon, \frac{7}{8}+\varepsilon\right),$$

which achieves the cost  $-\frac{4\varepsilon}{7+8\varepsilon}$  irrespective of the tie-breaking rule. Letting  $\varepsilon$  tend 0, we see that the encoder achieves a cost arbitrarily close to  $-\frac{4}{7}$ , which is the optimal robust cost.

### Restricted and noisy communication

We consider binary sets of signals  $\mathcal{X} = \{x_0, x_1\}$ ,  $\mathcal{Y} = \{y_0, y_1\}$  and we assume that the channel has a noise level  $\varepsilon \in [0, \frac{1}{2}]$ , that is  $\mathcal{T}_{Y|X}(y_j|x_i) = \varepsilon$  for  $j \neq i$ , see Fig. VI.4. The generic case is  $\varepsilon \in ]0, \frac{1}{2}[$  where the label of the signal (0 or 1) is changed with positive probability but observing a label 1 is still more

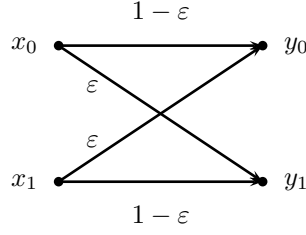


Figure VI.4 – Binary symmetric channel.

likely when the input label is 1. When  $\varepsilon = \frac{1}{2}$ , the distribution of the output signal is independent from the input signal, so the channel completely disrupts the communication.

A special case is the *binary perfect channel* when  $\varepsilon = 0$ : identifying together the sets  $\mathcal{X}$  and  $\mathcal{Y}$ , an input signal  $x$  is received with certainty. Communication is then restricted only by the number of available signals, i.e. the cardinality of  $\mathcal{X}$ .

The capacity of the binary symmetric channel<sup>8</sup> is  $1 - H(\varepsilon)$  where the entropy of the binary probability distribution  $(\varepsilon, 1 - \varepsilon)$  is defined by  $H(\varepsilon) = \varepsilon \log_2 \frac{1}{\varepsilon} + (1 - \varepsilon) \log_2 \frac{1}{1 - \varepsilon}$  for  $\varepsilon \in ]0, 1[$ , and  $H(\varepsilon) = 0$  for  $\varepsilon \in \{0, 1\}$ .

#### One-shot scenario $k = n = 1$

Let a strategy  $\sigma$  of the encoder be parametrized by  $\sigma(x_0|u_0) = 1 - \alpha$  and  $\sigma(x_1|u_1) = 1 - \beta$ ; see Fig. VI.5.

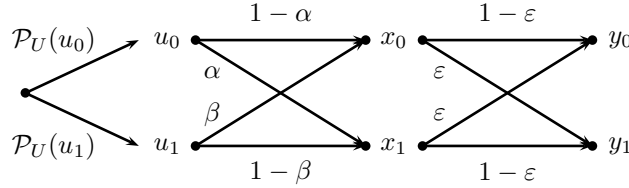


Figure VI.5 – Strategy on the binary symmetric channel.

Then,  $\Pr_\sigma(y_1|u_0) = \alpha(1 - \varepsilon) + (1 - \alpha)\varepsilon$ ,  $\Pr_\sigma(y_0|u_1) = \beta(1 - \varepsilon) + (1 - \beta)\varepsilon$  and from Bayes' rule,

$$\Pr_\sigma(u_1|y_1) = \frac{\mathcal{P}_U(u_1)(1 - \Pr_\sigma(y_0|u_1))}{\mathcal{P}_U(u_0)\Pr_\sigma(y_1|u_0) + \mathcal{P}_U(u_1)(1 - \Pr_\sigma(y_0|u_1))},$$

$$\Pr_\sigma(u_1|y_0) = \frac{\mathcal{P}_U(u_1)\Pr_\sigma(y_0|u_1)}{\mathcal{P}_U(u_0)(1 - \Pr_\sigma(y_1|u_0)) + \mathcal{P}_U(u_1)\Pr_\sigma(y_0|u_1)}.$$

It is easy to see that the numbers  $\Pr_\sigma(y_1|u_0)$ ,  $\Pr_\sigma(y_0|u_1)$ ,  $\Pr_\sigma(u_1|y_1)$ ,  $\Pr_\sigma(u_1|y_0)$  all belong to the interval  $[\varepsilon, 1 - \varepsilon]$ .

A pair of posteriors  $(\nu_0, \nu_1)$  is said to be *feasible* in the one-shot scenario if there exists a number  $\lambda \in [0, 1]$  such that:

$$(\mathcal{P}_U(u_0), \mathcal{P}_U(u_1)) = \lambda(\nu_0(u_0), \nu_0(u_1)) + (1 - \lambda)(\nu_1(u_0), \nu_1(u_1)).$$

The feasible splittings can be characterized as follows.

<sup>8</sup>Cover and Thomas, 2006, Example 2.1.1, p. 15.

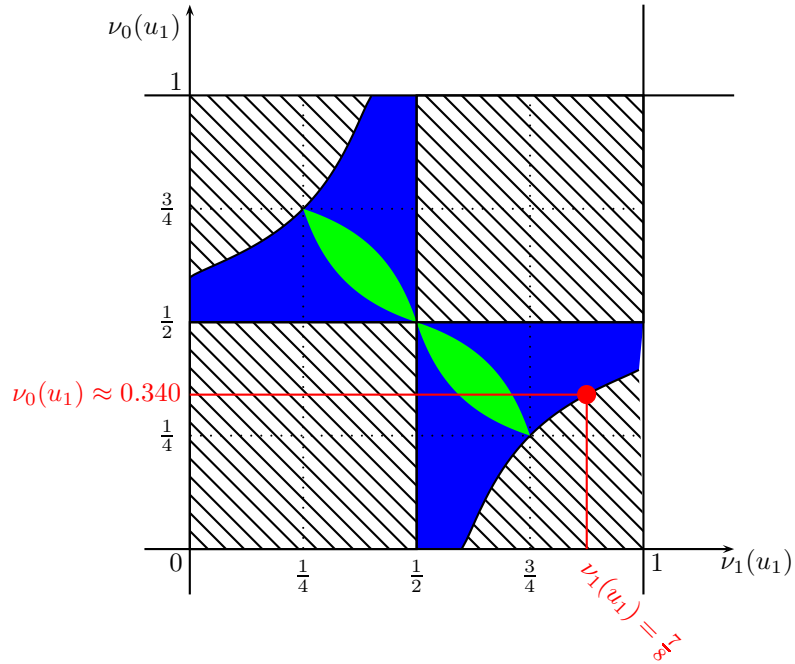


Figure VI.6 – For a noise parameter  $\varepsilon = \frac{1}{4}$ , the green lenses correspond to the feasible posteriors  $(\nu_0, \nu_1)$  characterized in Lemma 5 for the one-shot scenario  $k = n = 1$ . The blue and green regions correspond to the feasible posteriors  $(\nu_0, \nu_1)$  in the asymptotic scenario where  $k = n \rightarrow \infty$ . The red point corresponds to the optimal splitting, also depicted in Fig. VI.7. The hatched areas correspond to the nonfeasible posteriors  $(\nu_0, \nu_1)$ .

**Lemma 5** *We consider the one-shot problem where  $n = k = 1$ . A pair of posteriors  $(\nu_0, \nu_1)$  is feasible if and only if  $\nu_1 = \nu_0 = \mathcal{P}_U$  or,*

$$\varepsilon \leq \frac{\nu_0(u_1)(\nu_1(u_1) - \mathcal{P}_U(u_1))}{\mathcal{P}_U(u_1)(\nu_1(u_1) - \nu_0(u_1))} \leq 1 - \varepsilon$$

and

$$\varepsilon \leq \frac{(1 - \nu_0(u_1))(\mathcal{P}_U(u_1) - \nu_0(u_1))}{(1 - \mathcal{P}_U(u_1))(\nu_1(u_1) - \nu_0(u_1))} \leq 1 - \varepsilon.$$

The proof is in (Le Treust and Tomala, 2019, [J8], Appendix). As an illustration, take the uniform prior  $(\frac{1}{2}, \frac{1}{2})$  and a level of noise  $\varepsilon = \frac{1}{4}$ . The feasible posteriors are shown by the colored green regions on Fig. VI.6.

From the previous discussion, it is impossible to induce beliefs with  $\nu(u_1) > \frac{3}{4}$ . Therefore, the decoder will never be confident enough to invest and the cost is 0 for the encoder.

#### Asymptotic scenario with $k = n \rightarrow \infty$

We consider the case where  $k = n$  tends to infinity with a noise level of  $\varepsilon = \frac{1}{4}$  and compute the value of the optimal splitting problem with information constraint. The capacity of the channel is  $1 - H(\frac{1}{4})$ , the entropy of the uniform prior is 1; therefore, the information constraint is  $\sum_w \lambda_w H(\nu_w) \geq H(\frac{1}{4})$ . Under this constraint the optimal splitting for the encoder satisfies:

$$\left(\frac{1}{2}, \frac{1}{2}\right) = \lambda \left(\frac{1}{8}, \frac{7}{8}\right) + (1 - \lambda)(\nu_0(u_0), \nu_0(u_1))$$

and

$$H\left(\frac{1}{4}\right) = \lambda H\left(\frac{7}{8}\right) + (1 - \lambda)H(\nu_0(u_1)).$$

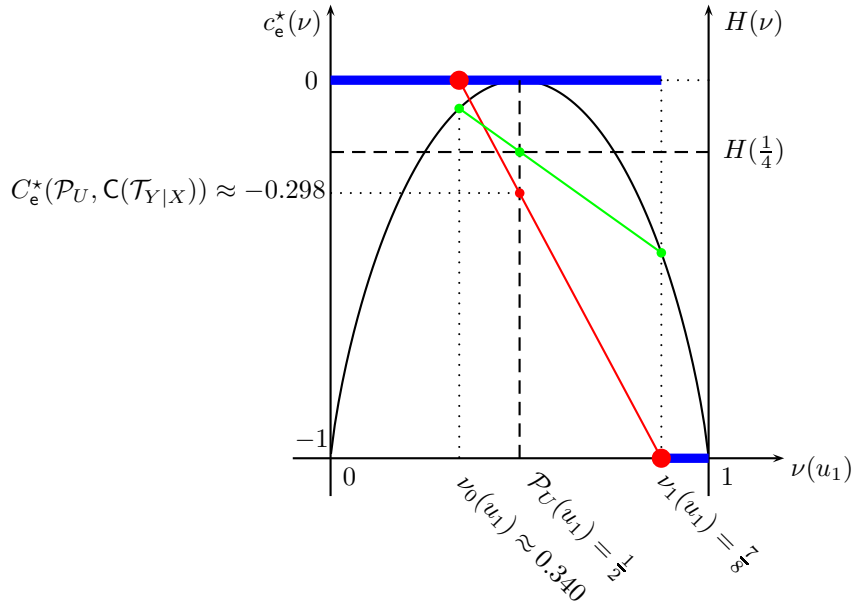


Figure VI.7 – For a noise parameter  $\varepsilon = \frac{1}{4}$ , the optimal splitting is given by  $\nu_0(u_1) \approx 0.340$  and  $\nu_1(u_1) = \frac{7}{8}$ .

To see why it is optimal, first observe that the encoder has to bring on some posterior, denoted by  $\nu_1$ , with  $\nu_1(u_1) > \frac{7}{8}$  in order to get some cost. To get it with the highest probability, he should aim for  $\nu_1(u_1) = \frac{7}{8}$ . Among the posteriors that induce investment, this is also the one with highest entropy. Second, to minimize expected costs, the remaining posteriors must be as far away as possible from the prior; that is, the information constraint should bind. Additionally, note that only one posterior, denoted by  $\nu_0$ , will be optimally generated in the region  $\nu_0(u_1) < \frac{7}{8}$ . Since the entropy is strictly concave, replacing two posteriors on this region by their average does not change the cost and increases the entropy.

Solving these two equations numerically we get,  $\nu_0(u_1) \approx 0.340$  and  $C_e^*(\mathcal{P}_U, \mathcal{T}_{Y|X}) = -\lambda \approx -0.298$  instead of the cost 0 for the one-shot scenario and about 52.1% of the unconstrained optimum  $-\frac{4}{7}$ .

This is shown in Fig. VI.7 which plots the cost function and the entropy function. The splitting of  $\mathcal{P}_U$  into  $\nu_0, \nu_1$  is shown by the three points on the horizontal axis. On the vertical line  $\mathcal{P}_U(u_1) = \frac{1}{2}$ , we can read the average cost with the red line and the average entropy with the green line. To see optimality on the picture, if we move  $\nu_0(u_1)$  to the right, then the average cost decrease, and if we move it to the left, the average entropy will fall below  $H(\frac{1}{4})$  and the information constraint will be violated.

The optimal splitting is also marked on Fig. VI.6 which shows the set of pairs of posteriors for the splittings that satisfy the information constraint (union of green and blue regions).

On Fig. VI.8, we represent the value  $C_e^*(\mathcal{P}_U, \mathcal{C}(\mathcal{T}_{Y|X}))$  of the optimal splitting problem as a function of the prior  $\mathcal{P}_U$ , for different values for the noise parameter  $\varepsilon \in \left\{ \frac{1}{20}, \frac{3}{20}, \frac{1}{4}, \frac{7}{20}, \frac{9}{20}, \frac{99}{200} \right\}$ . It is found by solving the following system for  $\nu_0$ :

$$(\mathcal{P}_U(u_0), \mathcal{P}_U(u_1)) = \lambda \left( \frac{1}{8}, \frac{7}{8} \right) + (1 - \lambda)(\nu_0(u_0), \nu_0(u_1))$$

and

$$H(\mathcal{P}_U(u_1)) - 1 + H(\varepsilon) = \lambda H\left(\frac{7}{8}\right) + (1 - \lambda)H(\nu_0(u_1)).$$

When  $\mathcal{P}_U(u_1) = \frac{1}{2}$  and  $\varepsilon = \frac{1}{4}$ , we recover the value  $C_e^*(\mathcal{P}_U, \mathcal{C}(\mathcal{T}_{Y|X})) \approx -0.298$  as in Fig. VI.7.

Observe that the function  $C_e^*(\mathcal{P}_U, \mathcal{C}(\mathcal{T}_{Y|X}))$  is *not* concave with respect to the prior  $\mathcal{P}_U$ . From Corollary 6,  $C_e^*(\mathcal{P}_U, \mathcal{C})$  is the convexification of the function  $c_e^H$  calculated at  $(\mathcal{P}_U, H(U) - \mathcal{C})$ , so this composed function need not be concave.

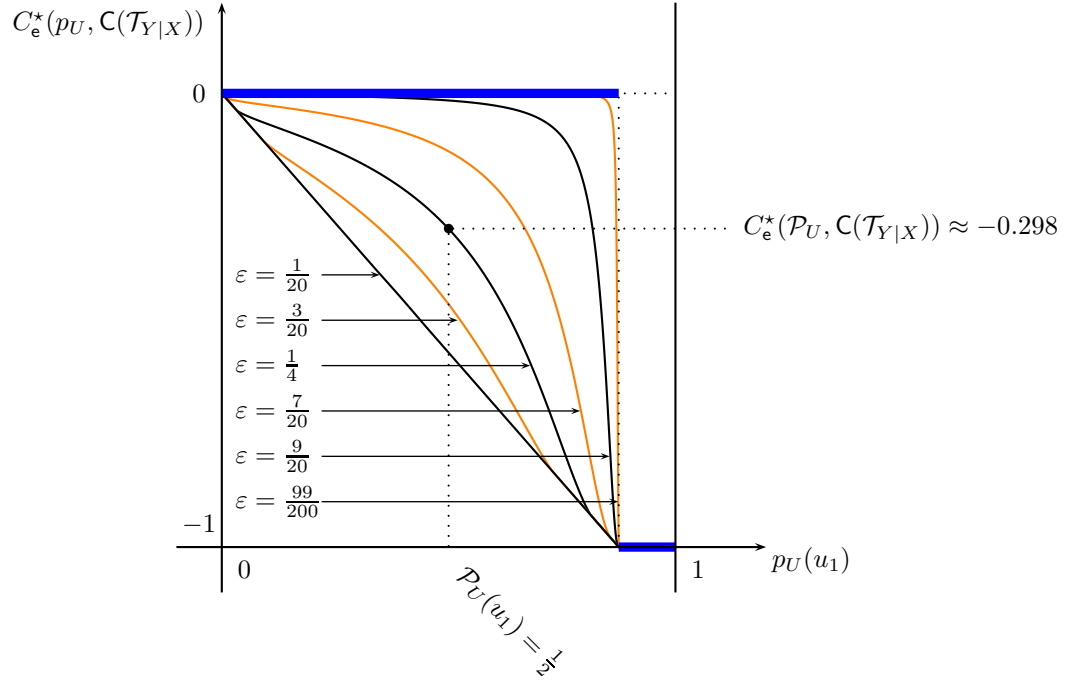


Figure VI.8 – Value of the optimal splitting problem as a function of the prior  $\mathcal{P}_U$ , for different noise parameters  $\varepsilon \in \left\{ \frac{1}{20}, \frac{3}{20}, \frac{1}{4}, \frac{7}{20}, \frac{9}{20}, \frac{99}{200} \right\}$ .

### Perfect binary channel with $k = \frac{n}{2} \rightarrow \infty$

We consider the same example as before, repeated  $n$  times with the uniform prior  $\mathcal{P}_U = (\frac{1}{2}, \frac{1}{2})$ . In line with the motivating example from the introduction, we consider a perfect channel and assume that the encoder has at its disposal half as many signals as needed to communicate perfectly, that is  $k = \frac{n}{2}$ . Since the capacity of the binary perfect channel is one,  $\frac{k}{n} = \frac{1}{2}$  and  $H(U) = 1$ , the information constraint is:

$$H(U) - \sum_w \lambda_w H(\nu_w) \leq \frac{1}{2} \iff \sum_w \lambda_w H(\nu_w) \geq \frac{1}{2}.$$

Observe that this constraint is identical to the one obtained with a binary symmetric channel with noise  $\varepsilon$  such that  $H(\varepsilon) = \frac{1}{2}$  (i.e.,  $\varepsilon \approx 0.110$ ). Therefore, the optimal splitting is given by the following system:

$$(\mathcal{P}_U(u_0), \mathcal{P}_U(u_1)) = \lambda \left( \frac{1}{8}, \frac{7}{8} \right) + (1 - \lambda)(\nu_0(u_0), \nu_0(u_1))$$

and

$$\frac{1}{2} = \lambda H\left(\frac{7}{8}\right) + (1 - \lambda)H(\nu_0(u_1)).$$

Solving numerically, we find  $C_e^*(\mathcal{P}_U, \frac{1}{2}) \approx -0.519$ .

### On the number of posteriors

We give now an example showing the tightness of the bound  $\min\{|\mathcal{V}|, |\mathcal{U}| + 1\}$  on the number of posteriors, given in Corollary 7. The cost table is as follows:

	$v_0$	$v_1$	$v_2$	
$u_0$	0, 0	-1, 7	-1, -1	$\frac{1}{2}$
$u_1$	0, 0	-1, -1	-1, 7	



There are two risky projects ( $v_1$  and  $v_2$ ) and the encoder wants to persuade the decoder to invest in any of them. The decoder invests only if  $\nu(u_1) > 7/8$  or  $\nu(u_1) < 1/8$ .

