# Essays in Microeconomics

Enrico Zanardo

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Essays in Microeconomics

Enrico Zanardo

This dissertation analyzes problems related to the the economics of incomplete information and to the theory of matching markets. Chapter 1 defines a family of functions that measure the distance between opinions; Chapter 2 investigates how to measure the cost of an experiment; and Chapter 3 studies a model of two-sided matching with countably many agents.

Chapter 1 introduces six axioms that a measure of disagreement should satisfy, and characterizes all the functions that satisfy them. The disagreement measures characterized generalize the Renyi divergences, and include the Kullback-Leibler divergence and the Bhattacharyya distance. Two applications are then studied. The first application provides a necessary and sufficient condition under which public information reduces expected disagreement between Bayesian agents. The second application shows that the measures of disagreement here defined are useful to understand trading under heterogeneous beliefs. Trade volume and gains from trade are increasing in some of the measures of disagreement.

Chapter 2 introduces seven postulates for a cost of information function. The main result of this chapter is the proof that there exists a unique function that satisfies these postulates. Differently from the cost functions commonly used, the function found in Chapter 2 is independent of the experimenter's beliefs, and it is additive in independent experiments. Similarly to other cost functions, it is increasing in the informativeness of

the experiment, and it is separable in the signal realizations.

Chapter 3 analyzes two-sided one-to-one matching with countably infinite agents. It shows that the set of stable matching is non-empty if and only if agents' preferences admit a maximum on all subsets. This requires generalizing the Deferred Acceptance algorithm, which also allows to find the man-optimal and woman-optimal stable matchings. It is then shown that, like in the finite model, the set of stable matchings is a complete lattice under the preferences induced by men (or women). Unlike in finite models, the set of matched agents may vary across stable matchings and some implications for dynamic matching markets are discussed.

# *Contents*

# *Acknowledgements*

I would like to thank my advisor, Professor Navin Kartik, who has supervised me during the PhD program. He rigorously evaluated my research, and has directed me toward ambitious and challenging projects. I would also like to thank the members of the defense committee, Professors Yeon-Koo Che, Pietro Ortoleva, Rajiv Sethi, and Paolo Siconolfi. They have always been available to me and it has been an honor to learn from them. Thanks also to all the professors and classmates who helped me succeed in the PhD program, for their help has been invaluable.

My warmest thanks go to Emily, who had the misfortune of meeting me during the most stressful period of the PhD program. Even so, she has provided me with unlimited support and I am looking forward to paying back the favor. I would also like to thank my family, Alberto, Zena, and Stefano for their encouragement. It goes without saying that without them I would have never been able to complete this program. Thanks also to Marian, Warren, Sydelle, Jo, Alex, and Kate for making me feel at home in New York.

Finally, I would like to deeply thank Shane, Jeanne, and Evan for all the time spent together, and for helping me transition smoothly into American culture. I would also like to thank all of my friends and in particular Paolo, Davide, Lucia, Andrea, and Manuel. It is a privilege to consider them friends.

A mia mamma.

Chapter 1

# *How to measure disagreement?*[1]

## 1.1 Introduction

*Measure what is measurable and make measurable what is not so.*

Attributed to Galileo Galilei (1564-1642)

People form opinions about uncertain events in any number of contexts: is global warming happening? Who will win the next elections? Will a business idea be successful? In all these situations, different agents hold different opinions and disagreement is ubiquitous. Mathematically, opinions, or beliefs, are defined as probability distributions. While mathematicians have defined numerous metrics on probability distributions, social scientists have not agreed on a metric that suitably quantifies opinion disagreement.

Besides being useful to social scientists, a quantitative measure of disagreement can be used in several practical problems. Political leaders can use it to measure polarization

[1]I am very grateful to Navin Kartik for his help and support. I would also like to thank Douglas Bernheim, Yeon-Koo Che, Tri Vi Dang, Mark Dean, Prajit Dutta, Emanuele Gerratana, Hari Govindan, Marina Halac, Harrison Hong, Pietro Ortoleva, Andrea Prat, Bernard Salanié, Jose Scheinkman, Rajiv Sethi, Paolo Siconolfi, Philipp Strack, Teck Yong Tan, Xingye Wu, and Alberto Zanardo for their invaluable feedback. Jeonghyun Jennifer Suh provided excellent research assistance.

of opinions in a community; managers can employ disagreement measures to form teams whose opinions are aligned (if a project requires consistent views), or misaligned (if one thinks that different opinions lead to more creative solutions); financial institutions can use it to measure the distance among investors' beliefs.

The goal of this chapter is to provide a quantitative measure of disagreement based on first principles, i.e., based on reasonable properties (axioms) one would like such a measure to satisfy. The main theorem characterizes all the functions that satisfy the axioms and the two applications show the usefulness of our measures of disagreement in two models. First, we prove that Bayesian agents expect to disagree less after observing the same piece of public information (under some conditions). Second, we show that trade of contingent assets among agents with heterogeneous beliefs is proportional to disagreement, as measured by some of our functions.

The family of functions we identify as disagreement measures includes several well-known divergence measures such as the Kullback-Leibler divergence and the Bhattacharyya distance. These measures, originally introduced in the information theory literature, are now used in several areas of economic research. The Kullback-Leibler divergence has been used to parametrize the cost of information in rational inattention models (see Sims (2003) and Sims (2010)). Hansen et al. (2014) use the Kullback-Leibler divergence and the Bhattacharyya coefficient to measure similarity in the content of speeches. Eliaz and Spiegler (2016) find a condition on the Bhattacharyya coefficient that implies the existence of a direct mechanism that allocates firms to search pools, in a Bayesian networks.

Let us describe our approach to measuring disagreement with an example. Suppose that three agents, Ann, Bob, and Carl, hold different opinions about four options, $1$, $2$, $3$, and $4$. The following table summarizes their opinions, by the probability they ascribe to each option being the best:

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Ann ($p_A$) | 8/10 | 1/10 | 1/90 | 8/90 |
| Bob ($p_B$) | 8/10 | 2/10 | 0 | 0 |
| Carl ($p_C$) | 9/10 | 1/10 | 0 | 0 |

Ann and Bob have different beliefs, more precisely they agree on the probability of state $1$ and they disagree on the likelihood of states $2, 3, 4$. Therefore, Ann and Bob agree on the reduced state space that is obtained by merging states $2, 3, 4$ into the same event $\{2, 3, 4\}$. In this new state space, the three beliefs are:

| | 1 | $\geq 2$ |
|---|---|---|
| Ann ($p_A$) | 8/10 | 2/10 |
| Bob ($p_B$) | 8/10 | 2/10 |
| Carl ($p_C$) | 9/10 | 1/10 |

Any metric $d$ (Euclidean distance, Total Variation, etc.[2]) yields that disagreement between Ann and Bob in the original state space is larger than their disagreement in the reduced space, as $d_{\{1,\geq 2\}}(p_A, p_B) = 0 < d_{\{1,2,3,4\}}(p_A, p_B)$. More generally, it is a desirable property of a disagreement measure to decrease whenever some states are merged into the same event, because by doing so any difference of opinions on those states is