With unrestricted communication, the solution is clear: the encoder fully discloses the state and gets a cost of -1. However, with a binary symmetric channel with noise  $\varepsilon = 1/4$ , the encoder gets 0 in the one-shot scenario. Consider now the case where  $k = n \rightarrow \infty$ .

The “one-sided” solution of Section VI.1.e is feasible. Recall that this is the splitting such that:

$$\left(\frac{1}{2}, \frac{1}{2}\right) = \lambda \left(\frac{1}{8}, \frac{7}{8}\right) + (1 - \lambda)(\nu_0(u_0), \nu_0(u_1))$$

and

$$H\left(\frac{1}{4}\right) = \lambda H\left(\frac{7}{8}\right) + (1 - \lambda)H(\nu_0(u_0), \nu_0(u_1)).$$

with  $\nu_0(u_1) \approx 0.340$  and  $\lambda \approx 0.298$ . It is easy to see that this is optimal among the splittings with two posteriors. Indeed, it is not possible that the two posteriors induce investment while satisfying the information constraint.

However, this is not optimal. The optimal splitting has three posteriors and is as follows:

$$\left(\frac{1}{2}, \frac{1}{2}\right) = (1 - \lambda) \left(\frac{1}{2}, \frac{1}{2}\right) + \frac{\lambda}{2} \left(\frac{1}{8}, \frac{7}{8}\right) + \frac{\lambda}{2} \left(\frac{7}{8}, \frac{1}{8}\right)$$

with

$$H\left(\frac{1}{4}\right) = (1 - \lambda)H\left(\frac{1}{2}\right) + \frac{\lambda}{2} H\left(\frac{1}{8}\right) + \frac{\lambda}{2} H\left(\frac{7}{8}\right).$$

This pins down a unique  $\lambda$  and solving numerically yields  $\lambda \approx 0.413$ . Since  $\lambda$  is the probability of investment, we get  $C_e^*(\mathcal{P}_U, \mathcal{T}_{Y|X}) \approx -0.413$  which is about 38% better than what is achieved with a splitting with two points.

To see that this is optimal, first since there are two states, we know that three posteriors are sufficient. Second, it is not possible to have all posteriors in the investment region and to satisfy the information constraint. If there is only one posterior in the investment region, then the splitting achieves no more than the “one-sided” solution. Therefore, it is optimal to have two posteriors in the investment region and one outside of it. However, then, it is optimal to choose the point in the middle region to be  $(\frac{1}{2}, \frac{1}{2})$ , since this is the one with the highest entropy.

### VI.1.f Beyond identical problems

The main result can be extended to series of persuasion problems which are not all identical, but such that each type of problem is repeated many times. Suppose that we have a family of persuasion problems indexed by a type parameter  $z$  in a finite set  $\mathcal{Z}$ . That is, for every  $z \in \mathcal{Z}$ , there is a prior probability distribution  $\mathcal{P}_{U|Z}(\cdot|z) \in \Delta(\mathcal{U})$  and cost functions  $c_i : \mathcal{U} \times \mathcal{V} \times \mathcal{Z} \rightarrow \mathbb{R}$  for each player  $i = e, d$ . The series of persuasion problems is given by a sequence  $z^n = (z_1, \dots, z_n)$  which is commonly known by both players. The distribution of states is as follows:

$$\mathcal{P}_U^n(u^n|z^n) := \prod_{t=1}^n \mathcal{P}_U(u_t|z_t).$$

If the sequence of states and actions are respectively  $u^n, v^n$ , the cost for player  $i$  is  $\frac{1}{n} \sum_{t=1}^n c_i(u_t, v_t, z_t)$ . The communication technology is still given by a channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$  used  $k$  times, so that the strategy sets are the same as before for both players. The optimal robust cost of the encoder is defined as before and is denoted by  $C_e^{n,k}(\mathcal{P}_U, \mathcal{T}_{Y|X}, z^n)$ .

For each posterior belief  $\nu \in \Delta(\mathcal{U})$  and type  $z \in \mathcal{Z}$ , the set of optimal actions of the decoder is  $\mathcal{V}^{*z}(\nu) = \operatorname{argmin}_{v \in \mathcal{V}} \sum_u \nu(u) c_d(u, v, z)$  and we denote by  $c_e^z(\nu) = \min_{v \in \mathcal{V}^{*z}(\nu)} \sum_u \nu(u) c_e(u, v, z)$  the robust cost of the encoder at the belief  $\nu$ .

**Definition 47** For  $\pi \in \Delta(\mathcal{Z})$  and  $C \geq 0$ , the optimal splitting problem with information constraint is as follows:

$$\begin{aligned} C_e^Z(\mathcal{P}_{U|Z}, C, \pi) &= \inf \sum_z \pi(z) \sum_m \lambda_w^z c_e^z(\nu_w^z) \\ \text{s.t.} \quad &\sum_w \lambda_w^z \nu_w^z = \mathcal{P}_{U|Z}(\cdot|z), \quad \forall z \in Z, \\ \text{and} \quad &\sum_z \pi(z) \left( H(\mathcal{P}_{U|Z}(\cdot|z)) - \sum_w \lambda_w^z H(\nu_w^z) \right) \leq C. \end{aligned}$$

The interpretation is as follows. Suppose that  $\pi(z)$  represents the probability, or frequency, of occurrence of  $z$ . Conditional on  $z$  which is known by both players, the encoder performs a spitting of  $\mathcal{P}_{U|Z}(\cdot|z)$ ,  $\sum_w \lambda_w^z \nu_w^z = \mathcal{P}_{U|Z}(\cdot|z)$ , and gets the cost  $\sum_m \lambda_w^z c_e^z(\nu_w^z)$ . The information constraint imposes the *conditional* mutual information  $I(U; W|Z)$  to be less than or equal to the capacity  $C$ .

Given a sequence  $z^n \in Z^n$ , let  $\pi_n \in \Delta(\mathcal{Z})$  be the empirical frequency induced by the sequence: for each  $z \in \mathcal{Z}$ ,  $\pi_n(z) = \frac{1}{n} |\{t : z_t = z\}|$ .

**Theorem 35** 1. The optimal robust cost of the encoder is no more than the value of the optimal splitting problem with information constraint. For each pair of integers  $n, k$ :

$$C_e^{m,k}(\mathcal{P}_U, \mathcal{T}_{Y|X}, z^n) \leq C_e^Z(\mathcal{P}_{U|Z}, \frac{k}{n} C(\mathcal{T}_{Y|X}), \pi_n).$$

2. The optimal robust cost of the encoder converges to the value of the optimal splitting problem with information constraint in the following sense. For each  $\pi \in \Delta(\mathcal{Z})$  and  $r \in [0, +\infty]$ , for each pair of sequences of integers  $(k_j, n_j)_{j \in \mathbb{N}^*}$  such that  $\lim_{j \rightarrow \infty} \min(n_j, k_j) = \infty$ ,  $\lim_{j \rightarrow \infty} \frac{k_j}{n_j} = r$  and  $\lim_{j \rightarrow \infty} \pi_{n_j} = \pi$ , we have:

$$\lim_{j \rightarrow \infty} C_e^{n_j, k_j}(\mathcal{P}_U, \mathcal{T}_{Y|X}, z^{n_j}) = C_e^Z(\mathcal{P}_{U|Z}, rC(\mathcal{T}_{Y|X}), \pi).$$

Given a sequence  $z^n$ ,  $\pi_n$  is the empirical distribution of types of problems. The optimal cost of the encoder is bounded above by the value of optimal splitting under information constraint. Suppose that the distribution of types is held fixed (or converges to)  $\pi$ , then when  $n$  and  $k$  grow large, the encoder is able to secure approximately this value.

This extension applies to the case where the proportions of types of problems are fixed. Alternatively, the sequence  $z^n$  could be drawn i.i.d. from a prior distribution  $\pi \in \Delta(\mathcal{Z})$ .

Notice the channel  $\mathcal{T}_{Y|X}^k$  is used for transferring information about all problems. Thus, Theorem 35 does more than merely patching up distinct families of problems together. The capacity of the channel bounds the total amount of information, across all problems. Thus, all problems, even of different types, are linked together in the signals.

We have analyzed a persuasion game where the encoder communicates with the decoder through a fixed and imperfect channel. The optimal cost of the encoder is bounded above by the value of the optimal splitting problem with information constraint. When the encoder and the decoder are engaged in many repetitions of identical persuasion games, the optimal cost for the encoder converges to the upper bound as the number of repetitions increases.

## VI.2 Persuasion game with side information at the decoder

In the previous section, it is assumed that the information about the state is fully controlled by the encoder. To model private information of the decoder, we consider that the nature draw pairs  $(u_t, z_t)_{t=1, \dots, n}$  and assume that  $u^n$  is the private information of the encoder and  $z^n$  is the private information of the decoder, as depicted in Fig. VI.9. Our methods generalize to this case provided that we use the correct information constraint. A random signal  $W$  can be transmitted over the channel provided that its mutual

information with the state, *conditional on the private information of the decoder*  $I(U;W|Z) \leq C$  is less than or equal to the capacity, where

$$I(U;W|Z) := \sum_z \Pr(z)I(U;W|Z=z)$$

is the expectation over  $Z$  of the mutual information conditional on  $\{Z=z\}$ .

Communication between autonomous devices that have distinct objectives is under study. This problem, referred to as the strategic communication problem, is at the crossroads of different disciplines such as Control Theory [Sarıtaş et al. \(2017\)](#), [Sarıtaş et al. \(2020\)](#), Computer Science [Dughmi et al. \(2019\)](#) and Information Theory ([Le Treust and Tomala, 2016, \[IC7\]](#)), ([Le Treust and Tomala, 2018b, \[IC20\]](#)), ([Le Treust and Tomala, 2018a, \[W2\]](#)), [Sarıtaş et al. \(2019\)](#), [Vora and Kulkarni \(2020\)](#), [Vora and Kulkarni \(2020\)](#), [Kazikh et al. \(2021\)](#), where it was introduced by Akyol et al. in [Akyol et al. \(2016\)](#), [Akyol et al. \(2017\)](#).

Three different formulations of the strategic communication problem are originally proposed in the Game Theory literature, see ([Le Treust and Tomala, 2020, \[IC26\]](#)). The cheap talk game of Crawford and Sobel [Crawford and Sobel \(1982\)](#) relies on the Nash equilibrium solution. In the mechanism design problem of Jackson and Sonnenschein [Jackson and Sonnenschein \(2007\)](#) the receiver commits to a prescribed decoding strategy, as the leader of a Stackelberg game. The hypothesis of decoder commitment is also related to the mismatched rate-cost problem in [Lapidoth \(1997\)](#), [Scarlett et al. \(2020\)](#). In the persuasion game of [Kamenica and Gentzkow \(2011\)](#), it is the sender who commits to a strategy whereas the decoder computes its Bayesian posterior belief and selects the optimal output symbol. In ([Le Treust and Tomala, 2019, \[J8\]](#)), we characterize the impact of the channel noise in the solution to the persuasion problem.

In this section, we extend these previous results with encoder commitment by considering that the decoder has side information. More specifically, we formulate a joint source-channel coding problem with decoder side information in which the encoder and the decoder are endowed with distinct cost functions. Given an encoder strategy, the decoder selects an optimal strategy for its cost function. The encoder anticipates the mismatch of the cost functions and commits to implementing the encoding strategy that minimizes its cost.

The technical novelty consists in controlling the distance of the posterior beliefs induced by Wyner-Ziv's coding to the target posterior beliefs. This demonstrates that the Wyner-Ziv's encoding reveals nothing but the exact amount of information needed to implement the optimal decoding strategy. Consequently at the optimum the decoder produces a sequence of outputs which is almost the same as the one generated by the coding of [Wyner and Ziv \(1976\)](#), for a specific probability distribution.

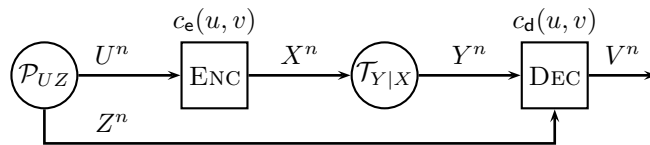


Figure VI.9 – The source  $\mathcal{P}_{UZ}$  is i.i.d., the channel  $\mathcal{T}_{Y|X}$  is memoryless. The encoder and the decoder have arbitrary distinct cost functions  $c_e(u, v) \neq c_d(u, v)$ .

## VI.2.a System model

We denote by  $\mathcal{U}$ ,  $\mathcal{Z}$ ,  $\mathcal{X}$ ,  $\mathcal{Y}$ ,  $\mathcal{V}$ , the finite sets of information source, side information, channel inputs, channel outputs and decoder's outputs. We consider the information source  $(\mathcal{U} \times \mathcal{Z}, \mathcal{P}_{UZ})$  and the channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ , as depicted in Fig. [VI.9](#).

**Definition 48** We define the encoding and the decoding strategies by

$$\sigma : \mathcal{U}^n \longrightarrow \Delta(\mathcal{X}^n), \quad (\text{VI.19})$$

$$\tau : \mathcal{Y}^n \times \mathcal{Z}^n \longrightarrow \Delta(\mathcal{V}^n). \quad (\text{VI.20})$$

The strategies  $(\sigma, \tau)$  induces the distribution

$$\mathcal{P}_{U^n Z^n X^n Y^n V^n}^{\sigma, \tau} = \left( \prod_{t=1}^n \mathcal{P}_{U_t Z_t} \right) \sigma_{X^n | U^n} \left( \prod_{t=1}^n \mathcal{T}_{Y_t | X_t} \right) \tau_{V^n | Y^n Z^n}, \quad (\text{VI.21})$$

where  $\sigma_{X^n | U^n}$ ,  $\tau_{V^n | Y^n Z^n}$  denote the distributions of the strategies  $\sigma$ ,  $\tau$ .

**Definition 49** The encoder and decoder are endowed with distinct cost functions

$$c_e : \mathcal{U} \times \mathcal{V} \longrightarrow \mathbb{R}, \quad c_d : \mathcal{U} \times \mathcal{V} \longrightarrow \mathbb{R}. \quad (\text{VI.22})$$

The long-run cost functions  $c_e^n(\sigma, \tau)$ ,  $c_d^n(\sigma, \tau)$  are defined by

$$c_e^n(\sigma, \tau) = \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n c_e(u_t, v_t) \right],$$

$$c_d^n(\sigma, \tau) = \sum_{u^n, v^n} \mathcal{P}^{\sigma, \tau}(u^n, v^n) \cdot \left( \frac{1}{n} \sum_{t=1}^n c_d(u_t, v_t) \right).$$

**Definition 50** Given  $n \in \mathbb{N}^*$ , we define:

1. the set of decoder best responses to strategy  $\sigma$  by

$$BR_d(\sigma) = \underset{\tau}{\operatorname{argmin}} c_d^n(\sigma, \tau), \quad (\text{VI.23})$$

2. the long-run encoder cost value by

$$C_e^n = \inf_{\sigma} \max_{\tau \in BR_d(\sigma)} c_e^n(\sigma, \tau). \quad (\text{VI.24})$$

In case  $BR_d(\sigma)$  is not a singleton, we assume that the decoder selects the worst strategy for the encoder cost  $\max_{\tau \in BR_d(\sigma)} c_e^n(\sigma, \tau)$ , so that the solution is robust to the exact specification of the decoding strategy. We aim at characterizing the asymptotic behavior of  $C_e^n$ .

**Definition 51** We consider an auxiliary random variable  $W \in \mathcal{W}$  with  $|\mathcal{W}| = \min(|\mathcal{U}| + 1, |\mathcal{V}|^{|\mathcal{Z}|})$  and we define

$$\mathbb{Q} = \left\{ \mathcal{P}_{UZ} \mathcal{Q}_{W|U}, \max_{\mathcal{P}_X} I(X; Y) - I(U; W|Z) \geq 0 \right\}. \quad (\text{VI.25})$$

Given  $\mathcal{Q}_{UZW}$ , we define the single-letter best responses

$$\mathbb{A}_d(\mathcal{Q}_{UZW}) = \underset{\mathcal{Q}_{V|ZW}}{\operatorname{argmin}} \mathbb{E}_{\mathcal{Q}_{UZW}} \left[ c_d(U, V) \right]. \quad (\text{VI.26})$$

The encoder optimal cost  $C_e^*$  is given by

$$C_e^* = \inf_{\mathcal{Q}_{UZW} \in \mathbb{Q}} \max_{\substack{\mathcal{Q}_{V|ZW} \in \\ \mathbb{A}_d(\mathcal{Q}_{UZW})}} \mathbb{E}_{\mathcal{Q}_{UZW}} \left[ c_e(U, V) \right]. \quad (\text{VI.27})$$

If  $\mathbb{A}_d(\mathcal{Q}_{UZW})$  is not a singleton, the decoder selects the worst distribution  $\mathcal{Q}_{V|ZW}$  from the encoder perspective.

**Theorem 36**

$$\forall n \in \mathbb{N}^*, \quad C_e^n \geq C_e^*, \quad (\text{VI.28})$$

$$\forall \varepsilon > 0, \exists \bar{n} \in \mathbb{N}^*, \forall n \geq \bar{n}, \quad C_e^n \leq C_e^* + \varepsilon. \quad (\text{VI.29})$$

The proof of Theorem 36 is stated in (Le Treust and Tomala, 2019, [W3], App. B and C).

*Sketch of proof of (VI.28).* For all  $n \in \mathbb{N}^*$ , the converse result relies on the identification of the auxiliary random variables  $W = (Y^n, Z^{T-1}, Z_{T+1}^n, T)$  and  $(U, Z, V) = (U_T, Z_T, V_T)$ , where  $T$  is uniformly distributed over  $\{1, \dots, n\}$ . We denote by  $\mathcal{Q}_{UZW}^\sigma, \mathcal{Q}_{V|ZW}^\tau$  the distributions induced by  $(\sigma, \tau)$  over  $(U, Z, W, V)$ . In Merhav and Shamai (2003), it is proved that the Markov chain  $Z \text{---} U \text{---} W$  holds and that  $I(U; W|Z) \leq \max_{\mathcal{P}_X} I(X; Y)$ , hence  $\mathcal{Q}_{UZW}^\sigma \in \mathcal{Q}$ . We show that

$$c_e^n(\sigma, \tau) = \mathbb{E}_{\substack{\mathcal{Q}_{UZW}^\sigma \\ \mathcal{Q}_{V|ZW}^\tau}} [c_e(U, V)], \quad (\text{VI.30})$$

$$\{\mathcal{P}_{V|ZW}, \exists \tau \in \text{BR}_d(\sigma), \mathcal{Q}_{V|ZW}^\tau = \mathcal{P}_{V|ZW}\} = \mathbb{A}_d(\mathcal{Q}_{UZW}^\sigma). \quad (\text{VI.31})$$

Then for any  $\tilde{\sigma}$ , we have

$$\max_{\tau \in \text{BR}_d(\tilde{\sigma})} c_e^n(\tilde{\sigma}, \tau) = \max_{\substack{\mathcal{P}_{V|ZW}, \exists \tau \in \text{BR}_d(\tilde{\sigma}) \\ \mathcal{Q}_{V|ZW}^\tau = \mathcal{P}_{V|ZW}}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW}^{\tilde{\sigma}} \\ \mathcal{P}_{V|ZW}}} [c_e(U, V)] \quad (\text{VI.32})$$

$$= \max_{\substack{\mathcal{Q}_{V|ZW} \in \\ \mathbb{A}_d(\mathcal{Q}_{UZW}^{\tilde{\sigma}})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW}^{\tilde{\sigma}} \\ \mathcal{Q}_{V|ZW}}} [c_e(U, V)] \quad (\text{VI.33})$$

$$\geq \inf_{\mathcal{Q}_{UZW} \in \mathcal{Q}} \max_{\substack{\mathcal{Q}_{V|ZW} \in \\ \mathbb{A}_d(\mathcal{Q}_{UZW})}} \mathbb{E}_{\substack{\mathcal{Q}_{UZW} \\ \mathcal{Q}_{V|ZW}}} [c_e(U, V)] = C_e^*, \quad (\text{VI.34})$$

which implies (VI.28).