[2]For any two beliefs $(p_1, \ldots, p_n)$, $(q_1, \ldots, q_n)$, the Euclidean distance is defined as $d(p, q) = \sqrt{\sum_{j=1}^{n}(p_j - q_j)^2}$, and the Total Variation is defined as $d(p, q) = \frac{1}{2}\sum_{j=1}^{n}|p_j - q_j|$.

not measured (the event " $\geq 2$" does not capture the fact that Ann and Bob assign different probabilities to states $2, 3, 4$). Consider then Ann and Carl: do they disagree less after merging states $2, 3, 4$? Not according to the Euclidean distance, because on the state space $\{1, 2, 3, 4\}$ their Euclidean distance is approximately $0.13$ and on the reduced state space $\{1, \{2, 3, 4\}\}$ it is larger than $0.14$. Our Axiom 4 imposes that disagreement between two agents cannot increase if we merge states, so this example shows that the Euclidean distance does not satisfy Axiom 4.

A distance that satisfies this property is the Hellinger distance, which is defined as: $d(p, q) = \sqrt{1 - \sum_i \sqrt{p_i q_i}}$. Nonetheless, there is another property, which we will impose, that the Hellinger distance does not satisfy. Observe that the agents' opinions in the example can be interpreted as the product of two independent opinions: i) is the best option an odd or an even number? ii) is the best option in $\{1, 2\}$ or in $\{3, 4\}$?

| | Odd | Even |
|---|---|---|
| Ann ($p_A$) | 8/9 | 1/9 |
| Bob ($p_B$) | 8/10 | 2/10 |
| Carl ($p_C$) | 9/10 | 1/10 |

| | {1,2} | {3,4} |
|---|---|---|
| Ann ($p_A$) | 9/10 | 1/10 |
| Bob ($p_B$) | 1 | 0 |
| Carl ($p_C$) | 1 | 0 |

We will require that whenever opinions can be written as a product of independent marginals, the disagreement is additively separable in issues (Axiom 5). In the example, the disagreement on the state $\{1, 2, 3, 4\}$ should be the sum of disagreement on the "Odd vs. Even" issue and the "$\{1, 2\}$ vs. {3,4}" issue. This is not the case for the the Hellinger distance, but if we instead consider the Bhattacharyya distance, $d(p, q) = -\log\left(\sum_i \sqrt{p_i q_i}\right)$, we have that total disagreement can be written as the sum of disagreement across issues.

Neither of the axioms we just described implies that on the "Odd vs. Even" issue, Bob should disagree less with Ann than with Carl (because $8/9$, the belief that Ann ascribes to the number being odd, lies between $8/10$ and $9/10$, respectively Bob's and Carl's belief). This will be assumed in Axiom 3, that implies that whenever an agent's opinion $p_A$ is a convex combination of the opinions of other two agents $p_B$ and $p_C$, then $d(p_B, p_A) \leq d(p_B, p_C)$.

In addition, we will assume in Axiom 1 that two agents with the same opinion have $0$ disagreement (Bob and Carl have $0$ disagreement on the "$\{1, 2\}$ vs. {3,4}" issue). We will also impose that the labelling of the state does not affect disagreement: if we permutate the columns of the tables above, disagreement among agents does not change (Axiom 2). We conclude with a separability axiom that restricts the way in which disagreement on state $i$ affects disagreement on other states (Axiom 6).

These axioms allow us to fully characterize the set of functions that measure disagreement (Theorem 1). The axioms we defined are based on geometric properties of opinions, and they are not explicitly tied to economic applications of disagreement. In the last section, then, we consider two models of incomplete information that show how our disagreement measures are useful to understand the interaction of rational agents with heterogeneous beliefs.

We first analyze how disagreement between two Bayesian agents changes when they observe a piece of public information. It will generically be the case that observing a signal realization might increase their disagreement. Nonetheless, disagreement will decrease *on average* if the disagreement measure satisfies a continuity axiom (Axiom 7). In other words, any rational agent expects public information to reduce disagreement with any other agent. In fact something stronger is true: if two experiments are ranked by statistical sufficiency, the more informative one will induce lower expected disagreement than the less informative one. This result matches the intuitive understanding of disagreement: whenever observing more precise public information, agents' opinions should converge. This would not be the case with most metrics on probability distributions, (Euclidean distance, Total Variation, any norm, etc.) whereas it is implied by some of our measures of disagreement.

In the second application, we show that trade in contingent assets among agents with heterogeneous beliefs is increasing in agents' disagreement. The relation between trade and disagreement has been documented extensively, but most metrics on probability distribution would not imply that more disagreement increases volume of trades. In our application, we show that if agents have constant relative risk aversion, the volume of trades is proportional to their disagreement. We conclude by showing that this result holds for any utility function, if the disagreement among agents is small.

### Related Literature

The set of results that are closest to those of this chapter come from the information theory literature. Even though those results are not motivated by the problem of measuring disagreement, they define functions that measure the distance between probability distributions, and often characterize them axiomatically. Rényi (1961) characterizes a family of divergence measures that is similar to our measures of disagreement. Some particular example of our measures are the Kullback-Leibler divergence (Kullback and Leibler (1951));

the Bhattacharyya distance (Bhattacharyya (1946)); and the logarithm of the Chernoff coefficients (Chernoff (1952)). Csiszár (2008) reviews different axiomatizations of some of these divergence measures, and comments on the axioms assumed in the literature. We postpone a detailed analysis of the overlapping between our measures of disagreement and well-known divergence measures to subsection 1.2. In Appendix A.4, we describe Renyi's axiomatization in detail, and comment on the differences with our setting.

More broadly, our results are related to economic models analyzing the interactions of agents with heterogeneous beliefs. A non-exhaustive list of papers in this literature includes: Mankiw et al. (2003) in the macroeconomics literature; Harrison and Kreps (1978) and Varian (1989) in the finance literature; Piketty (1995) in the political economy literature; Morris (1994) in the trade theory literature; Van Dan Steen (2010) in the firm theory literature; and more generally in the applied theory literature: Yildiz (2004), Che and Kartik (2009), Sethi and Yildiz (2012), Alonso and Camara (2016), Polemarchakis (2016), Sethi and Yildiz, etc. This chapter provides a way to measure the distance between two opinions, and therefore it can be related to papers studying the interaction of agents with heterogeneous beliefs.

Our application to the effect of public information on disagreement is related to the literature in political polarization (Sunstein (2002), Dixit and Weibull (2007), and Baliga et al. (2013)). Kartik et al. (2015) find conditions under which "information validates the prior", i.e. under which a Bayesian agent expects the posteriors of other agents to approach hers, as more information is observed. They find that this is the case if the experiment satisfies the Monotone Likelihood Ratio Property, and priors are Likelihood Ratio ranked. In our application, we study a similar questions, though instead of comparing expected posteriors, we analyze expected disagreement. Furthermore, we analyze our measures of disagreement, whereas Kartik et al. (2015) consider the expectations of monotone functions. The sufficiency ranking on information structures we consider is defined in Blackwell (1951) and Blackwell (1953). Other authors have investigated the relation between Blackwell's sufficiency and divergence measures (Taneja (1987), Kailath (1967), and Chambers and Healy (2010)).

The second application we study relates disagreement of investors to the volume of trade. The empirical relevance of disagreement on volume of trades has been shown, for example, in Cragg and Malkiel (1982), Kandel and Pearson (1995), Hong and Stein (2007), Cookson and Niessner (2016). Theoretically, many papers have proposed parametric models to explain those regularities: Kim and Verrecchia (1991), Harris and Raviv (1993), Kim and Verrecchia (1994), Kandel and Pearson (1995), and Hong and Stein (2007). Their results are related to ours in terms of motivation, but they differ in their setting. The model that is closest to ours is that of Varian (1985), who analyzes asset prices and volume of trades in a market with contingent assets. Similarly to us, he imposes no distributional assumption on the beliefs of the traders, but differently from us, he defines increasing disagreement as a *mean-preserving spread* of agents' beliefs. Owing to our measures of disagreement, we will be able to obtain the effect on trade of *any* change in beliefs, not only mean-preserving spreads.

## 1.2 Main Result

In this section, we introduce the model, axioms and main characterization of the chapter. We then study the properties of the family of disagreement functions we characterize in Theorem 1. The discussion and motivation of the axioms is postponed to Section 1.3.

### Model, Axioms and Main Theorem

Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ be a finite set of unknown states of the world. Define $\Delta(\Theta) := \{p \in \mathbb{R}^n_+ | \sum_j p_j = 1, p_j \geq 0, \forall j\}$ to be the set of beliefs on $\Theta$, and $\Delta^\circ(\Theta) = \{p \in \Delta(\Theta) \mid p_j > 0,\ \forall j\}$ to be its interior. We will denote by $p$ and $q$ two typical beliefs in $\Delta(\Theta)$, and let $p_i$ or $p(i)$ be $p$'s $i$-th component. The scope of this chapter is defining a function

$$D_\Theta : \Delta(\Theta) \times \Delta(\Theta) \to \mathbb{R}^+ \cup \{+\infty\}$$

that represents disagreement between two beliefs on $\Theta$. Observe that we allow disagreement to be infinite, i.e. the range of $D_\Theta$ includes $\{+\infty\}$. We will assume that $D_\Theta$ is three times continuously differentiable and finite on $\Delta^\circ(\Theta) \times \Delta^\circ(\Theta)$.

We say that a function $D_\Theta$ is a disagreement function if it satisfies the following axioms:

**Axiom 1** (Zero Disagreement). *For all $\Theta$ and $p, q \in \Delta(\Theta)$:*

$$D_\Theta(p, q) = 0 \Leftrightarrow p = q.$$

Plainly, two agents $p, q$ have zero disagreement $D(p, q) = 0$ if and only if they have the same opinion, $p = q$.

**Axiom 2** (Anonymity of the State Space). *Consider two state spaces* $\Theta_1, \Theta_2$*. If* $\gamma : \Theta_1 \to \Theta_2$ *is a bijection, then for all* $p, q \in \Delta(\Theta_1)$*:*

$$D_{\Theta_2}(p \circ \gamma^{-1}, q \circ \gamma^{-1}) = D_{\Theta_1}(p, q),$$

*where* $p \circ \gamma^{-1}$ *is the distribution on* $\Theta_2$ *defined by* $p \circ \gamma^{-1}(\theta_2) := p(\gamma^{-1}(\theta_2))$.

Axiom 2 implies that disagreement only depends on the cardinality of $\Theta$, $n = |\Theta|$ (see Lemma 2). Therefore, in the next axioms, we will denote by $D_n$ a disagreement function on a simplex of dimension $n$, $\Delta_n := \{x \in \mathbb{R}_n \mid \sum_i x_i = 1, x_j \geq 0\}$, regardless of the underlying state space $\Theta$.

If we consider permutations of the state space $\Theta$ ($\gamma : \Theta \to \Theta$), we obtain that Axiom 2 implies that disagreement $D_\Theta$ does not depend on the structure of the state space (order of the states, metric on the states, etc.). We discuss the implications of this axiom in detail in Subsection 1.3. Observe that the metrics that are typically used in $\mathbb{R}^n$ (Euclidean distance, Total Variation, p-norms) all satisfy Axiom 2.

**Axiom 3** (In Betweenness). *For all* $p^1, p^2, q^1, q^2 \in \Delta_n$*:*

$$D_n(\lambda p^1 + (1 - \lambda)p^2, \lambda q^1 + (1 - \lambda)q^2) \leq \max\{D_n(p^1, q^1), D_n(p^2, q^2)\},$$

*for all* $\lambda \in [0, 1]$.

Axiom 3 implies that the distance of two beliefs in the convex combination of four beliefs $p^1, q^1, p^2, q^2$ cannot be larger than the maximum pairwise disagreement. As a special case, we obtain some natural properties of disagreement: if $p^1 = p$ and $p^2 = q^1 = q^2$, Axiom 3 implies that $D_n(\lambda p + (1-\lambda)q, q) \leq D_n(p, q)$. If $p = p^1 = p^2$, we find that $D_n(p, \lambda q^1 + (1-\lambda)q^2) \leq \max\{D_n(p, q^1), D_n(p, q^2)\}$, which is equivalent to assuming that the balls in the topology induced by $D_n$ are convex.

**Axiom 4** (Coarsening). *For all* $p, q \in \Delta_n$*:*

$$D_{n-1}((p_1 + p_2, \ldots, p_n), (q_1 + q_2, \ldots, q_n)) \leq D_n((p_1, p_2, \ldots, p_n), (q_1, q_2, \ldots, q_n)).$$

This axiom implies that disagreement cannot increase after two states are merged. As observed in the introduction, this axiom is not satisfied by the Euclidean distance (unlike the previous axioms).

For any $n, m \in \mathbb{N}$, and for any pair of beliefs $p \in \Delta_n$ and $q \in \Delta_m$, we denote by $p * q \in \Delta_{nm}$ the belief, on a state space with $nm$ elements, with independent marginals $p$ and $q$. Formally:

$$p * q = (p_1 q_1, \ldots, p_1 q_m, \ldots, p_j q_1, \ldots, p_j q_m, \ldots, p_n q_1, \ldots, p_n q_m).$$

**Axiom 5** (Independence). *For all* $n, m \in \mathbb{N}$*, and all* $p^{(1)}, p^{(2)} \in \Delta_n$ *and* $q^{(1)}, q^{(2)} \in \Delta_m$*:*

$$D_{nm}(p^{(1)} * q^{(1)}, p^{(2)} * q^{(2)}) = D_n(p^{(1)}, p^{(2)}) + D_m(q^{(1)}, q^{(2)}).$$

This axiom means that if two agents consider two issues independent, then their disagreement on the product space is the sum of the disagreement across issues. This axiom is a departure from the previous ones in that it imposes a cardinal property of disagreement. In Subsection 1.3, we comment on how to relax this axiom to an ordinal property.