Note that the sequence  $(nC_e^n)_{n \in \mathbb{N}^*}$  is sub-additive. Indeed, when  $\sigma$  is the concatenation of several encoding strategies, the optimal  $\tau$  in (VI.24) is the concatenation of the optimal decoding strategies. Theorem 36 and Fekete's lemma, show that

$$C_e^* = \lim_{n \rightarrow +\infty} C_e^n = \inf_{n \in \mathbb{N}^*} C_e^n. \quad (\text{VI.35})$$

## VI.2.b Convex closure formulation

Recall that we denote by  $\text{vex } f : \mathcal{X} \rightarrow \mathbb{R} \cup \{-\infty\}$  the convex closure of a function  $f$ , i.e. the largest convex function everywhere smaller than  $f$  on  $X$ . We reformulate  $C_e^*$  in terms of a convex closure, similarly to (Kamenica and Gentzkow, 2011, Corollary 1).

**Lemma 6** For all  $\mathcal{Q}_{W|U} \in \Delta(\mathcal{W})^{|\mathcal{U}|}$ , for all  $(u, z, w) \in \text{supp}(\mathcal{P}_{UZ} \mathcal{Q}_{W|U})$  we have

$$\mathcal{Q}_{U|WZ}(u|w, z) = \frac{\mathcal{Q}_{U|W}(u|w) \mathcal{P}_{Z|U}(z|u)}{\sum_{u'} \mathcal{Q}_{U|W}(u'|w) \mathcal{P}_{Z|U}(z|u')}. \quad (\text{VI.36})$$

The proof is direct. The conditional distribution  $\mathcal{Q}_{U|WZ} \in \Delta(\mathcal{U})^{|\mathcal{W} \times \mathcal{Z}|}$  reformulates in terms of  $\mathcal{Q}_{U|W} \in \Delta(\mathcal{U})^{|\mathcal{W}|}$ .

**Definition 52** Given the probability distribution  $p \in \Delta(\mathcal{U})$ , the decoder selects a best response action  $v$  in

$$\mathcal{V}^*(p) = \underset{v \in \mathcal{V}}{\text{argmin}} \mathbb{E}_p [c_d(U, v)]. \quad (\text{VI.37})$$

Given  $p \in \Delta(\mathcal{U})$ , we define the robust cost function by

$$\psi_e(p) = \max_{v \in \mathcal{V}^*(p)} \mathbb{E}_p [c_e(U, v)]. \quad (\text{VI.38})$$

**Definition 53** For  $p \in \Delta(\mathcal{U})$ , the average cost function  $\Psi_e(p)$  and the average entropy function  $h(p)$  are defined by

$$\Psi_e(p) = \sum_{u,z} p(u) \mathcal{P}_{Z|U}(z|u) \cdot \psi_e \left( \frac{p(\cdot) \mathcal{P}_{Z|U}(z|\cdot)}{\sum_{u'} p(u') \mathcal{P}_{Z|U}(z|u')} \right), \quad (\text{VI.39})$$

$$h(p) = \sum_{u,z} p(u) \mathcal{P}_{Z|U}(z|u) \log_2 \frac{\sum_{u'} p(u') \mathcal{P}_{Z|U}(z|u')}{p(u) \mathcal{P}_{Z|U}(z|u)}. \quad (\text{VI.40})$$

The function  $h(p)$  is the conditional entropy  $H(U|Z)$  evaluated with respect to  $p \cdot \mathcal{P}_{Z|U}$  instead of  $\mathcal{P}_{UZ}$ .

**Lemma 7** The function  $h(p)$  is concave in  $p \in \Delta(\mathcal{U})$ .

*Proof.* [Lemma 7] The entropy  $H(U)$  is concave in  $p \in \Delta(\mathcal{U})$ , the mutual information  $I(U; Z)$  is convex in  $p \in \Delta(\mathcal{U})$  for fixed  $\mathcal{P}_{Z|U}$  and moreover  $H(U|Z) = H(U) - I(U; Z)$ .  $\square$

**Theorem 37**

$$C_e^* = \inf \left\{ \sum_{w \in \mathcal{W}} \lambda_w \Psi_e(p_w), \quad \sum_{w \in \mathcal{W}} \lambda_w p_w = \mathcal{P}_U, \right. \\ \left. \sum_{w \in \mathcal{W}} \lambda_w h(p_w) \geq H(U|Z) - \max_{\mathcal{P}_X} I(X; Y) \right\}, \quad (\text{VI.41})$$

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

The proof of Theorem 37 is stated in (Le Treust and Tomala, 2019, [W3], App. A). It is a consequence of the Markov chain property  $Z \ominus U \ominus W$ . Note that all the channels such that  $\max_{\mathcal{P}_X} I(X; Y) \geq H(U|Z)$  lead to the same value for  $C_e^*$ . The optimal parameters  $(\lambda_w^*, p_w^*)_{w \in \mathcal{W}}$  in (VI.41) are referred to as *the optimal splitting* of the prior distribution  $\mathcal{P}_U$ , see Aumann and Maschler (1995). When removing the decoder side information, e.g.  $|\mathcal{Z}| = 1$ , and changing the infimum into a supremum, we recover the value of the optimal splitting problem of (Le Treust and Tomala, 2019, [J8], Definition 2.4).

- Since  $\sum_w \lambda_w h(p_w) = H(U|Z, W)$ , the information constraint in (VI.41) is a reformulation of  $I(U; W|Z) \leq \max_{\mathcal{P}_X} I(X; Y)$ .
- The dimension of the problem (VI.41) is  $|\mathcal{U}| + 1$ . Caratheodory's Lemma (Rockafellar, 1970, Corollary 17.1.5, pp. 157) induces the cardinality bound  $|\mathcal{W}| \leq |\mathcal{U}| + 1$ .
- The cardinality of  $\mathcal{W}$  is also restricted by the vector of recommended symbols  $|\mathcal{W}| \leq |\mathcal{V}|^{|\mathcal{Z}|}$ , telling to the decoder which symbol  $v \in \mathcal{V}$  to select when the side information is  $z \in \mathcal{Z}$ .

The encoder optimal cost  $C_e^*$  can be reformulated in terms of Lagrangian and in terms of the convex closure of

$$\tilde{\Psi}_e(p, \nu) = \begin{cases} \Psi_e(p), & \text{if } \nu \leq h(p), \\ +\infty, & \text{otherwise.} \end{cases} \quad (\text{VI.42})$$

**Theorem 38**

$$C_e^* = \sup_{t \geq 0} \left\{ \text{vex} \left[ \Psi_e + t \cdot h \right] (\mathcal{P}_U) - t \cdot \left( H(U|Z) - \max_{\mathcal{P}_X} I(X; Y) \right) \right\} \quad (\text{VI.43})$$

$$= \text{vex} \tilde{\Psi}_e \left( \mathcal{P}_U, H(U|Z) - \max_{\mathcal{P}_X} I(X; Y) \right). \quad (\text{VI.44})$$

Equation (VI.43) is the convex closure of a Lagrangian with the information constraint. Equation (VI.44) corresponds to the convex closure of a bi-variate function where the information constraint requires an additional dimension. The proof follows directly from (Le Treust and Tomala, 2019, [J8], Theorem 3.3, pp. 37) by replacing concave closure by convex closure.

**Remark 30** When  $c_e = c_d$ , then  $\psi_e(p) = \min_v \mathbb{E}_p[c_e(U, v)]$  and we reformulate the solution of Wyner and Ziv (1976) in terms of a splitting problem, whose optimal solution is computed as in Sec. VI.2.c.

$$C_e^* = \inf_{\substack{(\lambda_w, p_w)_{w \in \mathcal{W}}, \\ \sum_w \lambda_w = 1}} \left\{ \sum_{w \in \mathcal{W}} \lambda_w \sum_{u, z} p_w(u) \mathcal{P}_{Z|U}(z|u) \times \min_v \mathbb{E} \frac{p_w(\cdot) \mathcal{P}_{Z|U}(z|\cdot)}{\sum_{u'} p_w(u') \mathcal{P}_{Z|U}(z|u')} [c_e(U, v)], \right. \\ \left. \sum_{w \in \mathcal{W}} \lambda_w p_w = \mathcal{P}_U, \quad \sum_{w \in \mathcal{W}} \lambda_w h(p_w) \geq H(U|Z) - \max_{\mathcal{P}_X} I(X; Y) \right\}. \quad (\text{VI.45})$$

When  $c_e = -c_d$ , then  $\mathcal{V}^*(p) = \operatorname{argmax}_v \mathbb{E}_p[c_e(U, v)]$  and both functions  $\psi_e(p)$ ,  $\Psi_e(p)$  are convex in  $p \in \Delta(\mathcal{U})$ . By Jensen's inequality, the infimum in (VI.41) is achieved by  $p_w^* = \mathcal{P}_U$ ,  $\forall w \in \mathcal{W}$ , i.e. no information is transmitted and  $C_e^* = \max_v \mathbb{E}_{\mathcal{P}_U}[c_e(U, v)]$ .

### VI.2.c Doubly symmetric binary source

We consider the doubly symmetric binary source (DSBS) example introduced in (Wyner and Ziv, 1976, Sec. II), depicted in Fig. VI.10 with parameters  $(p_0, \delta_0, \delta_1) \in [0, 1]^3$ . The cardinality bound is  $|\mathcal{W}| = \min(|\mathcal{U}| + 1, |\mathcal{V}|^{|\mathcal{Z}|}) = 3$ , 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  $\sum_k \alpha_k = \sum_k \beta_k = 1$ .

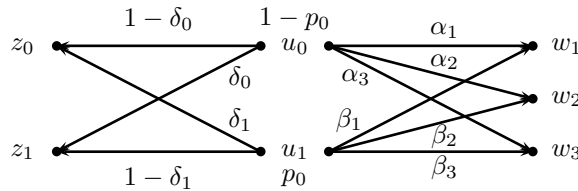


Figure VI.10 – Joint probability distribution  $\mathcal{P}_{UZ}(u, z) \mathcal{Q}_{W|U}(w|u)$  with  $|\mathcal{W}| = 3$  depending on parameters  $p_0, \delta_0, \delta_1, (\alpha_k, \beta_k)_{k \in \{1,2,3\}}$  that belong to  $[0, 1]$ .

	$v_0$	$v_1$
$u_0$	0	1
$u_1$	1	0

Figure VI.11 – Encoder cost  $c_e(u, v)$ .

	$v_0$	$v_1$
$u_0$	0	$1 + \kappa$
$u_1$	1	$\kappa$

Figure VI.12 – Decoder cost  $c_d(u, v)$ .

The cost functions are given by Fig. VI.11 and VI.12 for which the extra cost  $\kappa \in [0, 1]$  in the decoder cost may capture a computing cost, an energy cost, or the fact that an estimation error of the symbol  $v_1$  is more harmful than an estimation error of the symbol  $v_0$ .

The optimal decision for the decoder depends on the posterior belief  $\mathcal{Q}_{U|WZ}(\cdot|w, z) \in \Delta(\mathcal{U})$  after observing the symbols  $(w, z)$ . We denote by  $\gamma = \frac{1+\kappa}{2} = \frac{7}{8}$  the *belief threshold* at which the decoder changes from symbol  $v_0$  to  $v_1$ , as in Fig. VI.13. The decoder chooses  $v_0^*$  (resp.  $v_1^*$ ) when the posterior belief belongs to  $[0, \gamma]$  (resp.  $]\gamma, 1]$ ).

The correlation of  $(U, Z)$  is fixed whereas the correlation of  $(U, W)$  is selected by the encoder. Lemma 6 formulates the *posterior belief*  $\mathcal{Q}_{U|WZ}$  in terms of the *iterim belief*  $\mathcal{Q}_{U|W}$ . For the symbols  $w \in \mathcal{W}$ ,  $z_0 \in \mathcal{Z}$ ,  $z_1 \in \mathcal{Z}$  we have

$$\mathcal{Q}(u_1|w, z_0) = \frac{q\delta_1}{(1-q)(1-\delta_0) + q\delta_1} =: p_0(q), \quad (\text{VI.46})$$

$$\mathcal{Q}(u_1|w, z_1) = \frac{q(1-\delta_1)}{(1-q)\delta_0 + q(1-\delta_1)} =: p_1(q). \quad (\text{VI.47})$$

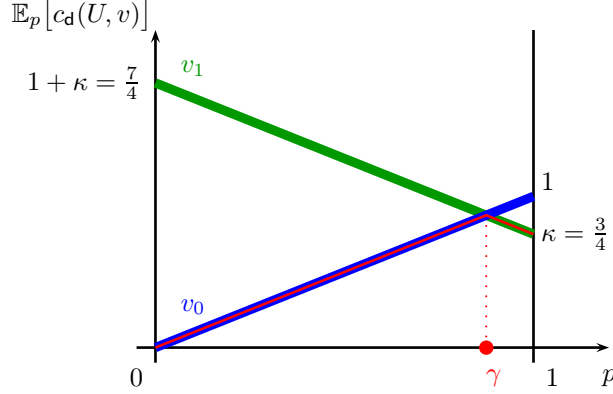


Figure VI.13 – Decoder's expected cost  $\mathbb{E}_p[c_d(U, v)] = (1 - p) \cdot c_d(u_0, v) + p \cdot c_d(u_1, v)$  for  $v \in \{v_0, v_1\}$  depending on the belief  $\mathcal{Q}(u_1|w, z) \in [0, 1]$ .

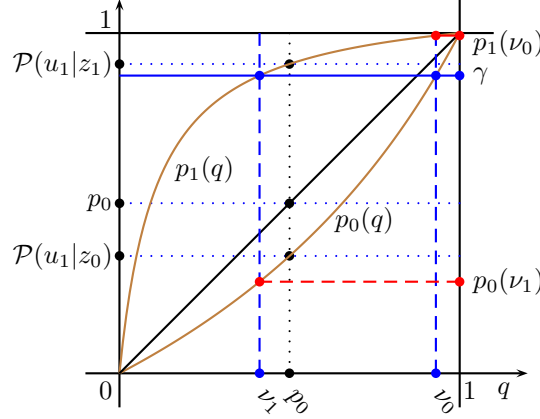


Figure VI.14 – Equations (VI.46) and (VI.47), depending on the *interim belief*  $q \in [0, 1]$ , for  $p_0 = 0.5$ ,  $\delta_1 = 0.05$ ,  $\delta_2 = 0.5$  and  $\gamma = 0.875$ .

Equations (VI.46) and (VI.47) are depicted on Fig. VI.14. Given the belief threshold  $\gamma = \frac{7}{8}$ , we define  $\nu_0$  and  $\nu_1$  such that

$$\gamma = p_0(\nu_0) \iff \nu_0 = \frac{\gamma(1 - \delta_0)}{\delta_1(1 - \gamma) + \gamma(1 - \delta_0)}, \quad (\text{VI.48})$$

$$\gamma = p_1(\nu_1) \iff \nu_1 = \frac{\gamma\delta_0}{\gamma\delta_0 + (1 - \delta_1)(1 - \gamma)}. \quad (\text{VI.49})$$

Without loss of generality, we assume that  $\delta_0 + \delta_1 < 1 \iff \nu_1 < \nu_0$ . The robust and average cost functions writes

$$\psi_{\epsilon}(p) = p \cdot \mathbf{1}(p \leq \gamma) + (1 - p) \cdot \mathbf{1}(p > \gamma), \quad (\text{VI.50})$$

$$\Psi_{\epsilon}(q) = \Pr_q(z_0) \cdot \psi_{\epsilon}(p_0(q)) + \Pr_q(z_1) \cdot \psi_{\epsilon}(p_1(q)) \quad (\text{VI.51})$$

$$= q \cdot \mathbf{1}(q \leq \nu_1) + (1 - q) \cdot \mathbf{1}(q > \nu_0) + (q\delta_1 + (1 - q)\delta_0) \cdot \mathbf{1}(\nu_1 < q \leq \nu_0). \quad (\text{VI.52})$$

The average cost function  $\Psi_{\epsilon}(q)$  is depicted by the orange lines in Fig. VI.15 and Fig. VI.16, where the black curve is the average entropy  $h(q) = H_b(q) + (1 - q) \cdot H_b(\delta_0) + q \cdot H_b(\delta_1) - H_b((1 - q)\delta_0 + q(1 - \delta_1))$  and  $H_b$  denotes the binary entropy.



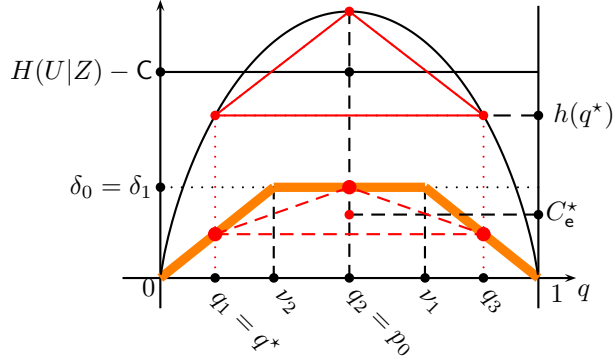


Figure VI.15 – The optimal splitting has three posteriors when  $C \in [0, H(U|Z) - h(q^*)]$ , with  $p_0 = 0.5$ ,  $\delta_0 = \delta_1 = 0.3$ ,  $C = 0.2$ ,  $\kappa = 0$ , then  $C_e^* = 0.2098$ .

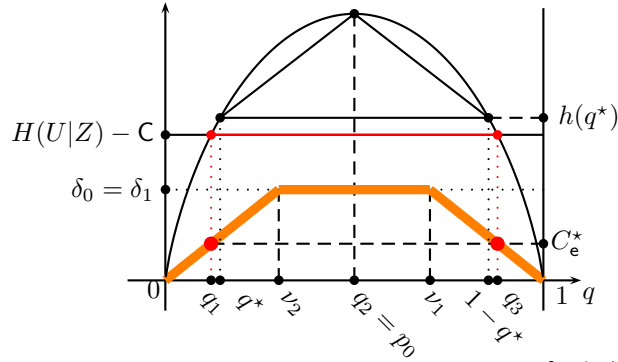


Figure VI.16 – The optimal splitting has only two posteriors when  $C \in [H(U|Z) - h(q^*), H(U|Z)]$ , with  $p_0 = 0.5$ ,  $\delta_0 = \delta_1 = 0.3$ ,  $C = 0.4$ ,  $\kappa = 0$ , then  $C_e^* = 0.1212$ .

The optimal splitting has posteriors  $(q_1, q_2, q_3) \in [0, 1]^3$  with respective weights  $(\lambda_1, \lambda_2, \lambda_3) \in [0, 1]^3$  that satisfy

$$1 = \lambda_1 + \lambda_2 + \lambda_3, \quad (\text{VI.53})$$

$$p_0 = \lambda_1 q_1 + \lambda_2 q_2 + \lambda_3 q_3, \quad (\text{VI.54})$$

$$H(U|Z) - C = \lambda_1 \cdot h(q_1) + \lambda_2 \cdot h(q_2) + \lambda_3 \cdot h(q_3). \quad (\text{VI.55})$$

Equation (VI.55) is satisfied when the information constraint is binding, therefore we obtain (Le Treust and Tomala, 2019, [J8], Eq. (58)-(60)). Without loss of generality, we assume that  $q_1 \in [0, \nu_1[$ ,  $q_2 \in [\nu_1, \nu_2[$ ,  $q_3 \in [\nu_2, 1]$  and characterize the optimal solution in three different scenarios. According to the Splitting Lemma Aumann and Maschler (1995), we have for  $k \in \{1, 2, 3\}$ ,

$$\mathcal{Q}(w_k|u_0) = \mathcal{Q}(w_k) \frac{1 - \mathcal{Q}(u_1|w_k)}{1 - \mathcal{P}(u_1)} = \lambda_k \frac{1 - q_k}{1 - p_0} = \alpha_k, \quad (\text{VI.56})$$

$$\mathcal{Q}(w_k|u_1) = \mathcal{Q}(w_k) \frac{\mathcal{Q}(u_1|w_k)}{\mathcal{P}(u_1)} = \lambda_k \frac{q_k}{p_0} = \beta_k. \quad (\text{VI.57})$$

### Wyner-Ziv's example with equal cost functions

We consider  $p_0 = 0.5$ ,  $\delta_0 = \delta_1 = 0.3$ ,  $\kappa = 0$ , hence both encoder and decoder minimize the Hamming cost and hence  $\gamma = \frac{1}{2}$ . The average cost and average entropy write

$$\Psi_e(q) = q \cdot \mathbf{1}(q \leq \delta) + \delta \cdot \mathbf{1}(\delta < q \leq 1 - \delta) + (1 - q) \cdot \mathbf{1}(q > 1 - \delta), \quad (\text{VI.58})$$

$$h(q) = H(U|Z) + H_b(q) - H_b(q \star \delta), \quad (\text{VI.59})$$

with the notation  $q \star \delta = (1 - q)\delta + q(1 - \delta)$ .

**Proposition 13** *We denote by  $q^*$  the unique solution to*

$$h'(q) = \frac{H(U|Z) - h(q)}{\delta - q}. \quad (\text{VI.60})$$

1) *If  $C \in [0, H(U|Z) - h(q^*)]$  then at the optimum (Fig. VI.15)*

$$q_1 = q^* = 1 - q_3, \quad q_2 = \frac{1}{2}, \quad (\text{VI.61})$$

$$\lambda_1 = \frac{1}{2} \cdot \frac{C}{H(U|Z) - h(q^*)} = \lambda_3 = \frac{1 - \lambda_2}{2}, \quad (\text{VI.62})$$

*which correspond to the distribution parameters  $\alpha_1 = (1 - q^*) \cdot \frac{C}{H(U|Z) - h(q^*)} = \beta_3$ ,  $\alpha_2 = 1 - \frac{C}{H(U|Z) - h(q^*)} = \beta_2$ ,  $\alpha_3 = q^* \cdot \frac{C}{H(U|Z) - h(q^*)} = \beta_1$ , and to the optimal cost*

$$C_e^* = \delta - C \cdot \frac{\delta - q^*}{H(U|Z) - h(q^*)}. \quad (\text{VI.63})$$

2) *If  $C \in [H(U|Z) - h(q^*), H(U|Z)]$  then at the optimum (Fig. VI.16)  $q_1 = h^{-1}(H(U|Z) - C) = 1 - q_3$ ,  $q_2 = \frac{1}{2}$ ,  $\lambda_1 = \frac{1}{2} = \lambda_3$ ,  $\lambda_2 = 0$ , which correspond to the distribution parameters  $\alpha_1 = 1 - h^{-1}(H(U|Z) - C) = 1 - \alpha_3 = \beta_3 = 1 - \beta_1$  and  $\alpha_2 = \beta_2 = 0$  and to the optimal cost*

$$C_e^* = h^{-1}(H(U|Z) - C), \quad (\text{VI.64})$$

*where the notation  $h^{-1}(H(U|Z) - C)$  stands for the unique solution  $q \in [0, 1]$  of the 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  $C_e^* = 0$ .*

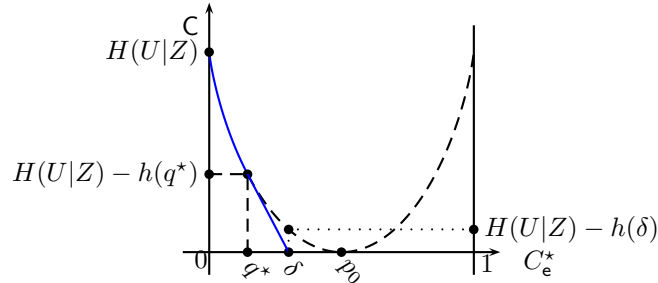


Figure VI.17 – Optimal trade-off between the capacity  $C$  and the optimal cost  $C_e^*$  for the DSBS with parameters  $p_0 = 0.5$ ,  $\delta_0 = \delta_1 = 0.3$ ,  $\kappa = 0$ .

The proof of Proposition 13 is provided in (Le Treust and Tomala, 2019, [W3], App. D). When  $C \leq H(U|Z) - h(q^*)$ , the optimal strategy consists of a time-sharing between  $(C_e^*, C) = (q^*, H(U|Z) - h(q^*))$  and the zero rate point  $(\delta, 0)$ , as depicted in Fig. VI.17.

According to Fig. VI.15 and Fig. VI.16, the optimal distribution of the Wyner-Ziv problem for DSBS has two regimes, either with two posteriors or with three posteriors, i.e. with  $|\mathcal{W}| = 2$  or  $|\mathcal{W}| = 3$ .

### Distinct cost functions without side information

We consider  $p_0 = 0.5$ ,  $C = 0.2$ ,  $\kappa = \frac{3}{4}$  and  $\delta_0 = \delta_1 = 0.5$  so that  $Z$  is independent of  $U$ , as in (Le Treust and Tomala, 2019, [J8]). We have  $H_b(\delta_0) = H_b(\delta_1) = H_b((1 - q)\delta_0 + q(1 - \delta_1)) = 1$  and  $\nu_1 = \nu_2 = \gamma = \frac{7}{8}$ . The average entropy and average cost write

$$h(q) = H_b(q), \quad (\text{VI.65})$$

$$\Psi_e(q) = \psi_e(q) = p \cdot \mathbf{1}(p \leq \gamma) + (1 - p) \cdot \mathbf{1}(p > \gamma), \quad (\text{VI.66})$$

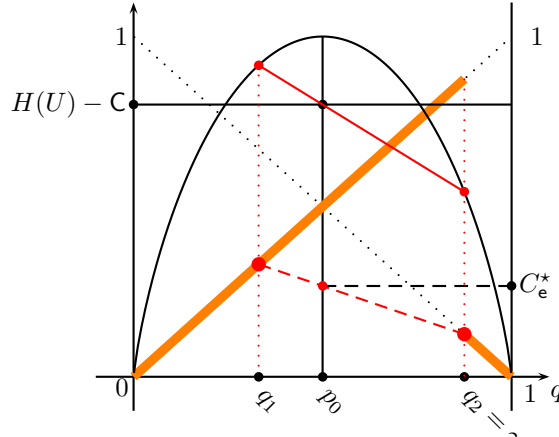


Figure VI.18 – For  $p_0 = 0.5$ ,  $\delta_0 = \delta_1 = 0.5$ ,  $C = 0.2$ ,  $\kappa = \frac{3}{4}$ ,  $C_e^* = 0.2668$ .

and are depicted in Fig. VI.18. Applying (Le Treust and Tomala, 2019, [J8], Corollary 3.5), the optimal splitting has two posteriors, i.e.  $|\mathcal{W}| = 2$ , and satisfy

$$\frac{p_0 - q_2}{q_1 - q_2} \cdot H_b(q_1) + \frac{q_1 - p_0}{q_1 - q_2} \cdot H_b(q_2) \geq H(U) - C. \quad (\text{VI.67})$$

By numerical optimization, the above inequality is satisfied for  $p_0 = 0.5$ ,  $\delta_0 = \delta_1 = 0.5$ ,  $C = 0.2$ ,  $\kappa = \frac{3}{4}$ , hence the optimal cost is achieved by using  $q_2 = \gamma$ , as in Fig. VI.18.

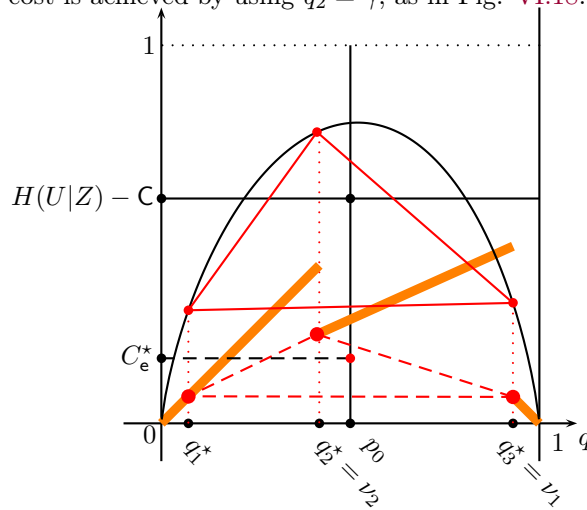


Figure VI.19 –  $p_0 = 0.5$ ,  $\delta_0 = 0.05$ ,  $\delta_1 = 0.5$ ,  $C = 0.2$ ,  $\kappa = \frac{3}{4}$ ,  $C_e^* = 0.1721$ .

### Distinct cost functions with side information

We consider  $p_0 = 0.5$ ,  $\delta_0 = 0.05$ ,  $\delta_1 = 0.5$ ,  $C = 0.2$ ,  $\kappa = \frac{3}{4}$ . By numerical simulation, we determine the optimal triple of posteriors  $(q_1, q_2, q_3)$  represented by the red dots in Fig. VI.19, corresponding to  $C_e^* = 0.1721$ . The parameters of the optimal strategy in Fig. VI.10, are given by the following table.

$q_1 = 0.0715$	$q_2 = 0.4118$	$q_3 = 0.9301$
$\lambda_1 = 0.1288$	$\lambda_2 = 0.6165$	$\lambda_3 = 0.2548$
$\alpha_1 = 0.2392$	$\alpha_2 = 0.7252$	$\alpha_3 = 0.0356$
$\beta_1 = 0.0184$	$\beta_2 = 0.5077$	$\beta_3 = 0.4739$

### VI.3 Bayesian persuasion with several decoders

In this section, we investigate the Bayesian persuasion problem with two decoders via a the communication network of Gray and Wyner (1974), as in Fig. VI.20. The encoder may use three restricted communication channels. One channel of capacity  $R_0$  is public, i.e. both decoder observe its output, and the two other channels of capacity  $R_1$  and  $R_2$  are private to each decoders. We assume that the encoder commits to an encoding strategy and the actions taken by each decoder do not affect the cost function of the other decoder.

Information design with multiple designers interacting with a set of agents is studied in Koessler et al. (2022). In Saritaş et al. (2017), Saritaş et al. (2019), the Nash equilibrium solution is investigated for multi-dimensional sources and quadratic cost functions, whereas the Stackelberg solution is studied in Saritaş et al. (2020). The strategic communication problem with a noisy channel is investigated in Akyol et al. (2015), Akyol et al. (2017). The case where the decoder privately observes a signal correlated to the state, also referred to as the setting of Wyner and Ziv (1976), is studied in Akyol et al. (2016), Bou Rouphael and Le Treust (2020) and Le Treust and Tomala (2021). Vora and Kulkarni investigate the achievable rates for the strategic communication problem in Vora and Kulkarni (2020), Vora and Kulkarni (2020) where the decoder is the Stackelberg leader.

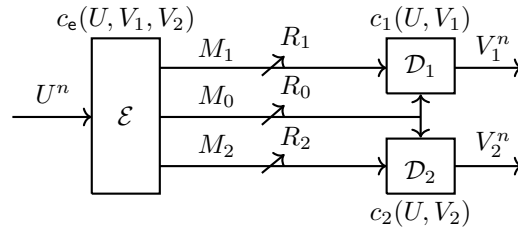


Figure VI.20 – Gray-Wyner Network with Cost-Independent Decoders.

#### VI.3.a System model

In this section, we introduce the coding problem depicted in Fig VI.20.

**Definition 54** Let  $R_0, R_1, R_2 \in \mathbb{R}_+^3 = [0, +\infty[^3$ , and  $n \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$ . We denote by  $M_0 \in \{1, \dots, 2^{\lfloor nR_0 \rfloor}\}$ ,  $M_1 \in \{1, \dots, 2^{\lfloor nR_1 \rfloor}\}$ ,  $M_2 \in \{1, \dots, 2^{\lfloor nR_2 \rfloor}\}$  the random variables of the public and private messages send by the encoder and perfectly received by the decoders, as in Fig. VI.20.

The encoding function  $\sigma$  and the decoding functions  $\tau_i$  of the encoder  $\mathcal{E}$  and decoders  $\mathcal{D}_i$  for  $i \in \{1, 2\}$  respectively, are given by

$$\begin{aligned} \sigma : \mathcal{U}^n &\mapsto \Delta(\{1, \dots, 2^{\lfloor nR_0 \rfloor}\} \times \{1, \dots, 2^{\lfloor nR_1 \rfloor}\} \times \{1, \dots, 2^{\lfloor nR_2 \rfloor}\}), \\ \tau_i : (\{1, 2, \dots, 2^{\lfloor nR_0 \rfloor}\} \times \{1, 2, \dots, 2^{\lfloor nR_i \rfloor}\}) &\mapsto \Delta(\mathcal{V}_i^n). \end{aligned}$$

The coding triplets  $(\sigma, \tau_1, \tau_2)$  are stochastic and induce a joint probability distribution

$$\mathcal{P}_{U^n M_0 M_1 M_2 V_1^n V_2^n}^{\sigma, \tau_1, \tau_2} = \left( \prod_{t=1}^n \mathcal{P}_{U_t} \right) \mathcal{P}_{M_0 M_1 M_2 | U^n}^{\sigma} \mathcal{P}_{V_1^n | M_0 M_1}^{\tau_1} \mathcal{P}_{V_2^n | M_0 M_2}^{\tau_2}. \quad (\text{VI.68})$$

**Definition 55** Single-letter cost functions  $c_e : \mathcal{U} \times \mathcal{V}_1 \times \mathcal{V}_2 \mapsto \mathbb{R}$  of the encoder and  $c_i : \mathcal{U} \times \mathcal{V}_i \mapsto \mathbb{R}$  of

decoder  $\mathcal{D}_i$  for  $i \in \{1, 2\}$  induce Long-run cost functions  $c_e^n(\sigma, \tau_1, \tau_2)$  and  $c_i^n(\sigma, \tau_i)$ ,

$$c_e^n(\sigma, \tau_1, \tau_2) = \sum_{u^n, v_1^n, v_2^n} \mathcal{P}_{U^n V_1^n V_2^n}^{\sigma, \tau_1, \tau_2}(u^n, v_1^n, v_2^n) \cdot \left[ \frac{1}{n} \sum_{t=1}^n c_e(u_t, v_{1,t}, v_{2,t}) \right],$$

$$c_i^n(\sigma, \tau_i) = \mathbb{E}_{\sigma, \tau_i} \left[ \frac{1}{n} \sum_{t=1}^n c_i(U_t, V_{i,t}) \right].$$

where  $\mathcal{P}_{U^n V_1^n V_2^n}^{\sigma, \tau_1, \tau_2}$  denotes the marginal distributions of  $\mathcal{P}^{\sigma, \tau_1, \tau_2}$  over the  $n$ -sequences  $(U^n, V_1^n, V_2^n)$ .

**Definition 56** Given  $\sigma$  the set of best-responses strategies of decoder  $\mathcal{D}_i$ ,  $i \in \{1, 2\}$  is defined by

$$\text{BR}_i(\sigma) = \left\{ \tau_i, \quad c_i^n(\sigma, \tau_i) \leq c_i^n(\sigma, \tilde{\tau}_i), \quad \forall \tilde{\tau}_i \right\}. \quad (\text{VI.69})$$

In case several pairs of strategies are best-responses, we consider the decoders select the worst  $(\tau_1, \tau_2) \in \text{BR}_1(\sigma) \times \text{BR}_2(\sigma)$  for the encoder cost function. Given  $(R_0, R_1, R_2) \in \mathbb{R}_+^3$  et  $n \in \mathbb{N}^*$ , the coding problem under study is defined by

$$C_e^n(R_0, R_1, R_2) = \inf_{\sigma} \max_{\substack{\tau_1 \in \text{BR}_1(\sigma), \\ \tau_2 \in \text{BR}_2(\sigma)}} c_e^n(\sigma, \tau_1, \tau_2). \quad (\text{VI.70})$$

Equation (VI.70) corresponds to the persuasion game that plays as follow:

- The encoder  $\mathcal{E}$  selects and announces the strategy  $\sigma$ .
- The sequence  $U^n$  is drawn i.i.d. according to  $\mathcal{P}_U$ .
- The messages  $(M_0, M_1, M_2)$  are encoded according to  $\mathcal{P}_{M_0 M_1 M_2 | U^n}^{\sigma}$ .
- The decoder  $\mathcal{D}_1$  implements  $\tau_1 \in \text{BR}_1(\sigma)$ : by knowing  $\sigma$ , it observes  $(M_0, M_1)$ , reactualises its Bayesian posterior belief  $\mathcal{P}_{U^n | M_0 M_1}^{\sigma}$  and selects a sequence of actions  $V_1^n$  that minimizes its expected cost function.
- Similarly,  $\mathcal{D}_2$  implements  $\tau_2 \in \text{BR}_2(\sigma)$ .
- We consider the players expected cost values given by  $c_e^n(\sigma, \tau_1, \tau_2)$ ,  $c_1^n(\sigma, \tau_1)$ ,  $c_2^n(\sigma, \tau_2)$ .

**Lemma 8** The sequence  $(nC_e^n(R_0, R_1, R_2))_{n \in \mathbb{N}^*}$  is sub-additive, i.e.  $\forall n, m \in \mathbb{Z}$ , we have

$$(n+m)C_e^{n+m}(R_0, R_1, R_2) \leq nC_e^n(R_0, R_1, R_2) + mC_e^m(R_0, R_1, R_2). \quad (\text{VI.71})$$

To prove Lemma 8, we show that the concatenation of the strategy  $\sigma^n$  that achieves  $nC_e^n(R_0, R_1, R_2)$ , with the strategy  $\sigma^m$  that achieves  $mC_e^m(R_0, R_1, R_2)$ , leads to an upper bound on  $(n+m)C_e^{n+m}(R_0, R_1, R_2)$ .