For all $i = 1, \ldots n-1$ and any two agents $p, q \in \Delta_n^\circ$ define the derivative of disagreement in the direction $i$ to $i+1$:

$$\partial_i D_n(p,q) := \lim_{\epsilon \to 0} \frac{D_n(p,(q_1,\ldots,q_i-\epsilon,q_{i+1}+\epsilon,\ldots,q_n)) - D_n(p,q)}{\epsilon}.$$

In words, $\partial_i D_n(p,q)$ denotes the marginal change in disagreement as the beliefs $q$ increase the probability of state $i+1$ and decrease the probability of state $i$. For any two states $i$ and $j$ the quantity $\frac{\partial_i D_n(p,q)}{\partial_j D_n(p,q)}$ then denotes the Marginal Rate of Substitution (MRS) of disagreement.

**Axiom 6** (MRS of Disagreement). *Consider any pair of states $i, j \in \{1, \ldots, n-1\}$. If $\partial_j D_n(p,q) \neq 0$, then:*

$$\frac{\partial_i D_n(p,q)}{\partial_j D_n(p,q)} = g(p_i, p_{i+1}, p_j, p_{j+1}; q_i, q_{i+1}, q_j, q_{j+1}),$$

*that is, such ratio does not depend on the belief on any other state.*

This axiom can be interpreted as imposing a separability property across states. The effect of a change in states $i, i+1, j, j+1$ on disagreement is independent of the beliefs on other states. Also, observe that Axiom 6 imposes a local condition. In Subsection 1.3, we provide a global condition that implies our axiom, and explains why Axiom 6 can be thought of as a local separability condition.

The following theorem, which is our main result, characterizes the functions that satisfy the above axioms.

**Theorem 1.** *The only functions $D_\Theta$ that satisfy Axioms 1, 2, 3, 4, 5, 6 have the following functional form (for all $\Theta$ finite, and all $p, q \in \Delta(\Theta)$):*[3]

1. *either:*

$$D_\Theta(p,q) = a \sum_{\theta \in \Theta} p(\theta) \log\left(\frac{p(\theta)}{q(\theta)}\right) + b \sum_{\theta} q(\theta) \log\left(\frac{q(\theta)}{p(\theta)}\right), \tag{1.1}$$

*for some $a, b \geq 0$ (not both zero);*

2. *or:*

$$D_\Theta(p,q) = a \log\left(\sum_{\theta} p(\theta)\left(\frac{p(\theta)}{q(\theta)}\right)^{z-0.5}\right) \quad \textit{where} \quad \begin{cases} a > 0 & \textit{if } |z| > 0.5 \\ a < 0 & \textit{if } |z| < 0.5 \end{cases} \tag{1.2}$$

*for $z \in \mathbb{R} \setminus \{-0.5, 0.5\}$.*

*Proof of Theorem 1.* The proof is in Appendix A.1.

□

*Remark* 1. Notice that the functions $D(p,q)$ are well-defined and finite for $p, q \in \Delta_n^\circ$. There is a unique way to extend these functions to $\Delta_n$ without violating any of the axioms. This extension uses the following conventions: if $p_j = q_j = 0$ then $0^\alpha/0^\beta = 0$

[3]The expression is not well-defined for non fully-mixed beliefs. See Remark 1 after the Theorem for a clarification.

and $0\log(0/0) = 0$. If $p_j > 0 = q_j$ then for all $\alpha, \beta > 0$, $p_j^\alpha/0 = +\infty$; $0/p_j^\beta = 0$; $p_j \log(p_j/0) = +\infty$; $0\log(0/p_j) = 0$.

## Analysis of the Disagreement Functions

We will denote by $Supp(p)$ the support of the distribution $p$: $Supp(p) = \{\theta \in \Theta \mid p_\theta > 0\}$. Unless specified otherwise, we will consider $p, q \in \Delta_n^\circ$, which implies that for all $j$, $p_j/q_j$ is finite (and then so is $D_n(p,q)$). Since in this section we will fix $\Theta$, we will drop the index $n$ writing simply $D(p,q)$ whenever unambiguous.

Notice that we have not assumed the disagreement functions to be symmetric. The next proposition shows which measures are symmetric in $p$ and $q$:

**Proposition 1**. *The only disagreement functions that satisfy Axioms 1–6 and are such that $D(p,q) = D(q,p)$ for all $p, q$, are proportional to:*

- *the symmetric divergence:*

$$D(p,q) = \sum_j (p_j - q_j) \log\left(\frac{p_j}{q_j}\right) = \sum_j p_j \log\left(\frac{p_j}{q_j}\right) + \sum_j q_j \log\left(\frac{q_j}{p_j}\right);$$

- *the Bhattacharyya distance:*

$$D(p,q) = -\log\left(\sum_j \sqrt{p_j q_j}\right).$$

*Furthermore, the only symmetric disagreement function that is additively separable in the states*[4] *is the symmetric divergence, and the only symmetric disagreement function such that*

[4]A divergence metric is said to be additively separable in the states if $D(p,q) = \sum_j f_j(p_j, q_j)$, for some

$D(p,q) < +\infty$ *if and only if* $Supp(p) \cap Supp(q) \neq \emptyset$ *is the Bhattacharyya distance.*

The symmetric divergence is also called J-divergence, or Jeffreys divergence, after Harold Jeffreys who first introduced it in Jeffreys (1946). It is also sometimes referred to as symmetrized Kullback-Leibler divergence, as it can be obtained as $D_{KL}(p\|q)+D_{KL}(q\|p)$, where $D_{KL}(p\|q) = \sum_j p_j \log\left(\frac{p_j}{q_j}\right)$ is the Kullback-Leibler divergence.

We allow the disagreement functions to be asymmetric because disagreement between agent $p$ and $q$ need not be the same as disagreement between agent $q$ and $p$. Symmetry is assumed for *metrics*, as metrics capture the distance between objects, an objective property of the geometry of the space. On the other hand, disagreement measures involve beliefs and thus are subjective evaluations. While we do not model the subjective process behind a definition of disagreement, we do allow for asymmetric disagreement measures. Furthermore, in the applications of Section 1.4, expected disagreement decreases in information only for asymmetric disagreement measures; and expected volume of trades are also captured by asymmetric disagreement measures. Namely, an agent $p$ who trades with agent $q$ might expect to receive a larger amount good than $q$ expects to give her.