### VI.3.b Main result

We investigate the asymptotic behaviour of  $C_e^n(R_0, R_1, R_2)$ .

**Definition 57** We consider three auxiliary random variables  $W_0 \in \mathcal{W}_0$ ,  $W_1 \in \mathcal{W}_1$  and  $W_2 \in \mathcal{W}_2$  with  $|\mathcal{W}_0| = |\mathcal{V}_1| \times |\mathcal{V}_2| + 1$  and  $|\mathcal{W}_i| = |\mathcal{V}_i|$ , for  $i \in \{1, 2\}$ . Given  $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ , we define

$$\mathbb{Q}_0(R_0, R_1, R_2) = \left\{ \mathcal{Q}_{W_0 | U} \mathcal{Q}_{W_1 | W_0 U} \mathcal{Q}_{W_2 | W_0 U}, \quad R_0 \geq I(U; W_0), \right. \\ \left. R_0 + R_1 \geq I(U; W_1, W_0), \quad R_0 + R_2 \geq I(U; W_2, W_0) \right\}, \quad (\text{VI.72})$$

$$\tilde{\mathbb{Q}}_0(R_0, R_1, R_2) = \left\{ \mathcal{Q}_{W_0 W_1 W_2 | U}, \quad R_0 \geq I(U; W_0), \right. \\ \left. R_0 + R_1 \geq I(U; W_1, W_0), \quad R_0 + R_2 \geq I(U; W_2, W_0) \right\} \quad (\text{VI.73})$$

A distribution in  $\mathbb{Q}_0(R_0, R_1, R_2)$  must satisfy the Markov chain property  $W_1 \ominus (U, W_0) \ominus W_2$ . For each  $\mathcal{Q}_{W_0 W_1|U}$  and  $\mathcal{Q}_{W_0 W_2|U}$ , we define

$$\mathbb{Q}_1(\mathcal{Q}_{W_0 W_1|U}) = \operatorname{argmin}_{\mathcal{Q}_{V_1|W_0 W_1}} \mathbb{E} \left[ c_1(U, V_1) \right], \quad (\text{VI.74})$$

$$\mathbb{Q}_2(\mathcal{Q}_{W_0 W_2|U}) = \operatorname{argmin}_{\mathcal{Q}_{V_2|W_0 W_2}} \mathbb{E} \left[ c_2(U, V_2) \right]. \quad (\text{VI.75})$$

We define lower and upper bounds on the encoder optimal cost

$$C_e^*(R_0, R_1, R_2) = \inf_{\substack{\mathcal{Q}_{W_0|U} \mathcal{Q}_{W_1|W_0 U} \mathcal{Q}_{W_2|W_0 U} \\ \in \mathbb{Q}_0(R_0, R_1, R_2)}} \max_{\substack{\mathcal{Q}_{V_1|W_0 W_1} \in \mathbb{Q}_1(\mathcal{Q}_{W_0 W_1|U}) \\ \mathcal{Q}_{V_2|W_0 W_2} \in \mathbb{Q}_2(\mathcal{Q}_{W_0 W_2|U})}} \mathbb{E} \left[ c_e(U, V_1, V_2) \right], \quad (\text{VI.76})$$

$$\tilde{C}_e(R_0, R_1, R_2) = \inf_{\substack{\mathcal{Q}_{W_0 W_1 W_2|U} \\ \in \mathbb{Q}_0(R_0, R_1, R_2)}} \max_{\substack{\mathcal{Q}_{V_1|W_0 W_1} \in \mathbb{Q}_1(\mathcal{Q}_{W_0 W_1|U}) \\ \mathcal{Q}_{V_2|W_0 W_2} \in \mathbb{Q}_2(\mathcal{Q}_{W_0 W_2|U})}} \mathbb{E} \left[ c_e(U, V_1, V_2) \right]. \quad (\text{VI.77})$$

Therefore the infimum in (VI.77) has a larger domain which implies  $\tilde{C}_e(R_0, R_1, R_2) \leq C_e^*(R_0, R_1, R_2)$ .

**Lemma 9** If the encoder cost function satisfies

$$c_e(u, v_1, v_2) = c_{e1}(u, v_1) + c_{e2}(u, v_2), \quad \forall (u, v_1, v_2), \quad (\text{VI.78})$$

for some  $c_{e1} : \mathcal{U} \times \mathcal{V}_1 \mapsto \mathbb{R}$  and  $c_{e2} : \mathcal{U} \times \mathcal{V}_2 \mapsto \mathbb{R}$ , then

$$C_e^*(R_0, R_1, R_2) = \tilde{C}_e(R_0, R_1, R_2), \quad (\text{VI.79})$$

for all  $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ .

**Theorem 39** Let  $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ , we have

$$\forall \varepsilon > 0, \exists \bar{n} \in \mathbb{N}, \forall n \geq \bar{n}, \quad C_e^n(R_0, R_1, R_2) \leq C_e^*(R_0, R_1, R_2) + \varepsilon, \quad (\text{VI.80})$$

$$\forall n \in \mathbb{N}, \quad C_e^n(R_0, R_1, R_2) \geq \tilde{C}_e(R_0, R_1, R_2). \quad (\text{VI.81})$$

The proof of Theorem 39 is stated for the case  $R_2 = 0$  in (Bou Rouphael and Le Treust, 2021, [IC28]) and (Bou Rouphael and Le Treust, 2022d, [W5]), and for the general case in (Bou Rouphael and Le Treust, 2022a, [NC5]). By applying Fekete's Lemma for the sub-additive sequence  $(nC_e^n(R_0, R_1, R_2))_{n \in \mathbb{N}^*}$ , for all  $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ , we have

$$\tilde{C}_e(R_0, R_1, R_2) \leq \lim_{n \rightarrow \infty} C_e^n(R_0, R_1, R_2) = \inf_{n \in \mathbb{N}^*} C_e^n(R_0, R_1, R_2) \leq C_e^*(R_0, R_1, R_2).$$

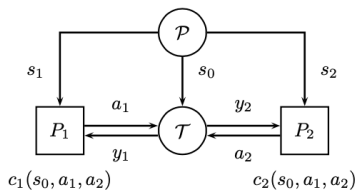
## VI.4 Chapter summary

In this chapter, we characterize the optimal solutions of Bayesian persuasion games when the communication is passed through a noisy channel, when the decoder has side information and when there are two decoders and a common message.

In the scenarios with one encoder and one decoder, the optimal solution reformulates as the convexification of an auxiliary function, that depends on the decoder posterior beliefs. As a side result, this new formulation allows us to determine when Wyner-Ziv's auxiliary random variable for the DSBS, must be binary.

# Chapter VII

## Conclusion and perspectives



Information is crucial in decision making problems, the information asymmetries in Game Theory raise new scenarios and new open questions. In this chapter, we present on-going research directions and a long-term perspectives that will deepen the connection between Game Theory and Information Theory, and also with Graph Theory and Convex Optimization.

In Sec. VII.1, we present some of Nicolas Charpenay’s PhD contributions on zero-error source coding problems with decoder side information. This question is related to the long-standing open problem of the zero-error Shannon capacity of a graph, see Shannon (1956). We study the zero-error version of the coding for computing problem of Orlitsky and Roche (2001), by coloring an appropriate product of graphs. Nicolas’s results are stated in (Charpenay et al., 2022b, [IC30]), (Charpenay and Le Treust, 2020, [IC25]), (Charpenay et al., 2022a, [NC11]), (Charpenay et al., 2022c, [W6]), (Charpenay et al., 2023, [W9]). In Sec. VII.2, we study the convex optimization problem of (IV.28) that involves the information constraint of Gossner et al. (2006), by using a recent convex optimization algorithm that computes efficiently the optimal cost. We will adapt this algorithm in order to compute the solution of various coordination solutions, stated in Chap. IV. In Sec. VII.3, we present a sender-receiver game in which the commitment is at the decoder side, instead of being at the encoder side. This assumption changes the strategic nature of the problem, it is a mechanism design problem, similar to the one of Jackson and Sonnenschein (2007), in which the communication is restricted. This problem can also be seen as a reformulation of the mismatched rate-distortion problem of Lapidoth (1997). The goal is to understand the incentive constraints of both players in order to generalize the cheap talk problem of Crawford and Sobel (1982).

### Contents

---

VII.1 Zero-error coding for computing problem . . . . .	115
VII.1.a Problem statement . . . . .	115
VII.1.b General setting . . . . .	116
VII.1.c Pairwise shared side information . . . . .	117
VII.2 Convex optimization algorithm for coordination problems . . . . .	118
VII.2.a Problem formulation . . . . .	118
VII.2.b Connection with Blahut-Arimoto’s algorithm . . . . .	119
VII.2.c Original optimization algorithm . . . . .	120
VII.3 Mechanism design and cheap talk problems . . . . .	120
VII.3.a System model . . . . .	121
VII.3.b Upper bound . . . . .	121
VII.4 Repeated games with imperfect monitoring and incomplete information . . . . .	122

---

## VII.1 Zero-error coding for computing problem

Consider the source coding scenario depicted in Fig. VII.1 where two correlated sequences  $(X^n, Y^n)$  of discrete i.i.d. random source symbols are drawn with a source distribution  $P_{X,Y} \in \Delta(\mathcal{X} \times \mathcal{Y})$ . The encoder knows the source sequence  $X^n$ , and has the sequence  $(g(Y_t))_{t \leq n}$  as side information, where  $g$  is a deterministic function. The encoder transmits information to the decoder through a perfect channel of capacity  $R \geq 0$ . The decoder has the side information  $Y^n$  and wants to reconstruct the sequence  $(f(X_t, Y_t))_{t \leq n}$ , where  $f$  is deterministic. What is the minimal capacity  $R \geq 0$  such that  $(f(X_t, Y_t))_{t \leq n}$  can be retrieved by the decoder with probability of error 0?

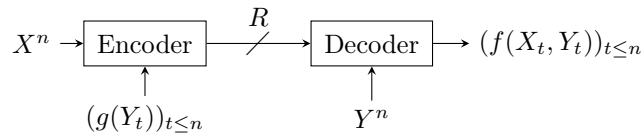


Figure VII.1 – Zero-error coding for computing with side information at the encoder.

This problem can be seen as a zero-error variant of the “coding for computing” problem of [Orlitsky and Roche \(2001\)](#) with side information at the encoder. This problem is difficult, as it has the “restricted inputs” zero-error problem of [Alon and Orlitsky \(1996\)](#) as a particular case, i.e. by taking  $g$  constant and  $f(X, Y) = X$ .

More precisely, the optimal rate in “restricted inputs” is given by asymptotic chromatic entropies of graph products, for which no single-letter expression is known. It is shown in [Kougi et al. \(2003\)](#) that this optimal rate is equal to the complementary graph entropy introduced in [Körner and Longo \(1973\)](#). In [Marton \(1993\)](#), the author show that finding a single-letter expression for the complementary graph entropy would give a single-letter formula for the Shannon capacity of a graph, see [Shannon \(1956\)](#), which is a wide open problem. Bounds for the “restricted inputs” zero-error setting can be obtained with achievability results in the “unrestricted inputs” zero-error setting of [Alon and Orlitsky \(1996\)](#), where  $(X, Y)$  can take values outside the support of  $P_{X,Y}$ . For the “unrestricted inputs” case, the authors gave a single letter characterization of the optimal rate: the Körner graph entropy, introduced in [Körner \(1973\)](#).

Now the scheme of Fig. 1 has been studied for lossless, lossy and zero-error “unrestricted inputs” coding. More precisely, in [Orlitsky and Roche \(2001\)](#) the authors give a single-letter characterization of the lossless variant of this problem; in [Yamamoto \(1982\)](#), the author studies the lossy variant and determines the rate-distortion function; and the optimal rate in the zero-error “unrestricted inputs” based on an extension of the Körner graph entropy is characterized in [Shayevitz \(2014\)](#). The results from “unrestricted inputs” settings can only be used as bounds here: the zero-error “restricted inputs” problem depicted in Fig. VII.1 does not have a characterization of the optimal rate.

Under some hypothesis on  $P_{X,Y}$  and  $g$ , we derive single letter characterization of the optimal rate. Our hypothesis, called “pairwise shared side information”, requires that every pair of source symbols “share” at least one side information symbol for all output of  $g$ . Under this hypothesis, we show that the characteristic graph is a as disjoint union of OR products for which a single-letter characterization is available.

### VII.1.a Problem statement

Given the sequence length  $n \in \mathbb{N}^*$ , we denote by  $\Delta_n(\mathcal{X}) \subset \Delta(\mathcal{X})$  the set of empirical distributions of sequences from  $\mathcal{X}^n$ . We denote by  $\{0, 1\}^*$  the set of binary words. We say that two words  $w \neq w'$  are *not prefix* if  $w$  is not a prefix of  $w'$  and  $w'$  is not a prefix of  $w$ . Let  $\mathcal{S} \subset \{0, 1\}^*$ , we say that  $\mathcal{S}$  is *prefix-free* if for all  $w, w' \in \mathcal{S}$ , the words  $w, w'$  are not prefix. The setting of Fig. VII.1 is described by:

- Four finite sets  $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{Z}$ , a tuple of random variables  $(X, Y) \in \mathcal{X} \times \mathcal{Y}$  drawn from the distribution



$P_{X,Y}$  (with  $P_X$  and  $P_Y$  full-support), and deterministic functions

$$f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{U}, \quad (\text{VII.1})$$

$$g : \mathcal{Y} \rightarrow \mathcal{Z}. \quad (\text{VII.2})$$

- The encoder observes the realizations of  $X^n, (g(Y_t))_{t \leq n}$  and sends information to the decoder over a noiseless channel of capacity  $R \geq 0$ .
- The decoder recover  $(f(X_t, Y_t))_{t \leq n}$  based on the encoder message and the side information  $Y^n$ .

**Definition 58 (Zero-error source code, achievable rates)** *Given  $n \in \mathbb{N}^*$ , a  $(n, R_n)$ -zero-error source code consists of an encoding function  $\phi_e : \mathcal{X}^n \times \mathcal{Z}^n \rightarrow \{0, 1\}^*$  and a decoding function  $\phi_d : \mathcal{Y}^n \times \{0, 1\}^* \rightarrow \mathcal{U}^n$  such that:*

1. the set  $\phi_e(\mathcal{X}^n \times \mathcal{Z}^n)$  is prefix-free;
2.  $R_n = \frac{1}{n} \mathbb{E}[l \circ \phi_e(X^n, (g(Y_t))_{t \leq n})]$ , where  $l(\cdot)$  denotes the length of a binary word;
3. the “restricted inputs” zero-error property is satisfied:

$$\forall (x^n, y^n) \in \text{supp } P_{X,Y}^n, \quad \phi_d(y^n, \phi_e(x^n, (g(y_t))_{t \leq n})) = (f(x_t, y_t))_{t \leq n}. \quad (\text{VII.3})$$

A rate  $R$  is achievable if there exists a sequence of  $(n, R_n)$ -zero-error source codes such that  $\lim_n R_n = R$ . The optimal rate is denoted by  $R^* = \inf\{R \geq 0 \mid R \text{ is achievable}\}$ .

The prefix-free hypothesis guarantees that the decoder knows when the encoder’s message stops and decode correctly with probability one.

### VII.1.b General setting

A probabilistic graph  $G$  is a tuple  $(\mathcal{V}, \mathcal{E}, P_V)$ , where  $\mathcal{V}$  is the set of vertices,  $\mathcal{E}$  is the set of edges, and  $P_V \in \Delta(\mathcal{V})$  is an underlying probability distribution on the vertices. A graph coloring is a mapping that assigns different colors to adjacent vertices. We will make use of graph colorings as the colorings of a specific graph correspond directly to zero-error encoding functions.

**Definition 59 (Coloring, independent subset)** *Let  $G = (\mathcal{V}, \mathcal{E}, P_V)$  be a probabilistic graph. A subset  $\mathcal{S} \subseteq \mathcal{V}$  is independent if  $xx' \notin \mathcal{E}$  for all  $x, x' \in \mathcal{S}$ . Let  $\mathcal{C}$  be a finite set (the set of colors), a mapping  $c : \mathcal{V} \rightarrow \mathcal{C}$  is a coloring if  $c^{-1}(i)$  is an independent subset for all  $i \in \mathcal{C}$ .*

**Definition 60 (Chromatic entropy  $H_X$ )** *The chromatic entropy of a probabilistic graph  $G = (\mathcal{V}, \mathcal{E}, P_V)$  is defined by*

$$H_X(G) = \inf \left\{ H(c(V)) \mid c \text{ is a coloring of } G \right\}. \quad (\text{VII.4})$$

The chromatic entropy of a probabilistic graph is the smallest entropy of a coloring of that graph. As in [Witsenhausen \(1976\)](#), we build the characteristic graph  $G_{[n]}$  that captures the zero-error encoding constraints on a given number  $n$  of source uses.

**Definition 61 (Characteristic graph  $G_{[n]}$ )** *The characteristic graph  $G_{[n]}$  is defined by:*

- $\mathcal{X}^n \times \mathcal{Z}^n$  as set of vertices with distribution  $P_{X,g(Y)}^n$ ,
- $(x^n, z^n)(x'^n, z'^n)$  are adjacent if  $z^n = z'^n$  and there exists  $y^n \in g^{-1}(z^n)$  such that:

$$\forall t \leq n, P_{X,Y}(x_t, y_t) P_{X,Y}(x'_t, y_t) > 0, \quad (\text{VII.5})$$

$$\text{and } \exists t \leq n, f(x_t, y_t) \neq f(x'_t, y_t); \quad (\text{VII.6})$$

where  $g^{-1}(z^n) = \{y^n \in \mathcal{Y}^n \mid (g(y_t))_{t \leq n} = z^n\}$  is the preimage of  $z^n$ .

The characteristic graph  $G_{[n]}$ , differs from the graphs used in [Shayevitz \(2014\)](#), as we do not need a cartesian representation of these graphs to study the optimal rates. Furthermore, it has vertices that correspond to the possible realizations of  $(X^n, (g(Y_t))_{t \leq n})$  known at the encoder, instead of  $\mathcal{X}^n$  like in [Witsenhausen \(1976\)](#) and [Alon and Orlicsky \(1996\)](#). It is also designed to take  $f$  and  $g$  into account, with the same core idea: if two pairs of symbols  $(x^n, z^n), (x'^n, z'^n)$  are adjacent in  $G_{[n]}$ , then the encoder must map them to different codewords, otherwise the decoder will decode erroneously with positive probability.

**Theorem 40 (Optimal rate in general setting)** *The optimal rate writes:*

$$R^* = \lim_{n \rightarrow \infty} \frac{1}{n} H_{\chi}(G_{[n]}). \quad (\text{VII.7})$$

A general single-letter expression for  $R^*$  is missing, due to the lack of intrinsic structure of  $G_{[n]}$ . In Section [VII.1.c](#), we introduce a hypothesis called “pairwise shared side information”, that gives structure to  $G_{[n]}$  and allows us to derive a single-letter expressions for  $R^*$ .

**Remark 31** *In the case  $f(X, Y) = X$  and  $g$  constant, we obtain the “restricted inputs” source coding problem from [Alon and Orlicsky \(1996\)](#). The adjacency condition in the graph  $G_{[n]}$  becomes:*

$$x^n x'^n \text{ are adjacent if } x^n \neq x'^n \text{ and } \forall t \leq n, \exists y_t \in \mathcal{Y}, P_{X,Y}(x_t, y_t) P_{X,Y}(x'_t, y_t) > 0. \quad (\text{VII.8})$$

### VII.1.c Pairwise shared side information

**Definition 62** *The distribution  $P_{X,Y}$  and the function  $g$  satisfy the “pairwise shared side information” if*

$$\forall z \in \mathcal{Z}, \forall x, x' \in \mathcal{X}, \exists y \in g^{-1}(z), P_{XY}(x, y) P_{XY}(x', y) > 0. \quad (\text{VII.9})$$

*Every pair of source symbols “share” at least one side information symbol for all output of  $g$ .*

Note that any distribution  $P_{X,Y}$  with full support satisfies the “pairwise shared side information” hypothesis.

**Definition 63 (Characteristic graph  $G_z^f$ )** *For all  $z \in \mathcal{Z}$ , we define the characteristic graph  $G_z^f$  by*

- $\mathcal{X}$  as set of vertices with distribution  $P_{X|g(Y)=z}$ ,
- $xx'$  are adjacent if  $f(x, y) \neq f(x', y)$  for some  $y \in g^{-1}(z) \cap \text{supp } P_{Y|X=x} \cap \text{supp } P_{Y|X=x'}$ .

Under the “restricted inputs” assumption of [Alon and Orlicsky \(1996\)](#) (resp. “unrestricted inputs”), the  $n$ -shot characteristic graph can be expressed as an iterated AND (resp. OR) product of the one-shot characteristic graph.

**Definition 64 (AND, OR product)** *Let  $G_1 = (\mathcal{V}_1, \mathcal{E}_1, P_{V_1})$ ,  $G_2 = (\mathcal{V}_2, \mathcal{E}_2, P_{V_2})$  be two probabilistic graphs; their AND (resp. OR) product denoted by  $G_1 \wedge G_2$  (resp.  $G_1 \vee G_2$ ) is defined by:  $\mathcal{V}_1 \times \mathcal{V}_2$  as set of vertices,  $P_{V_1} P_{V_2}$  as probability distribution on the vertices, and  $(v_1 v_2), (v'_1 v'_2)$  are adjacent if*

$$\begin{aligned} &v_1 v'_1 \in \mathcal{E}_1 \text{ AND } v_2 v'_2 \in \mathcal{E}_2, \\ &\text{resp. } (v_1 v'_1 \in \mathcal{E}_1 \text{ and } v_1 \neq v'_1) \text{ OR } (v_2 v'_2 \in \mathcal{E}_2 \text{ and } v_2 \neq v'_2); \end{aligned} \quad (\text{VII.10})$$

*with the convention that all vertices are self-adjacent. We denote by  $G_1^{\wedge n}$  (resp.  $G_1^{\vee n}$ ) the  $n$ -th AND (resp. OR) power.*

Introduced in [Körner \(1973\)](#), the Körner graph entropy relates to the limit of the normalized chromatic entropy of an OR product of graphs.

**Definition 65** For all probabilistic graph  $G = (\mathcal{V}, \mathcal{E}, P_V)$ , let  $\Gamma(G)$  be the collection of independent sets of vertices in  $G$ . The Körner graph entropy of  $G$  is defined by

$$H_\kappa(G) = \min_{W \in \Gamma(G)} I(W; V), \quad (\text{VII.11})$$

where the minimum is taken over all distributions  $P_{W|V} \in \Delta(\mathcal{W})^\mathcal{V}$ , with  $\mathcal{W} = \Gamma(G)$  and with the constraint that the random vertex  $V$  belongs to the random independent set  $W$  with probability 1; this is denoted by  $V \in W \in \Gamma(G)$  in (VII.11).

**Proposition 14** (Alon and Orlitsky, 1996, Theorem 5) For all probabilistic graphs  $G$  and  $G'$ ,

$$H_\kappa(G) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G^{\vee n}), \quad (\text{VII.12})$$

$$H_\kappa(G \vee G') = H_\kappa(G) + H_\kappa(G'). \quad (\text{VII.13})$$

There does not exist such a characterization for  $\lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G^{\wedge n})$  defined with respect to the AND product. By using a convex combination of Körner graph entropies, we provide a single-letter expression for the optimal rate  $R^*$ .

**Theorem 41 (Optimal rate with pairwise shared side information)** If  $P_{X,Y}$  and  $g$  satisfy (VII.9), the optimal rate writes:

$$R^* = \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G_z^f). \quad (\text{VII.14})$$

The proof of Theorem 41 is stated in (Charpenay et al., 2023, [W8]).

**Perspective 1** Nicolas Charpenay's PhD research work focuses on zero-error coding for computing problems with two-sided state information. We will further investigate several extensions of this problem, e.g. by considering the joint source-channel coding and the coding with secrecy constraints.

**Perspective 2** Zero-error problems in Information Theory raise new questions in Graph Theory and in Combinatorial Optimization. We will investigate the design of algorithms in order to compute the obtained graph-based solutions, as in (VII.14).

## VII.2 Convex optimization algorithm for coordination problems

In this section, we propose a numerical approach for solving the cost optimization problem of (IV.28) in the empirical coordination framework, under the information constraint  $I(U; V) \leq H(X|U, V)$  of Gossner et al. (2006). This is a joint work with Luis Briceno Arias, Associate Prof. at Univ. Técnica Federico Santa María, Santiago, Chile and member of CMM IRL 2807.

### VII.2.a Problem formulation

**Problem 1** Let  $\mathcal{U}$ ,  $\mathcal{X}$ , and  $\mathcal{V}$  be nonempty finite sets such that  $|\mathcal{X}| \geq 2$  and, let  $(c_{xv}^u)_{u \in \mathcal{U}, x \in \mathcal{X}, v \in \mathcal{V}}$  that corresponds to the cost function  $c: \mathcal{U} \times \mathcal{X} \times \mathcal{V} \rightarrow \mathbb{R}$  and denote by  $p = (p_u)_{u \in \mathcal{U}} = \mathcal{P}_U \in \text{ri } \Delta(\mathcal{U})$  the source distribution, where  $\text{ri}$  stands for the relative interior. The problem is to

$$\text{minimize}_{q \in C} \sum_{u,x,v} p_u q_{xv}^u, \quad (\text{VII.15})$$

where

$$C = \left\{ q \in \Delta(\mathcal{X} \times \mathcal{V})^{|\mathcal{U}|} \mid \sum_{u,x,v} p_u q_{xv}^u \ln \left( \frac{q_{xv}^u}{\sum_{x',u'} p_{u'} q_{x'v}^{u'}} \right) \leq 0 \right\} \quad (\text{VII.16})$$

with the convention  $0 \cdot \ln 0 = 0$  and the logarithm has basis  $e$ .

Problem 1 belongs to the class of optimization problems under entropy constraints, that comes to light in several fields such as Information Theory via *rate-distortion* problems Shannon (1959), *coordination* problems Cuff et al. (2010); Cuff and Zhao (2011); (Le Treust, 2017, [J6]), see Chap IV, in Economics via *rational inattention* problems Sims (2003), *Bayesian persuasion* problems (Le Treust and Tomala, 2019, [J8]) and Repeated Games with *asymmetric information* Gossner et al. (2006), with *finite automata and bounded recall* Neyman and Okada (1999, 2000), with *private observation* Gossner and Vieille (2002), or with *imperfect monitoring* Gossner et al. (2009); Gossner and Tomala (2006, 2007).

We now interpret Problem (1). For every  $(u, x, v) \in \mathcal{U} \times \mathcal{X} \times \mathcal{V}$ , the variable  $q_{xv}^u$  represents a target conditional probability of  $(x, v)$  given  $u$ , thus the objective function is the expected cost with respect to the joint probability distribution  $\{p_u q_{xv}^u\}_{xv} \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{V})$ . In Gossner et al. (2006), the authors prove that the probability distribution  $\{p_u q_{xv}^u\}_{xv}$  satisfies the non-linear constraint of (VII.16) *if and only if* there exists a coding strategy  $(\sigma, \tau) \in \Sigma$  that induces an empirical distribution over the sequences  $(u_t, x_t, v_t)_{t \in \mathbb{N}^*} \in \mathcal{U}^\infty \times \mathcal{X}^\infty \times \mathcal{V}^\infty$ , that converges to  $\{p_u q_{xv}^u\}_{xv}$ . We denote by  $U, X, V$  the random variables drawn according to  $\{p_u q_{xv}^u\}_{xv}$ , thus the function appearing in (VII.16) reformulates as:

$$\sum_{u,x,v} p_u q_{xv}^u \ln \left( \frac{q_{xv}^u}{\sum_{x',v'} p_{u'} q_{x'v'}^u} \right) = I(U; V) - H(X|U, V), \quad (\text{VII.17})$$

where

$$I(U; V) = \sum_{u,x,v} p_u q_{xv}^u \ln \frac{\sum_{a'} q_{x'v}^s}{\sum_{x',v'} p_{u'} q_{x'v'}^u} \quad \text{and} \quad H(X|U, V) = \sum_{u,x,v} p_u q_{xv}^u \ln \frac{\sum_{a'} q_{x'v}^s}{q_{xv}^u} \quad (\text{VII.18})$$

have the following interpretation: the mutual information  $I(U; V)$  measures the correlation between random variables  $V$  and  $U$ ; the conditional entropy  $H(X|U, V)$  measures the capacity of a perfect channel where the input  $X$  must be correlated with  $(U, V)$ . More details are stated in Chap. III and IV.

## VII.2.b Connection with Blahut-Arimoto's algorithm

Now assume that, for every  $(u, x, v) \in \mathcal{U} \times \mathcal{X} \times \mathcal{V}$ ,  $c_{xv}^u = c_v^u \in \mathbb{R}$  and define  $q_v^u := \sum_x q_{xv}^u$ . In this particular instance, it can be proved that Problem 1 is equivalent to

$$\min_{q \in \Xi} \sum_{u,v} p_u c_v^u q_v^u, \quad (\text{VII.19})$$

where

$$\Xi = \left\{ q \in \Delta(\mathcal{V})^{|\mathcal{U}|} \mid \sum_{u,v} p_u q_v^u \ln \left( \frac{q_v^u}{\sum_{u'} p_{u'} q_v^{u'}} \right) \leq \ln |\mathcal{X}| \right\}.$$

This problem is deeply related to the Shannon's lossy source coding problem (see, e.g., (Cover and Thomas, 2006, Chap. 10)) where the non-linear constraint reformulates as:  $I(V; U) - \ln |\mathcal{X}| \leq 0$ . The set of solutions of (VII.19) is nonempty (El Gamal and Kim, 2011, Section 3.6.1) and every solution  $q^*$  to (VII.19) satisfies

$$(\forall v \in \mathcal{V})(\forall u \in \mathcal{U}) \quad q_v^{u*} = \frac{t_v^* e^{-c_v^u/\lambda^*}}{\sum_{v'} t_{v'}^* e^{-c_{v'}^u/\lambda^*}}, \quad (\text{VII.20})$$

where

$$(\forall v \in \mathcal{V}) \quad t_v^* = \sum_u p_u q_v^{u*}$$

and  $\lambda^* \geq 0$  is the Lagrange multiplier associated to the nonlinear constraint. In Arimoto (1972); Blahut (1972), the authors provide an algorithm for solving (VII.19) whose convergence is proved in (Csiszár,

1974, Theorem 1). Given a starting probability  $(t_{v,0})_{v \in \mathcal{V}} \in \Delta(\mathcal{V})$ , the algorithm iterates

$$\text{For } n = 0, 1, 2, \dots \quad (\text{VII.21})$$

$$\left[ \begin{array}{l} \text{For } v \in \mathcal{V} \\ \quad \left[ \begin{array}{l} \text{For } u \in \mathcal{U} \\ \quad \left[ \begin{array}{l} q_{v,n}^u = \frac{t_{v,n} e^{-c_v^u / \lambda^*}}{\sum_{v'} t_{v',n} e^{-c_{v'}^u / \lambda^*}} \\ t_{v,n+1} = \sum_u p_u q_{v,n}^u. \end{array} \right. \end{array} \right. \end{array} \right. \quad (\text{VII.22})$$

Note that, in order to solve (VII.19), the previous algorithm needs the Lagrange multiplier  $\lambda^*$ , which is not known a priori. In order to estimate  $\lambda^*$  usually some heuristics, as bisection methods, are implemented. This means that Blahut-Arimoto's algorithm has to be implemented for several values of  $\lambda$ . As far as we know, the problem (VII.19) cannot be solved directly by methods proposed in the literature without adding additional loops for estimating  $\lambda^*$ .

### VII.2.c Original optimization algorithm

Problem 1 has been solved explicitly for a specific cost function with  $|\mathcal{U}| = |\mathcal{X}| = |\mathcal{V}| = 2$  (Gossner et al., 2006, Example 2.1) but, it seems that Problem 1 has not been explicitly found in its whole generality. On the other hand, since the set  $C$  is convex and the objective function is smooth (even linear), Problem 1 can be solved by the projected gradient method. However, this algorithm needs to compute efficiently the projection onto  $C$ , which is not easy in our case.

**Perspective 3** *We will provide an explicit optimization algorithm that exploits the structure of the problem for solving it.*

**Perspective 4** *We will generalize our optimization algorithm by considering the information constraints of the various scenario of Chap. IV, corresponding to Theorems 6, 9-18.*

## VII.3 Mechanism design and cheap talk problems

We consider the sender-receiver game described in Fig. VII.2. In the Bayesian persuasion game of Kamenica and Gentzkow (2011), see Chap. VI, the encoder commits to implementing a strategy, as the leader of the Stackelberg game.

In this section, it is the decoder that is the leader of the Stackelberg game, i.e. the decoder commits to implement a decoding strategy before the game plays. This new model is closely related to the mechanism design of Jackson and Sonnenschein (2007) and to the lossy source coding with mismatch distortion functions of Lapidath (1997).

By studying the mechanism design problem, we investigate the incentive constraints of the decoder. Then, the goal will be to characterize the set of the Nash equilibria of the cheap talk game, as in Crawford and Sobel (1982). A detailed study of the incentive constraints of both encoder and decoder is mandatory.

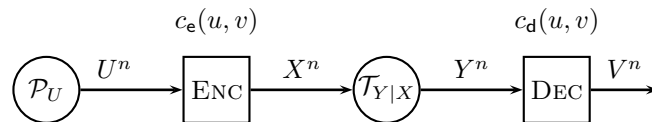


Figure VII.2 – The source  $(\mathcal{U}, \mathcal{P}_U)$  is i.i.d. and the channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$  is memoryless. The encoder and the decoder are endowed with distinct cost functions  $c_e(u, v) \neq c_d(u, v)$ .

### VII.3.a System model

We consider an i.i.d. information source  $(\mathcal{U}, \mathcal{P}_U)$  and a memoryless channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ , as depicted in Fig. VII.2.

**Definition 66** We define the encoding and decoding strategies by

$$\sigma : \mathcal{U}^n \longrightarrow \Delta(\mathcal{X}^n), \quad (\text{VII.23})$$

$$\tau : \mathcal{Y}^n \longrightarrow \Delta(\mathcal{V}^n). \quad (\text{VII.24})$$

The strategies  $(\sigma, \tau)$  induce a distribution defined by

$$\mathcal{P}_{U^n X^n Y^n V^n}^{\sigma, \tau} = \left( \prod_{t=1}^n \mathcal{P}_{U_t} \right) \sigma_{X^n|U^n} \left( \prod_{t=1}^n \mathcal{T}_{Y_t|X_t} \right) \tau_{V^n|Y^n}, \quad (\text{VII.25})$$

where  $\sigma_{X^n|U^n}$ ,  $\tau_{V^n|Y^n}$  denote the distributions of  $\sigma$ ,  $\tau$ .

**Definition 67** The encoder and decoder cost functions  $c_e : \mathcal{U} \times \mathcal{V} \longrightarrow \mathbb{R}$  and  $c_d : \mathcal{U} \times \mathcal{V} \longrightarrow \mathbb{R}$  induce long-run cost functions  $c_e^n(\sigma, \tau)$  and  $c_d^n(\sigma, \tau)$  defined by

$$c_e^n(\sigma, \tau) = \mathbb{E}_{\sigma, \tau} \left[ \frac{1}{n} \sum_{t=1}^n c_e(U_t, V_t) \right], \quad (\text{VII.26})$$

$$c_d^n(\sigma, \tau) = \sum_{u^n, v^n} \mathcal{P}_{U^n V^n}^{\sigma, \tau}(u^n, v^n) \cdot \left[ \frac{1}{n} \sum_{t=1}^n c_d(u_t, v_t) \right]. \quad (\text{VII.27})$$

**Definition 68** Given  $n \in \mathbb{N}^*$ , we define

1. the set of encoder best responses to strategy  $\tau$  by

$$BR_e(\tau) = \underset{\sigma}{\operatorname{argmin}} c_e^n(\sigma, \tau), \quad (\text{VII.28})$$

2. the long-run decoder cost value by

$$C_d^n = \inf_{\tau} \max_{\sigma \in BR_e(\tau)} c_d^n(\sigma, \tau). \quad (\text{VII.29})$$

The value  $C_d^n$  corresponds to the optimal cost of the decoder for fixed  $n \in \mathbb{N}^*$ . In case there are several best responses, we assume the encoder selects the worst strategy  $\sigma$  for the decoder cost. We aim at characterizing the asymptotic behaviour of  $C_d^n$

### VII.3.b Upper bound

**Definition 69** Given an auxiliary random variable  $W \in \mathcal{W}$  with  $|\mathcal{W}| = \min(|\mathcal{U}|+1, |\mathcal{V}|)$  with distribution  $\mathcal{P}_W$ , we define

$$\mathbb{P}(\mathcal{P}_W) = \left\{ \mathcal{Q}_{UW} \in \Delta(\mathcal{U} \times \mathcal{W}), \quad \mathcal{Q}_U = \mathcal{P}_U, \quad \mathcal{Q}_W = \mathcal{P}_W, \quad \max_{\mathcal{P}_X} I(X; Y) - I(U; W) \geq 0 \right\}. \quad (\text{VII.30})$$

Given  $\mathcal{P}_{WV}$ , we define the single-letter encoder best responses

$$A_e(\mathcal{P}_{WV}) = \underset{\mathcal{Q}_{UW} \in \mathbb{P}(\mathcal{P}_W)}{\operatorname{argmin}} \mathbb{E}_{\mathcal{Q}_{UW}} \mathbb{E}_{\mathcal{P}_{V|W}} [c_e(U, V)]. \quad (\text{VII.31})$$

The decoder optimal cost  $C_d^*$  is given by

$$C_d^* = \inf_{\mathcal{P}_{WV}} \max_{\substack{\mathcal{Q}_{UW} \in \\ A_e(\mathcal{P}_{WV})}} \mathbb{E}_{\mathcal{Q}_{UW}} \mathbb{E}_{\mathcal{P}_{V|W}} [c_d(U, V)]. \quad (\text{VII.32})$$

In (VII.32), it is the decoder, i.e. the Stackelberg leader, who selects the marginal distribution  $\mathcal{P}_W$ , whereas the incentive constraints affect the Stackelberg follower, i.e. the encoder. In both settings, the encoder selects the distribution  $\mathcal{Q}_{UW} \in \mathbb{P}(\mathcal{P}_W)$  that satisfies the information constraint and the decoder selects  $\mathcal{P}_{V|W}$ .