The next proposition illustrates the relation between the Likelihood Ratio (LR) order and our disagreement functions. Given an order $\preceq$ on $\Theta$ we say that $q$ likelihood ratio dominates $p$ (and write $p \leq_{LR} q$) if:

$$p(\theta)q(\theta') \leq p(\theta')q(\theta), \quad \forall \theta \preceq \theta'.$$

Recall that our state space $\Theta$ is not endowed with an order, so in the following theorem

functions $f_j : [0,1] \times [0,1] \to \mathbb{R}$.

we write that $p \leq_{LR} q \leq_{LR} r$ meaning that *there exists an order* $\succeq$ *on* $\Theta$ *such that*:

$$p_i q_j \leq q_i p_j \quad \text{and} \quad q_i r_j \leq r_i q_j, \quad \forall i \succeq j.$$

**Proposition 2.** *Let* $p, q, r \in \Delta_n^\circ$ *be beliefs ranked by Likelihood Ratio:*

$$p <_{LR} q <_{LR} r,$$

*then for any disagreement function we have that:*

$$D(p, q) \leq D(p, r).$$

Therefore, our measures of disagreement are compatible with the LR order on beliefs, regardless of the underlying order on $\Theta$. Since the likelihood ratio order is often related to First Order Stochastic Dominance (FOSD),[5] it is useful to notice that in general it is *not* true that $p <_{FOSD} q <_{FOSD} r$ implies $D(p, q) \leq D(p, r)$.

In order to study the disagreement functions characterized in Theorem 1, let us rewrite them in order to avoid redundancy. If a function $D$ satisfies Axioms 1–6, then so does $\alpha D$, for any $\alpha > 0$, and these functions are ordinally equivalent. Let us define a representative in each of these classes of equivalence:

[5]We say that $q$ first order stochastic dominates $p$ if:

$$\sum_{i=1}^{j} q_i \leq \sum_{i=1}^{j} p_i, \quad \forall j,$$

and write $p \leq_{FOSD} q$. It is well-known that LR implies FOSD, see e.g. Shaked and Shanthinkumar (2006).

$$D^z(p,q) := \begin{cases} \log\left(\sum_j p_j \left(\frac{p_j}{q_j}\right)^{z-0.5}\right) & \text{if } |z| > 0.5, \\ -\log\left(\sum_j p_j \left(\frac{p_j}{q_j}\right)^{z-0.5}\right) & \text{if } |z| < 0.5. \end{cases} \tag{1.3}$$

**Observation 1.** *For all $z \neq \tilde{z}$, the disagreement functions $D^z$ and $D^{\tilde{z}}$ are not ordinally equivalent, i.e. there exist $p, q, r \in \Delta_n$ such that:*

$$D^z(p,q) < D^z(p,r) \quad \textit{and} \quad D^{\tilde{z}}(p,q) > D^{\tilde{z}}(p,r).$$

There is a natural graphical interpretation of the functions $D^z$: the argument of the logarithm can be written as an average of a function of the likelihood ratio,

$$\sum_j p_j \left(\frac{p_j}{q_j}\right)^{z-0.5} = \sum_j p_j \phi_z\left(\frac{q_j}{p_j}\right), \quad \text{where} \quad \phi_z(x) = x^{0.5-z}.^{6} \tag{1.4}$$

Our disagreement functions can then be interpreted as average dispersion of the likelihood ratio between $p$ and $q$, as Figure 1.1 shows.[7] This implies that the key statistic for our measures of disagreement is the vector of likelihood ratios $(p_1/q_1, \ldots, p_n/q_n)$, and this property sets our measures apart from other metrics on the space of beliefs (all norms, Euclidean distance, Total Variation, etc.). Figure 1.2 shows graphically the function $\phi_z$,

[6]Observe that this average is computed with respect to $p$, but analogously we could have written it with respect to $q$, after rescaling the parameter $z$ accordingly. Formally:

$$\sum_j p_j \phi_z(q_j/p_j) = \sum_j p_j \left(\frac{p_j}{q_j}\right)^{z-0.5} = \sum_j q_j \left(\frac{q_j}{p_j}\right)^{-z-0.5} = \sum_j q_j \phi_{-z+1}(p_j/q_j).$$

[7]Notice also that having defined $\phi_z$ on $q_j/p_j$, for all $p, q$, $\sum_j p_j \frac{q_j}{p_j} = 1$, so we *normalized* the mean of the likelihood ratios. Therefore if $\frac{q_j}{p_j}$ is a "spread" of $\frac{q'_j}{p_j}$, it will be a *mean preserving spread.*

for $z \neq 0.5, -0.5$. Observe that if $|z| > 0.5$, $\phi_z(\cdot)$ is a convex function; while if $|z| < 0.5$, $\phi_z(\cdot)$ is concave. This explains the different sign in the definition of $D^z$, equation (1.3).

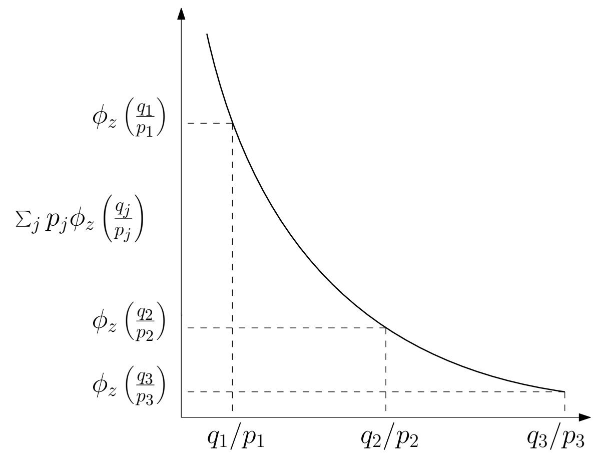
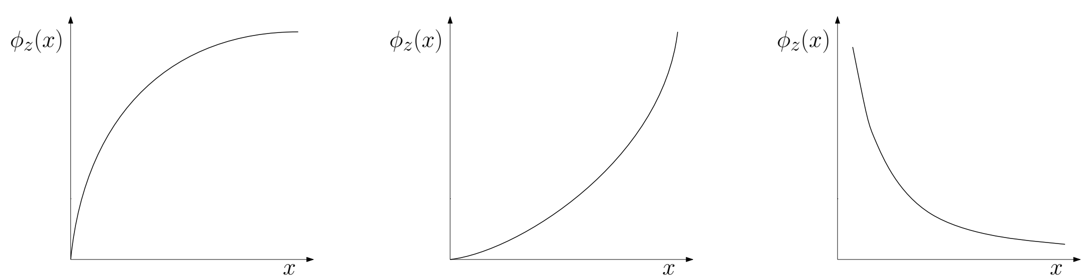



Figure 1.1: Graphical Interpretation of equation (1.4), for $z > 0.5$.




(a) $|z| < 0.5$, $\phi_z$ concave. (b) $z < -0.5$, $\phi_z$ convex. (c) $z > 0.5$, $\phi_z$ convex.