**Theorem 42 (Achievability)**

$$\forall \varepsilon > 0, \exists \bar{n} \in \mathbb{N}^*, \forall n \geq \bar{n}, \quad C_d^n \leq C_d^* + \varepsilon. \quad (\text{VII.33})$$

The achievability proof of Theorem 42, see (Le Treust and Tomala, 2020, [IC26]), relies on similar arguments as in (Lapidoth, 1997, Step 1) and (Scarlett et al., 2020, Lemma 4.3). The sequence  $(nC_d^n)_{n \in \mathbb{N}^*}$  is sub-additive, thus Theorem 42 and Fekete's lemma show that

$$C_d^* \geq \lim_{n \rightarrow +\infty} C_d^n = \inf_{n \in \mathbb{N}^*} C_d^n. \quad (\text{VII.34})$$

**Perspective 5** We investigate the converse argument in order to show that

$$\forall n \in \mathbb{N}^*, \quad C_d^* \leq C_d^n. \quad (\text{VII.35})$$

We already know that (VII.35) is valid under several hypothesis: when the encoder cost function  $c_e(u, v)$  does not depend on the state variable  $u \in \mathcal{U}$ , when  $c_e(u, v) = -c_d(u, v)$  or when  $|\mathcal{U}| = |\mathcal{V}| = 2$ .

**Perspective 6** Inspired by the model of Crawford and Sobel (1982), the cheap talk version of this problem corresponds to the scenario where neither the encoder nor the decoder is able to commit to a strategy. We will investigate the cheap talk solution of this problem when the communication is passed through an arbitrary channel  $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ . The goal is to build a theory that encompasses the standard results for the joint source-channel coding problem in Information Theory, see Shannon (1948, 1959), and for the repeated sender-receiver game model, see Crawford and Sobel (1982), Kamenica and Gentzkow (2011), Jackson and Sonnenschein (2007).

## VII.4 Repeated games with imperfect monitoring and incomplete information

The information asymmetries in Game Theory raise new communication scenarios and new open questions. In a general formulation of the problem, each player  $P_k$ ,  $k \in \mathcal{K}$  senses its environment, observes signals  $s_k$  that depend on the system state  $s_0$  and signals  $y_k$  that depend on the actions of the other players, and takes action  $a_k$  in order to minimize its own cost function  $c_k(s_0, a_1, \dots, a_k)$ , as depicted in Fig. VII.3.

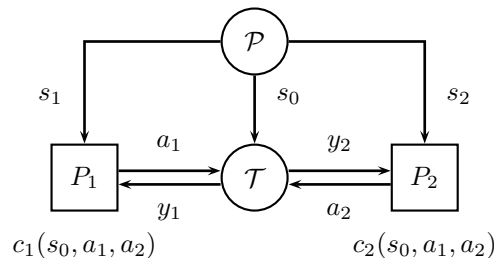


Figure VII.3 – Repeated game with imperfect monitoring and incomplete information. The states are distributed according to  $\mathcal{P}_{S_0 S_1 S_2} \in \Delta(\mathcal{S}_0 \times \mathcal{S}_1 \times \mathcal{S}_2)$  and the observation structure  $\mathcal{T}_{Y_1 Y_2 | S_0 A_1 A_2} \in \Delta(\mathcal{Y}_1 \times \mathcal{Y}_2)^{|\mathcal{S}_0 \times \mathcal{A}_1 \times \mathcal{A}_2|}$  depends on the state  $s_0$ . The players have distinct cost functions  $c_1 : \mathcal{S}_0 \times \mathcal{A}_1 \times \mathcal{A}_2 \rightarrow \mathbb{R}$  and  $c_2 : \mathcal{S}_0 \times \mathcal{A}_1 \times \mathcal{A}_2 \rightarrow \mathbb{R}$ .

Repeated game models focus on incentives of players involved in long-term interactions, see [Sorin \(1992\)](#) and [Mertens et al. \(2015\)](#). The players strategies induce sequences of actions which are evaluated via long-term average cost functions. The perfect monitoring of the actions is crucial to incite the players to cooperate and make the threat of long-term punishment credible. Indeed, the set of Nash equilibrium costs in repeated games of *imperfect monitoring* is unknown. In [Lehrer \(1991\)](#), the players use the observation structure, e.g.  $\mathcal{T}_{Y_1 Y_2 | S_0 A_1 A_2}$  in [Fig. VII.3](#), in order to construct a correlation device robust to unilateral deviations, that induces a correlated equilibrium, see [Aumann \(1974\)](#). Note that a correlated equilibrium is a Nash equilibrium of an extended game, in which players observe correlated signals before choosing their actions.

**Perspective 7** *We will revisit the jointly controlled correlation procedure of [Lehrer \(1991\)](#) through the lens of Information Theory. By using coding arguments, we will try to describe the set of correlation devices robust to the unilateral deviations, that the players can induce via their sequence of plays, when they have imperfect monitoring.*

The players are said to have *incomplete information* when the cost functions depend on a state variable which is uncertain to some players, as in [Fig. VII.3](#). In a repeated game of incomplete information, see [Aumann and Maschler \(1995\)](#), [Forges \(1992\)](#), and [Renault \(2012\)](#), the state variable remains fixed along the course of the game, whereas it evolves dynamically in response of the players' actions in stochastic repeated games, see [Neyman and Sorin \(2003\)](#).

The concept of communication equilibrium, introduced in [Forges \(1986\)](#) and [Myerson \(1986\)](#), generalizes the notion of Nash equilibrium and correlated equilibrium. After partially observing the state variable, the players communicate via a communication device that generates correlated signals before they select their actions.

**Definition 70** *A communication device  $(\{\mathcal{X}_k\}_{k \in \mathcal{K}}, \{\mathcal{Y}_k\}_{k \in \mathcal{K}}, \mathcal{T}_{Y|X})$  is a collection of inputs  $\{\mathcal{X}_k\}_{k \in \mathcal{K}}$  and outputs  $\mathcal{Y}_k$ , for each player  $k \in \mathcal{K}$ , and a conditional probability distribution  $\mathcal{T}_{Y|X} \in \Delta(\mathcal{Y})^{|\mathcal{X}|}$  where  $\mathcal{X} = \prod_{k \in \mathcal{K}} \mathcal{X}_k$  and  $\mathcal{Y} = \prod_{k \in \mathcal{K}} \mathcal{Y}_k$ .*

In the two-player case, the communication device of [Def. 70](#) is a two-way channel, see [Han \(1984\)](#), for which the capacity region is unknown. A communication device is also referred to as the multi-way channel, see [van der Meulen \(1977\)](#)

For infinitely repeated game with imperfect monitoring, the set of communication equilibrium costs is characterized in [Renault and Tomala \(2004\)](#). Such a characterization is not available for Nash and correlated equilibria. The authors reformulate the communication equilibrium of a  $|\mathcal{K}|$  players repeated game by using a correlated equilibrium of an auxiliary two-player repeated game, for which the solution is characterized in [Lehrer \(1992\)](#).

**Perspective 8** *The communication equilibrium concept of [Forges \(1986\)](#) and [Myerson \(1986\)](#) gave us keys to understand the strategic aspects in games with imperfect monitoring and incomplete information. To which extend is it possible to implement a communication equilibrium when the players have a restricted number of available messages? Are such capacity restrictions useful for characterizing the sets of correlated equilibria and Nash equilibria solution? We will explore the impact of the restricted number of messages for communication devices through the lens of information theoretical results for the two-way channel, see [Han \(1984\)](#) and for the multiway channel, see [van der Meulen \(1977\)](#).*



# Bibliography

- Abroshan, M., Gohari, A., and Jaggi, S. (2015). Zero error coordination. In *Proceedings of the IEEE Information Theory Workshop - Fall (ITW)*, pages 202–206.
- Ahlsvede, R. and Körner, J. (1975). Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, 21(6):629–637.
- Akyol, E., Langbort, C., and Başar, T. (2015). Strategic compression and transmission of information. In *Proceedings of the IEEE Information Theory Workshop - Fall (ITW)*, pages 219–223.
- Akyol, E., Langbort, C., and Başar, T. (2016). On the role of side information in strategic communication. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 1626–1630.
- Akyol, E., Langbort, C., and Başar, T. (2017). Information-theoretic approach to strategic communication as a hierarchical game. *Proceedings of the IEEE*, 105(2):205–218.
- Alon, N. and Orlitsky, A. (1996). Source coding and graph entropies. *IEEE Transactions on Information Theory*, 42(5):1329–1339.
- Anantharam, V. and Borkar, V. (2007). Common randomness and distributed control: A counterexample. *Systems & Control Letters*, 56(7-8):568–572.
- Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18(1):14–20.
- Asnani, H., Permuter, H., and Weissman, T. (2013). Successive refinement with decoder cooperation and its channel coding duals. *IEEE Transactions on Information Theory*, 59(9):5511–5533.
- Aumann, R. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96.
- Aumann, R. and Maschler, M. (1995). *Repeated Games with Incomplete Information*. MIT Press, Cambridge, MA.
- Belmega, E.-V., Lasaulce, S., and Debbah, M. (2009). Power allocation games for MIMO multiple access channels with coordination. *IEEE Transactions on Wireless Communications*, 8(6):3182–3192.
- Bennett, C., Shor, P., Smolin, J., and Thapliyal, A. (2002). Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Transactions on Information Theory*, 48(10):2637–2655.
- Bereyhi, A., Bahrami, M., Mirmohseni, M., and Aref, M. (2013). Empirical coordination in a triangular multiterminal network. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 2149–2153.
- Bergemann, D. and Morris, S. (2016). Information design, Bayesian persuasion, and Bayes correlated equilibrium. *American Economic Review Papers and Proceedings*, 106(5):586–591.
- Bergemann, D. and Morris, S. (2017). Information design: a unified perspective. *Cowles Foundation Discussion Paper No 2075*.

- Berger, T. (1978). *Multi-terminal source coding*. Springer-Verlag, chapter in *The Information Theory Approach to Communications* (g. longo) edition.
- Berry, R. and Tse, D. (2011). Shannon meets Nash on the interference channel. *IEEE Transactions on Information Theory*, 57(5):2821–2836.
- Blahut, R. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473.
- Blasco-Serrano, R., Thobaben, R., and Skoglund, M. (2012). Polar codes for coordination in cascade networks. In *Proceedings of the International Zurich Seminar on Communication, Zurich, Switzerland*, pages 55–58.
- Bloch, M. and Barros, J. (2011). *Physical Layer Security - From Information Theory to Security Engineering*. Cambridge University Press.
- Bloch, M. and Kliewer, J. (2013). Strong coordination over a line network. *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 2319–2323.
- Bloch, M. and Kliewer, J. (2014). Strong coordination over a three-terminal relay network. In *Proceedings of the IEEE Information Theory Workshop (ITW)*, pages 646–650.
- Bloch, M., Luzzi, L., and Kliewer, J. (2012). Strong coordination with polar codes. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, pages 565–571.
- Blume, A., Board, O. J., and Kawamura, K. (2007). Noisy talk. *Theoretical Economics*, 2:395–440.
- Boleslavsky, R. and Cotton, C. (2015). Grading standards and education quality. *American Economic Journal: Microeconomics*, 7(2):248–279.
- Boleslavsky, R. and Kim, K. (2018). Bayesian persuasion and moral hazard. *Working paper*.
- Bou Rouchael, R. and Le Treust, M. (2020). Impact of private observation in the Bayesian persuasion game. In *Network Games, Control and Optimization*, pages 229–238. Springer International Publishing.
- Bou Rouchael, R. and Le Treust, M. (2021). Strategic successive refinement coding for Bayesian persuasion with two decoders. In *Proceedings of the IEEE Information Theory Workshop (ITW)*.
- Bou Rouchael, R. and Le Treust, M. (2022a). Communication stratégique via le réseau Gray-Wyner. In *Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI22), Lyon, France*.
- Bou Rouchael, R. and Le Treust, M. (2022b). Strategic communication via cascade multiple-description network. [on-line] available: <https://arxiv.org/abs/2210.00533>.
- Bou Rouchael, R. and Le Treust, M. (2022c). Strategic communication with cost-dependent decoders via the Gray-Wyner network. In *Proceedings of the IEEE Information Theory Workshop (ITW)*.
- Bou Rouchael, R. and Le Treust, M. (2022d). Strategic successive refinement for Bayesian persuasion with cost-dependent decoders. Technical report, working paper.
- Bross, S. I. and Lapidoth, A. (2018). The rate-and-state capacity with feedback. *IEEE Transactions on Information Theory*, 64(3):1893–1918.
- Caire, G. and Tuninetti, D. (2001). Throughput of hybrid-ARQ protocols for Gaussian collision channel. *IEEE Transactions on Information Theory*, 47(5):1971–1988.
- Caplin, A. and Dean, M. (2013). Behavioral implications of rational inattention with Shannon entropy. *NBER Working Papers 19318*.

- Cervia, G., Luzzi, L., Bloch, M. R., and Le Treust, M. (2016). Polar coding for empirical coordination of signals and actions over noisy channels. In *Proceedings of the IEEE Information Theory Workshop (ITW)*.
- Cervia, G., Luzzi, L., Le Treust, M., and Bloch, M. R. (2017a). Polar codes for empirical coordination over noisy channels with strictly causal encoding. In *Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI17)*, Juan-les-Pins, France.
- Cervia, G., Luzzi, L., Le Treust, M., and Bloch, M. R. (2017b). Strong coordination of signals and actions over noisy channels. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*.
- Cervia, G., Luzzi, L., Le Treust, M., and Bloch, M. R. (2018). Strong coordination over noisy channels with strictly causal encoding. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*.
- Cervia, G., Luzzi, L., Le Treust, M., and Bloch, M. R. (2020). Strong coordination of signals and actions over noisy channels with two-sided state information. *IEEE Transactions on Information Theory*, 66(8):4681–4708.
- Charpenay, N. and Le Treust, M. (2020). Zero-error coding with a generator set of variable-length words. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*.
- Charpenay, N., Le Treust, M., and Roumy, A. (2022a). Codage de source zéro-erreur pour un réseau de Gray-Wyner lorsque l’information adjacente peut être présente. In *Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI22)*, Lyon, France.
- Charpenay, N., Le Treust, M., and Roumy, A. (2022b). Zero-error source coding when side information may be present. In *Proceedings of the International Zurich Seminar on Information and Communication (IZS 2022)*.
- Charpenay, N., Le Treust, M., and Roumy, A. (2022c). Zero-error source coding with two-sided side information. Technical report, IRISA UMR 6074.
- Charpenay, N., Le Treust, M., and Roumy, A. (2023). Zero-error coding for computing with encoder side-information. In [on-line] available: <https://arxiv.org/abs/2211.03649>, submitted to *IEEE Information Theory Workshop (ITW)*, Saint-Malo, France.
- Chou, R., Bloch, M., and Kliewer, J. (2015). Polar coding for empirical and strong coordination via distribution approximation. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*.
- Choudhuri, C., Kim, Y.-H., and Mitra, U. (2010). Capacity-distortion trade-off in channels with state. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, pages 1311–1318.
- Choudhuri, C., Kim, Y.-H., and Mitra, U. (2011). Causal state amplification. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 2110–2114.
- Choudhuri, C., Kim, Y.-H., and Mitra, U. (2013). Causal state communication. *IEEE Transactions on Information Theory*, 59(6):3709–3719.
- Choudhuri, C. and Mitra, U. (2012a). On non-causal side information at the encoder. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, pages 648–655.
- Choudhuri, C. and Mitra, U. (2012b). On Witsenhausen’s counterexample: the asymptotic vector case. In *Proceedings of the IEEE Information Theory Workshop (ITW)*, pages 162–166. IEEE.
- Costa, M. H. M. (1983). Writing on dirty paper. *IEEE Transactions on Information Theory*, 29(3):439–441.

- Cover, T. and Chiang, M. (2002). Duality between channel capacity and rate distortion with two-sided state information. *IEEE Transactions on Information Theory*, 48(6):1629–1638.
- Cover, T. and Permuter, H. (2007). Capacity of coordinated actions. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 2701–2705.
- Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. 2nd. Ed., Wiley-Interscience, New York.
- Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6):1431–1451.
- Csiszár, I. (1974). On the computation of rate-distortion functions (corresp.). *IEEE Transactions on Information Theory*, 20(1):122–124.
- Csiszár, I. and Körner, J. (2011). *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press.
- Cuff, P. (2008). Communication requirements for generating correlated random variables. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 1393–1397.
- Cuff, P. (2009). *Communication in Networks for Coordinating Behavior*. PhD thesis, Stanford University.
- Cuff, P. (2013). Distributed channel synthesis. *IEEE Transactions on Information Theory*, 59(11):7071–7096.
- Cuff, P., Permuter, H., and Cover, T. (2010). Coordination capacity. *IEEE Transactions on Information Theory*, 56(9):4181–4206.
- Cuff, P. and Schieler, C. (2011). Hybrid codes needed for coordination over the point-to-point channel. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, pages 235–239.
- Cuff, P. and Zhao, L. (2011). Coordination using implicit communication. In *Proceedings of the IEEE Information Theory Workshop (ITW)*, pages 467–471.
- Doval, L. and Skreta, V. (2018). Constrained information design: Toolkit. *Working paper*, <http://www.laura-doval.com/constrained-information-design-toolkit>.
- Dughmi, S., Niazadeh, R., Psomas, A., and Weinberg, S. M. (2019). Persuasion and incentives through the lens of duality. In *Proceedings of the International Conference on Web and Internet Economics*, pages 142–155. Springer.
- El Gamal, A. and Kim, Y.-H. (2011). *Network Information Theory*. Cambridge University Press.
- Equitz, W. and Cover, T. (1991). Successive refinement of information. *IEEE Transactions on Information Theory*, 37(2):269–275.
- Fekete, M. (1923). Über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten. *Mathematische Zeitschrift*, 17(1):228–249.
- Forges, F. (1986). An approach to communication equilibria. *Econometrica: Journal of the Econometric Society*, pages 1375–1385.
- Forges, F. (1992). Chapter 6: Repeated Games of Incomplete Information: Non-Zero-Sum. In *Handbook of Game Theory with Economic Applications*, volume 1, pages 155–177. Elsevier.
- Gel'fand, S. I. and Pinsker, M. S. (1980). Coding for channel with random parameters. *Problems of Control and Information Theory*, 9(1):19–31.
- Gentzkow, M. and Kamenica, E. (2014). Costly persuasion. *American Economic Review*, 104:457–462.

- Gohari, A. and Anantharam, V. (2011). Generating dependent random variables over networks. In *Proceedings of the IEEE Information Theory Workshop (ITW)*, pages 698–702.
- Goldfeld, Z., Permuter, H., and Kramer, G. (2014). The Ahlswede-Körner coordination problem with one-sided encoder cooperation. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 1341–1345.
- Gossner, O., Hernández, P., and Neyman, A. (2006). Optimal use of communication resources. *Econometrica*, 74(6):1603–1636.
- Gossner, O., Laraki, R., and Tomala, T. (2009). Informationally optimal correlation. *Mathematical Programming*, 116(1-2):147–172.
- Gossner, O. and Tomala, T. (2006). Empirical distributions of beliefs under imperfect observation. *Mathematics of Operation Research*, 31(1):13–30.
- Gossner, O. and Tomala, T. (2007). Secret correlation in repeated games with imperfect monitoring. *Mathematics of Operation Research*, 32(2):413–424.
- Gossner, O. and Vieille, N. (2002). How to play with a biased coin? *Games and Economic Behavior*, 41(2):206–226.
- Gray, R. M. and Wyner, A. D. (1974). Source coding for a simple network. *The Bell System Technical Journal*, 53(9):1681–1721.
- Grover, P., Park, S. Y., and Sahai, A. (2013). Approximately optimal solutions to the finite-dimensional witsenhausen counterexample. *IEEE Transactions on Automatic Control*, 58(9):2189–2204.
- Grover, P. and Sahai, A. (2008). A vector version of witsenhausen’s counterexample: A convergence of control, communication and computation. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 1636–1641.
- Grover, P. and Sahai, A. (2010). Witsenhausen’s counterexample as assisted interference suppression. *International Journal of Systems, Control and Communications*, 2(1-3):197–237.
- Grover, P., Wagner, A. B., and Sahai, A. (2015). Information embedding and the triple role of control. *IEEE Transactions on Information Theory*, 61(4):1539–1549.
- Gülcü, T. C. and Barg, A. M. (2016). Interactive function computation via polar coding. *Problems of Information Transmission*, 52(1):66–91.
- Haddadpour, F., Yassaee, M., Gohari, A., and Aref, M. (2012). Coordination via a relay. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 3048–3052.
- Haddadpour, F., Yassaee, M. H., Beigi, S., Gohari, A., and Aref, M. R. (2017). Simulation of a channel with another channel. *IEEE Transactions on Information Theory*, 63(5):2659–2677.
- Han, T. (1984). A general coding scheme for the two-way channel. *IEEE Transactions on Information Theory*, 30(1):35–44.
- Han, T. and Verdú, S. (1993). Approximation theory of output statistics. *IEEE Transactions on Information Theory*, 39(3):752–772.
- Hebert, B. and Woodford, M. (2018). Information costs and sequential information sampling. *NBER Working Paper 25316*.
- Inostroza, N. and Pavan, A. (2021). Persuasion in global games with application to stress testing. *CEPR Discussion Paper No. DP16484*.
- Jackson, M. O. and Sonnenschein, H. F. (2007). Overcoming incentive constraints by linking decisions. *Econometrica*, 75(1):241–257.

- Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101:2590–2615.
- Kazıklı, E., Sarıtaş, S., Gezici, S., and Yüksel, S. (2021). Optimal signaling with mismatch in priors of an encoder and decoder. [on-line] available: <https://arxiv.org/abs/2101.00799>.
- Keshet, G., Steinberg, Y., and Merhav, N. (2008). Channel coding in the presence of side information. *Foundations and Trends in Communications and Information Theory*, 4(6):445–586.
- Kim, Y.-H., Sutivong, A., and Cover, T. (2008). State amplification. *IEEE Transactions on Information Theory*, 54(5):1850–1859.
- Koessler, F., Laclau, M., and Tomala, T. (2022). Interactive information design. *Mathematics of Operations Research*, 47(1):153–175.
- Körner, J. (1973). Coding of an information source having ambiguous alphabet and the entropy of graphs. In *6th Prague conference on information theory*, pages 411–425.
- Körner, J. and Longo, G. (1973). Two-step encoding for finite sources. *IEEE Transactions on Information Theory*, 19(6):778–782.
- Kouligi, P., Tuncel, E., Regunathan, S., and Rose, K. (2003). On zero-error source coding with decoder side information. *IEEE Transactions on Information Theory*, 49(1):99–111.
- Koyluoglu, O. O., Sundararajan, R., and Vishwanath, S. (2016). State amplification subject to masking constraints. *IEEE Transactions on Information Theory*, 62(11):6233–6250.
- Kramer, G. and Savari, S. (2002). Quantum data compression with commuting density operators. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 44–.
- Kramer, G. and Savari, S. (2007). Communicating probability distributions. *IEEE Transactions on Information Theory*, 53(2):518–525.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.
- Lapidoth, A. (1997). On the role of mismatch in rate distortion theory. *IEEE Transactions on Information Theory*, 43(1):38–47.
- Lapidoth, A. and Wigger, M. (2016). Conditional and relevant common information. In *Proceedings of the IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, pages 1–5.
- Larrousse, B. and Lasaulce, S. (2013). Coded power control: Performance analysis. *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*.
- Larrousse, B., Lasaulce, S., and Bloch, M. R. (2018). Coordination in distributed networks via coded actions with application to power control. *IEEE Transactions on Information Theory*, 64(5):3633–3654.
- Larrousse, B., Lasaulce, S., and Wigger, M. (2015a). Coordinating partially-informed agents over state-dependent networks. In *Proceedings of the IEEE Information Theory Workshop (ITW)*.
- Larrousse, B., Lasaulce, S., and Wigger, M. (2015b). Coordination in state-dependent distributed networks: The two-agent case. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*.
- Le Treust, M. (2014). Correlation between channel state and information source with empirical coordination constraint. In *Proceedings of the IEEE Information Theory Workshop (ITW)*, pages 272–276.
- Le Treust, M. (2015). Coding Theorems for Empirical Coordination. <https://hal.archives-ouvertes.fr/hal-01865569>.

- Le Treust, M. (2015a). Coordination d'appareils autonomes dans un réseau décentralisé. In *Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI15)*, Lyon, France.
- Le Treust, M. (2015b). Empirical coordination with channel feedback and strictly causal or causal encoding. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*.
- Le Treust, M. (2015c). Empirical coordination with two-sided state information and correlated source and state. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 466–470.
- Le Treust, M. (2017). Joint empirical coordination of source and channel. *IEEE Transactions on Information Theory*, 63(8):5087–5114.
- Le Treust, M. and Bloch, M. (2016). Empirical coordination, state masking and state amplification: Core of the decoder's knowledge. *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 895–899.
- Le Treust, M. and Bloch, M. (2019). Jeu d'estimation de l'état de canal et coordination. In *Actes de la Conférence du Groupement de Recherche en Traitement du Signal et des Images (GRETSI19)*, Lille, France.
- Le Treust, M. and Bloch, M. R. (2021). State leakage and coordination with causal state knowledge at the encoder. *IEEE Transactions on Information Theory*, 67(2):805–823.
- Le Treust, M. and Lasaulce, S. (2010). A repeated game formulation of energy-efficient decentralized power control. *IEEE Transactions on Wireless Communication*, 9(9):2860–2869.
- Le Treust, M. and Lasaulce, S. (2013). Transforming monitoring structure with resilient encoders - application to repeated games. *Springer Dynamic Games and Applications*, 3(1):38–67.
- Le Treust, M. and Oechtering, T. (2021). Continuous random variable estimation is not optimal for the Witsenhausen counterexample. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*.
- Le Treust, M. and Oechtering, T. J. (2018). Optimal control designs for vector-valued Witsenhausen counterexample setups. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, pages 532–537. IEEE.
- Le Treust, M. and Oechtering, T. J. (2022). Power-estimation trade-off of vector-valued witsenhausen counterexample with causal decoder. [on-line] available: <http://arxiv.org/abs/2211.09200>, submitted to *IEEE Transactions on Information Theory*.
- Le Treust, M. and Tomala, T. (2016). Information design for strategic coordination of autonomous devices with non-aligned utilities. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, pages 233–242.
- Le Treust, M. and Tomala, T. (2017). 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.
- Le Treust, M. and Tomala, T. (2018a). Information-theoretic limits of strategic communication. [on-line] available: <https://arxiv.org/abs/1807.05147>.
- Le Treust, M. and Tomala, T. (2018b). Strategic coordination with state information at the decoder. *Proceedings of the International Zurich Seminar on Information and Communication*.
- Le Treust, M. and Tomala, T. (2019). Persuasion with limited communication capacity. *Journal of Economic Theory*, 184:104940.

- Le Treust, M. and Tomala, T. (2019). Strategic communication with side information at the decoder. *[on-line] available: <https://arxiv.org/abs/1911.04950>*.
- Le Treust, M. and Tomala, T. (2020). Point-to-point strategic communication. In *Proceedings of the IEEE Information Theory Workshop (ITW)*.
- Le Treust, M. and Tomala, T. (2021). Strategic communication with decoder side information. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*.
- Le Treust, M., Zaidi, A., and Lasaulce, S. (2011). An achievable rate region for the broadcast wiretap channel with asymmetric side information. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, pages 68–75.
- Lehrer, E. (1991). Internal correlation in repeated games. *International Journal of Game Theory*, 19(4):431–456.
- Lehrer, E. (1992). Correlated equilibria in two-player repeated games with nonobservable actions. *Mathematics of Operations Research*, 17(1):175–199.
- Lim, S., Minero, P., and Kim, Y.-H. (2010). Lossy communication of correlated sources over multiple access channels. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, pages 851–858.
- Martin, D. (2017). Strategic pricing with rational inattention to quality. *Games and Economic Behavior*, 104:131–145.
- Marton, K. (1993). On the Shannon capacity of probabilistic graphs. *Journal of Combinatorial Theory, Series B*, 57(2):183–195.
- Matejka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.
- Mathevet, L., Perego, J., and Taneva, I. (2020). On information design in games. *Journal of Political Economy*, 128(4):1370–1404.
- Matyskova, L. (2018). Bayesian persuasion with costly information acquisition. *Working Paper*.
- Merhav, N. and Shamai, S. (2003). On joint source-channel coding for the Wyner-Ziv source and the Gel'fand-Pinsker channel. *IEEE Transactions on Information Theory*, 49(11):2844–2855.
- Merhav, N. and Shamai, S. (2007). Information rates subject to state masking. *IEEE Transactions on Information Theory*, 53(6):2254–2261.
- Mertens, J., Sorin, S., and Zamir, S. (2015). *Repeated Games*. Cambridge University Press.
- Minero, P., Lim, S., and Kim, Y.-H. (2011). Joint source-channel coding via hybrid coding. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 781–785.
- Morris, S. and Strack, P. (2019). The Wald problem and the equivalence of sequential sampling and static information costs. *Working Paper*.
- Moulin, P. and O'Sullivan, J. (2003). Information-theoretic analysis of information hiding. *IEEE Transactions on Information Theory*, 49(3):563–593.
- Myerson, R. B. (1986). Multistage games with communication. *Econometrica: Journal of the Econometric Society*, pages 323–358.
- Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, 54:286–295.
- Neyman, A. and Okada, D. (1999). Strategic entropy and complexity in repeated games. *Games and Economic Behavior*, 29(1–2):191–223.



- Neyman, A. and Okada, D. (2000). Repeated games with bounded entropy. *Games and Economic Behavior*, 30(2):228–247.
- Neyman, A. and Sorin, S. (2003). *Stochastic games and applications*, volume 570. Springer Science & Business Media.
- Oechtering, T. J. and Le Treust, M. (2019). Coordination coding with causal decoder for vector-valued Witsenhausen counterexample setups. In *Proceedings of the IEEE Information Theory Workshop (ITW)*.
- Orlitsky, A. and Roche, J. (2001). Coding for computing. *IEEE Transactions on Information Theory*, 47(3):903–917.
- Perez-Richet, E. and Skreta, V. (2022). Test design under falsification. *Econometrica*, 90(3):1109–1142.
- Perlaza, S., Tandon, R., Poor, H., and Han, Z. (2012). The Nash equilibrium region of the linear deterministic interference channel with feedback. In *Proceedings of the IEEE Annual Allerton Conference on Communication, Control, and Computing*, pages 1790–1797.
- Raginsky, M. (2010). Empirical processes and typical sequences. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 1458–1462.
- Raginsky, M. (2013). Empirical processes, typical sequences, and coordinated actions in standard borel spaces. *IEEE Transactions on Information Theory*, 59(3):1288–1301.
- Renault, J. (2012). *Repeated Games with Incomplete Information*, pages 2635–2655. Springer New York, New York, NY.
- Renault, J. and Tomala, T. (2004). Communication equilibrium payoffs in repeated games with imperfect monitoring. *Games and Economic Behaviour*, 49:313–344.
- Rimoldi, B. (1994). Successive refinement of information: characterization of the achievable rates. *IEEE Transactions on Information Theory*, 40(1):253–259.
- Rockafellar, R. (1970). *Convex Analysis*. Princeton landmarks in mathematics and physics. Princeton University Press.
- Sarıtaş, S., Furrer, P., Gezici, S., Linder, T., and Yüksel, S. (2019). On the number of bins in equilibria for signaling games. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 972–976.
- Sarıtaş, S., Yüksel, S., and Gezici, S. (2017). Quadratic multi-dimensional signaling games and affine equilibria. *IEEE Transactions on Automatic Control*, 62(2):605–619.
- Sarıtaş, S., Yüksel, S., and Gezici, S. (2020). Dynamic signaling games with quadratic criteria under Nash and Stackelberg equilibria. *Automatica*, 115:108883.
- Satpathy, S. and Cuff, P. (2013). Secure cascade channel synthesis. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 2955–2959.
- Satpathy, S. and Cuff, P. (2016). Secure cascade channel synthesis. *IEEE Transactions on Information Theory*, 62(11):6081–6094.
- Scarlett, J., Guillén i Fàbregas, A., Somekh-Baruch, A., and Martinez, A. (2020). Information-theoretic foundations of mismatched decoding. *Foundations and Trends in Communications and Information Theory*, 17(2–3):149–401.
- Schieler, C. and Cuff, P. (2014). Rate-distortion theory for secrecy systems. *IEEE Transactions on Information Theory*, 60(12):7584–7605.

- Shamai, S. and Verdú, S. (1997). The empirical distribution of good codes. *IEEE Transactions on Information Theory*, 43(3):836–846.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.
- Shannon, C. (1958). Channels with side information at the transmitter. *IBM Journal of Research and Development*, 2(4):289–293.
- Shannon, C. E. (1956). The zero error capacity of a noisy channel. *IRE Trans. Inf. Theory*, 2:8–19.
- Shannon, C. E. (1959). Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec*, 4(142-163):1.
- Shayevitz, O. (2014). Distributed computing and the graph entropy region. *IEEE transactions on information theory*, 60(6).
- Sims, C. (2003). Implication of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.
- Sion, M. (1958). On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176.
- Soljanin, E. (2002). Compressing quantum mixed-state sources by sending classical information. *IEEE Transactions on Information Theory*, 4(8):2263–2275.
- Sorin, S. (1992). *Repeated Games with Complete Information*, in *Hanbook of Game Theory with Economic Applications*, volume 1. Elsevier Science Publishers.
- Steiner, J., Stewart, C., and Matejka, F. (2017). Rational inattention dynamics: Inertia and delay in decision-making. *Econometrica*, 84(2):521–553.
- Sutivong, A. (2003). Channel capacity and state estimation for state-dependent channel. *Ph.D. dissertation, Stanford Univ., Palo Alto, CA, USA*.
- Sutivong, A., Chiang, M., Cover, T. M., and Kim, Y.-H. (2005). Channel capacity and state estimation for state-dependent Gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1486–1495.
- Taneva, I. (2019). Information design. *American Economic Journal: Microeconomics*, 11(4):151–85.
- Tang, X., Liu, R., Spasojevic, P., and Poor, H. (2009). On the throughput of secure hybrid-ARQ protocols for Gaussian block-fading channels. *IEEE Transactions on Information Theory*, 55(4):1575–1591.
- Tian, C., Bandemer, B., and Shitz, S. S. (2015). Gaussian state amplification with noisy observations. *IEEE Transactions on Information Theory*, 61(9):4587–4597.
- Tsakas, E. and Tsakas, N. (2021). Noisy persuasion. *Games and Economic Behavior*, 130:44–61.
- Tung, S. (1978). *Multi-terminal source coding*. Ph. D. Thesis.
- Tutuncuoglu, K., Ozel, O., Yener, A., and Ulukus, S. (2014). State amplification and state masking for the binary energy harvesting channel. In *Proceedings of the IEEE Information Theory Workshop (ITW)*, pages 336–340.
- van der Meulen, E. (1977). A survey of multi-way channels in information theory: 1961-1976. *IEEE Transactions on Information Theory*, 23(1):1–37.
- Vellambi, B. N., Kliewer, J., and Bloch, M. R. (2015). Strong coordination over multi-hop line networks. In *Proceedings of the IEEE Information Theory Workshop - Fall (ITW)*, pages 192–196.
- Vellambi, B. N., Kliewer, J., and Bloch, M. R. (2016). Strong coordination over a line when actions are markovian. In *Proceedings of the Annual Conference on Information Science and Systems (CISS)*, pages 412–417.

- von Stackelberg, H. (1934). *Marketform und Gleichgewicht*. Oxford University Press.
- Vora, A. S. and Kulkarni, A. A. (2020). Achievable rates for strategic communication. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 1379–1384.
- Vora, A. S. and Kulkarni, A. A. (2020). Information extraction from a strategic sender: The zero error case. [on-line] available: <https://arxiv.org/abs/2006.10641>.
- Vu, M. T., Oechtering, T. J., and Skoglund, M. (2019). Hierarchical identification with pre-processing. *IEEE Transactions on Information Theory*, 66(1):82–113.
- Wang, L. and Wornell, G. W. (2020). Communication subject to state obfuscation. In *Proceedings of the International Zurich Seminar on Information and Communication (IZS 2020)*. ETH Zurich.
- Weissman, T. and Ordentlich, E. (2004). The empirical distribution of rate-constrained source codes. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, page 464.
- Weissman, T. and Ordentlich, E. (2005). The empirical distribution of rate-constrained source codes. *IEEE Transactions on Information Theory*, 51(11):3718–3733.
- Winter, A. (2002). Compression of sources of probability distributions and density operators. [on-line] available: <http://arxiv.org/abs/quant-ph/0208131>.
- Witsenhausen, H. (1968). A counterexample in stochastic optimum control. *SIAM J. Control*, 6:131–147.
- Witsenhausen, H. (1976). The zero-error side information problem and chromatic numbers (corresp.). *IEEE Transactions on Information Theory*, 22(5):592–593.
- Wyner, A. (1975). The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179.
- Wyner, A. D. (1978). The rate-distortion function for source coding with side information at the decoder-II: General sources. *Information and Control*, 38:60–80.
- Wyner, A. D. and Ziv, J. (1976). The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22(1):1–11.
- Yamamoto, H. (1981). Source coding theory for cascade and branching communication systems. *IEEE Transactions on Information Theory*, 27(3):299–308.
- Yamamoto, H. (1982). Wyner-Ziv theory for a general function of the correlated sources (corresp.). *IEEE Transactions on Information Theory*, 28(5):803–807.
- Yassaee, M. H., Gohari, A., and Aref, M. R. (2015). Channel simulation via interactive communications. *IEEE Transactions on Information Theory*, 61(6):2964–2982.
- Yüksel, S. and Başar, T. (2013). *Stochastic Networked Control Systems Stabilization and Optimization under Information Constraints*. Systems & Control: Foundations & Applications, Birkhäuser New York, NY.