Figure 1.2: Plots of the functions $\phi_z$, for different values of $z \in \mathbb{R} \setminus \{0.5, 0.5\}$.

The following proposition summarizes some noteworthy properties of the functions $(D^z)_{z \in \mathbb{R} \setminus \{-0.5, 0.5\}}$.

**Proposition 3.** *For all $p, q \in \Delta_n^\circ$ and for all $z \neq 0.5, -0.5$:*

1. $D^z(p, q) = D^{-z}(q, p)$;
2. *for all $z \in \mathbb{R}$, $z \neq 0.5$:*

$$\frac{1}{z - 0.5} D^z(p, q) = \log(\|p(\theta)/q(\theta)\|^p_{z-0.5}),$$

*where the $w$-norm of a function $f : \Theta \to \mathbb{R}^+$, is defined by $\|f(\theta)\|_w^p := \left(\sum_i |f(\theta_i)|^w p(\theta_i)\right)^{1/w}$;*

3. *$D^z$ can be extended by continuity at $z = 0.5$ and $z = -0.5$:*

$$\lim_{z \to 0.5} \frac{D^z(p,q)}{z - 0.5} = \sum_j p_j \log\left(\frac{p_j}{q_j}\right) =: D^{0.5}(p,q)$$

$$\lim_{z \to -0.5} \frac{D^z(p,q)}{-z - 0.5} = \sum_j q_j \log\left(\frac{q_j}{p_j}\right) =: D^{-0.5}(p,q).$$

The first point of the proposition highlights a symmetry in the family disagreement functions: for all $z$, the disagreement functions $D^z$ is equivalent to $D^{-z}$ after inverting $p$ and $q$. The second point shows that $D^z$ can be interpreted as the rescaled logarithm of a norm. This allows us to compare the distance between two beliefs as $z$ changes, because $\|f\|_p^z \leq \|f\|_p^{z'}$ for all $z < z'$, and for all $f$. Finally, the third point shows that even though $D^z$ are defined on $\mathbb{R} \setminus \{-0.5, 0.5\}$, they can be extended to the whole real line, once an appropriately rescaled limit is considered. Notice that $D^{0.5}$ corresponds to the Kullback-Leibler divergence, as $D^{0.5}(p,q) = D_{KL}(p\|q)$, whereas $D^{-0.5}(p,q) = D_{KL}(q\|p)$.

We stated Proposition 3 only for $p, q$ in the interior of $\Delta_n$. The following Lemma clarifies how the functions $D^z$ differ at the boundary of $\Delta_n$.[8]

**Lemma 1.** *If $|z| < 0.5$, then $D^z(p,q) = +\infty$ if and only if $Supp(p) \cap Supp(q) = \emptyset$. If $|z| \geq 0.5$ then:*

- *if $z \geq 0.5$ then $D^z(p,q) = +\infty$ if and only if $Supp(p) \setminus Supp(q) \neq \emptyset$.*
- *if $z \leq -0.5$ then $D^z(p,q) = +\infty$ if and only if $Supp(q) \setminus Supp(p) \neq \emptyset$.*

[8]The boundary of $\Delta_n$ is defined by $\Delta_n \setminus \Delta_n^\circ$.

This Lemma has some important implications regarding the continuity of $D^z$ on $\Delta_n \times \Delta_n$. Recall that by assumption we have that all disagreement measures are smooth on $\Delta_n^\circ \times \Delta_n^\circ$, and hence continuous. This corollary shows when they are continuous also on the entire simplex $\Delta_n$. For a fixed $p$, we denote by $D^z(p,\cdot) : \Delta_n \to \mathbb{R}$ the function $q \mapsto D^z(p,q)$.

**Corollary 1.**

1. *if* $z \geq 0.5$, $D^z(p,\cdot) : \Delta_n \to \mathbb{R}$ *is continuous for all fixed* $p \in \Delta_n$*. Furthermore* $D^z$ *depends only on the states* $\theta \in Supp(p)$*:*

$$D^z(p,q) = \begin{cases} \log\left(\sum_{j\in Supp(p)} p_j \left(\frac{p_j}{q_j}\right)^{z-0.5}\right) & \textit{if } z > 0.5 \\ \sum_{j\in Supp(p)} p_j \log\left(\frac{p_j}{q_j}\right) & \textit{if } z = 0.5 \end{cases}.$$

   *On the other hand,* $D^z(\cdot,q) : \Delta_n \to \mathbb{R}$ *is continuous if and only if* $q \in \Delta_n^\circ$*.*

2. *if* $z \leq -0.5$*,* $D^z(\cdot,q) : \Delta_n \to \mathbb{R}$ *is continuous for all fixed* $q \in \Delta_n$*. Furthermore* $D^z$ *depends only on the states* $\theta \in Supp(q)$*:*

$$D^z(p,q) = \begin{cases} \log\left(\sum_{j\in Supp(q)} p_j \left(\frac{p_j}{q_j}\right)^{z-0.5}\right) & \textit{if } z > 0.5 \\ \sum_{j\in Supp(q)} p_j \log\left(\frac{p_j}{q_j}\right) & \textit{if } z = 0.5 \end{cases}.$$

   *On the other hand,* $D^z(p,\cdot) : \Delta_n \to \mathbb{R}$ *is continuous if and only if* $p \in \Delta_n^\circ$*.*

3. *If* $|z| < 0.5$*, then* $D^z : \Delta_n \times \Delta_n \to \mathbb{R}$ *is continuous, so it is continuous on both*

*variables separately, and:*

$$D^z(p,q) = -\log\left(\sum_{j\in Supp(p)\cap Supp(q)} p_j \left(\frac{p_j}{q_j}\right)^{z-0.5}\right),$$

*with the convention that* $\sum_{j\in\emptyset} = 0$.

Besides differing at the boundary of $\Delta_n$, the disagreement functions $D^z$ have very different properties also for $p,q \in \Delta_n^\circ$. The next proposition shows that for $|z| \geq 0.5$ the topology induced by $D^z$ is very different from that induced by any norm on $\Delta_n$. If $|z| < 0.5$, instead, the disagreement topology and the norm topology are equivalent. We state the proposition using the sup-norm on $\Delta_n$, which is defined as $\|x\|_\infty := \sup_j |x_j|$. Since $\Delta_n$ is finite dimensional, the same result holds for any norm.[9]

**Proposition 4.** *Let* $|z| \geq 0.5$ *then for any* $\epsilon, \delta > 0$ *(arbitrary small) there exist* $\bar{p}, \bar{q}, \underline{p}, \underline{q} \in \Delta_n^\circ$ *such that:*

$$\|\bar{p} - \bar{q}\|_\infty > \max_{x,y\in\Delta_n} \|x-y\|_\infty - \delta, \qquad \|\underline{p} - \underline{q}\|_\infty < \epsilon,$$

*and* $D^z(\underline{p}, \underline{q}) > D^z(\bar{p}, \bar{q})$.

*If* $|z| < 0.5$*, instead,* $D^z(p,q)$ *is uniformly continuous with respect to the metric induced by* $\|\cdot\|_\infty$*. Namely, for all* $\epsilon$ *there exist a* $\delta$ *such that if* $\|p-q\|_\infty \leq \delta$ *then* $D^z(p,q) < \epsilon$.

The first statement says that if we use disagreement measures $D^z$, with $|z| > 0.5$, we can find two pairs of beliefs such that the first pair is arbitrarily distant in the norm sense, and the second is arbitrarily close; and yet disagreement between the former is smaller

[9] All norms in finite dimensional spaces $\mathbb{R}^n$ are topologically equivalent, see Shores (2007).

than disagreement in the latter. On the other hand, if $|z| < 0.5$, the topology induced by $D^z$ is equivalent to that induced by any norm.

The apparently counterintuitive property of measures $D^z$ for $|z| > 0.5$ follows from the fact that our measures of disagreement are dispersions of the likelihood ratio (see equation (1.4)), unlike norms. To highlight this point consider the following example. Let $p^\epsilon = (1-\epsilon, \epsilon)$ and $q^\epsilon = (1-\epsilon^3, \epsilon^3)$. Plainly, as $\epsilon \to 0$ we have that $p^\epsilon, q^\epsilon \to (1,0)$ (in any norm), and so $\|p^\epsilon - q^\epsilon\| \to 0$. On the other hand:

$$D^z(p^\epsilon, q^\epsilon) = (1-\epsilon)\left(\frac{1-\epsilon}{1-\epsilon^3}\right)^z + \epsilon\left(\frac{\epsilon}{\epsilon^3}\right)^z \to +\infty, \quad \text{as } \epsilon \to 0,$$

for any $z > 0.5$. Therefore even though two beliefs converge to agreeing that the true distribution is $(1,0)$, their disagreement diverges. To rephrase this result in terms of volume of trade in financial markets, suppose that two agents with beliefs $p^\epsilon$ and $q^\epsilon$ trade an asset that pays $1$ unit of good in state $\theta_2$. Even as they converge to agreeing that state $\theta_2$ will not be realized (i.e. that the asset is worthless), they will still be willing to trade the asset, and in fact the volume of trade will diverge, if one agent's speed of convergence is exponentially larger than the other's.

Proposition 4 might suggest that the disagreement functions $D^z$ for $|z| < 0.5$ are preferable, in that they relate better to commonly used distances on $\mathbb{R}^n$. Nevertheless, the results we present in Section 1.4 will show that the case $|z| > 0.5$ satisfies other desirable properties of disagreement. For example, disagreement measures $D^z(p,q)$ with $z > 0.5$ decrease on average when agents observe the same piece of public information, and the ex-ante probability of a signal is taken to be $p$'s (see Theorem 3).

## Technical Literature Review

In this subsection, we will relate our disagreement measures to other divergence measures introduced in the information theory literature. Our measures overlap with several others in the literatures, so we cite the papers that introduced them and the ones that axiomatized them. The family that is most closely related to our disagreement measures is the Renyi's divergences. We defer a detailed description of Renyi's axiomatization and its differences from ours to Appendix A.4.

| Name | Definition |
|---|---|
| Renyi's divergences | $R_\alpha(p,q) = \begin{cases} \frac{1}{\alpha-1}\log\left(\sum_j p_j \left(\frac{p_j}{q_j}\right)^{\alpha-1}\right) & \alpha \neq 1 \\ \sum_j p_j \log\left(\frac{p_j}{q_j}\right) & \alpha = 1 \end{cases}$ |
| Chernoff coefficients | $C_\beta(p,q) = \sum_j p_j^\beta q_j^{1-\beta}$, $\beta \in (0,1)$ |
| Kullback-Leibler divergence | $D_{KL}(p\|q) = \sum_j p_j \log\left(\frac{p_j}{q_j}\right)$ |
| Symmetric divergence | $D_S(p,q) = \sum_j (p_j - q_j)\log\left(\frac{p_j}{q_j}\right)$ |
| Bhattacharyya distance | $D_B(p,q) = -\log\left(\sum_j \sqrt{p_j q_j}\right)$ |

| Name | Function $D^z$ |
|---|---|
| Renyi's divergences | $R_\alpha(p,q) = \begin{cases} \frac{1}{\alpha-1} D^{\alpha-0.5}(p,q) & \text{if } \alpha \neq 0,1 \\ D^{0.5}(p,q) & \text{if } \alpha = 1 \end{cases}$ |
| Chernoff coefficients | $C_\beta(p,q) = \exp(-D^{\beta-0.5}(p,q))$ |
| Kullback-Leibler divergence | $D_{KL}(p\|q) = D^{0.5}(p,q)$ |
| Symmetric divergence | $D_S(p,q) = D^{0.5}(p,q) + D^{-0.5}(p,q)$ |
| Bhattacharyya distance | $D_B(p,q) = D^0(p,q)$ |

Renyi's divergences were introduced and axiomatized in Rényi (1961), as a generalization of Kullback-Leibler divergence. The Chernoff coefficients (Chernoff (1952)) were introduced as a bound for the asymptotic efficiency of a test, and axiomatized in Kannappan and Rathie (1972). The Kullback-Leibler divergence (Kullback and Leibler (1951)) was also obtained as a bound on the error when testing $p$ versus $q$; it was axiomatized by several authors (Kullback and Khairat (1966), Campbell (1972), Kannappan and Rathie (1973)). The symmetric divergence was first studied by Jeffreys (1946) as a differential form invariant for all transformations of the distributions, and later axiomatized by Kannappan and Rathie (1988). The Bhattacharyya distance, Bhattacharyya (1946), was introduced to measure the divergence between multinomial samples, and it was never axiomatized (so far as we know). Csiszár (2008) reviews several different axiomatization and provides a guide through the different axioms assumed in the literature. This chapter differs from all the other axiomatizations both in terms of the axioms themselves, and in terms of the motivation.

Renyi's divergences differ from our disagreement measures in three ways: for $\alpha < 0$ the Renyi divergences are negative (unlike our disagreement measures); Renyi's divergences do not include $D^{-0.5}(p, q) = \sum_j q_j \log\left(\frac{p_j}{q_j}\right)$; and Renyi's divergence do not include all the positive linear combinations of $D^{0.5}(p, q)$ and $D^{-0.5}(p, q)$. More importantly, the axioms of Rényi (1961) are very different from ours. First off, Renyi assumes that divergences satisfy a strengthening of our independence axiom, which implies that divergence of an event is the logarithm of the ratio. Furthermore, Renyi assumes that the divergence be a generalized mean, i.e. that there exists an increasing function $g$ such that:

$$D(p,q) = g^{-1}\left(\sum_j p_j g(D(p_i,q_i))\right),$$

where $D(p_i,q_i)$ represents the divergence on state $i$. This assumption constraints the functional form of Renyi's divergences and therefore makes his axiomatization very different from ours. A more detailed comparison of the two approaches can be found in Appendix A.4.

## 1.3 Discussion of the Axioms

### Zero Disagreement

**Axiom 1** (Zero Disagreement). *For all* $\Theta$ *and* $p,q \in \Delta(\Theta)$*:*

$$D_\Theta(p,q) = 0 \Leftrightarrow p = q.$$

Since we assume $D$ to be weakly positive, this axiom says that agents with the same opinion have the minimum disagreement. Assuming that $D(p,q) > 0$ for $p \neq q$ amounts to *separate* different opinions.

## Anonymity of the state space

**Axiom 2** (Anonymity of the State Space). *Consider two state spaces* $\Theta_1, \Theta_2$. *If* $\gamma : \Theta_1 \to \Theta_2$ *is a bijection, then for all* $p, q \in \Delta(\Theta_1)$*:*

$$D_{\Theta_2}(p \circ \gamma^{-1}, q \circ \gamma^{-1}) = D_{\Theta_1}(p, q),$$

*where* $p \circ \gamma^{-1}$ *is the distribution on* $\Theta_2$ *defined by* $p \circ \gamma^{-1}(\theta_2) := p(\gamma^{-1}(\theta_2))$.

This axiom formalizes the idea that our disagreement functions are independent of the structure of the state space. In other words, our disagreement measures do not distinguish between any two state spaces with the same cardinality.

We impose this axiom because we are interested in disagreement functions that depend only on the relative probability that two agents assign to a particular state (without being concerned about what the state stands for, or how states are related). A lot of issues of economic relevance do involve a "structured" state spaces,[10] but we leave such analysis to a future project. The disagreement functions characterized in this chapter can be used also to measure distance of opinions on spaces with a structure, but the structure of the space will not be captured by them. The next lemma formalizes the fact that any $D_\Theta$ satisfying Axiom 2 depends only on the cardinality of the state space, $n = |\Theta|$:

**Lemma 2.** *If a family of disagreement functions* $D_\Theta$ *for any state space* $\Theta$ *satisfies Axiom 2*

[10]For example, disagreement over the price of a good tomorrow will depend on the labeling of the states.

*then so does the family* $(\tilde{D}_n)_{n\geq 2, n\in\mathbb{N}}$*, where:*

$$\tilde{D}_n : \Delta(\Theta) \times \Delta(\Theta) \to \mathbb{R}^+ \cup \{+\infty\}, \qquad \tilde{D}_n(p,q) := D_\Theta(p,q),$$

*for any* $\Theta$ *with* $|\Theta| = n$*.*

*Proof.* All proofs of the results in this section are in Appendix A.2. $\square$

Axiom 2 then allows to redefine the goal of the chapter as defining a family of functions $(D_n)_{n\geq 2}$, where $n$ is the cardinality of the state space. In the rest of the chapter, we will drop the dependence of $D_\Theta$ on $\Theta$.

## In Betweenness

Suppose that there is a set $A \subset \Theta$ on whose probability two agents agree, $p(A) = q(A)$, and with $p(A) \in (0,1)$. Define $p(\cdot|A)$ as the conditional belief given $A$ and $p(\cdot|A^c)$ the conditional belief given its complement. It is natural to assume that $D_n(p,q)$ be smaller than the maximum disagreement on the conditional beliefs (resp. $D_{|A|}(p(\cdot|A), q(\cdot|A))$ and $D_{|A^c|}(p(\cdot|A^c), q(\cdot|A^c))$). Formally:

$$D_n(p,q) \leq \max\{D_{|A|}(p(\cdot|A), q(\cdot|A)), D_{|A^c|}(p(\cdot|A^c), q(\cdot|A^c))\}, \tag{1.5}$$

for all $A \subset \Theta$ with $p(A) = q(A) \in (0,1)$. This amounts to imposing that if we *agree* on the probability of a event $A$, and observing such event leads us to reduce our disagreement, then *not observing* $A$ will have the effect of increasing disagreement.

Axiom 3 generalizes this property to any four beliefs $p^1, p^2, q^1, q^2$:[11]

**Axiom 3** (In Betweenness). *For all* $p^1, p^2, q^1, q^2 \in \Delta_n$*:*

$$D_n(\lambda p^1 + (1-\lambda)p^2, \lambda q^1 + (1-\lambda)q^2) \leq \max\{D_n(p^1, q^1), D_n(p^2, q^2)\},$$

*for all* $\lambda \in [0, 1]$.

Axiom 3 is equivalent to assuming $D_n : \Delta_n \times \Delta_n \to \mathbb{R} \cup \{+\infty\}$ quasi-convex. The next result shows that Axiom 3 implies three convexity properties of disagreement:

**Lemma 3.** *If* $D_n$ *satisfies Axiom 3 and Axiom 1 then:*

1. *for all* $p, q \in \Delta_n$ *and for all* $\lambda \in [0, 1]$*:*

$$D_n(p, \lambda p + (1-\lambda)q) \leq D_n(p, q).$$

2. *For all* $p, q, r \in \Delta_n$ *and* $\lambda \in [0, 1]$*:*

$$D_n(\lambda p + (1-\lambda)r, \lambda q + (1-\lambda)r) \leq D_n(p, q).$$

3. *For all* $p, q^1, q^2$ *and* $\lambda \in [0, 1]$*:*

$$D_n(p, \lambda q^1 + (1-\lambda)q^2) \leq \max\{D_n(p, q^1), D_n(p, q^2)\}.$$

[11] Another noteworthy particular case is that in which instead of considering an event $A$, we consider a public signal $s$. Then, defining $p^1 = p(\cdot|s)$, $p^2 = p(\cdot|s^c)$, $q^1 = q(\cdot|s)$, $q^2 = q(\cdot|s^c)$, and $\lambda = \mathbb{P}_p(s) = \mathbb{P}_p(s)$, we find that if observing a public signal reduces disagreement ($D(p(s), q(s)) < D(p, q)$), then *not* observing it must increase it: $D(p, q) \leq D(p(s^c), q(s^c))$ (if we agree on the ex-ante probability of the signal, $\mathbb{P}_p(s) = \mathbb{P}_p(s)$).

*Also, this last property is equivalent to assuming that the balls:*

$$B(p,\rho] := \{q \in \Delta_n \mid D_n(p,q) \leq \rho\}$$

*are convex for all $D_n$.*

As a corollary we have that the maximum disagreement between two agents in any convex set $C \subset \Delta_n$ must be reached at the boundary of $C$.

**Corollary 2.** *For every $C \subset \Delta_n$ convex:*

$$\sup_{p,q\in C} D_n(p,q) = D_n(\bar{p},\bar{q}),$$

*for some $\bar{p},\bar{q} \in \partial C$. This implies that for all $p,q \in \Delta_n$ we have that:*

$$D_n(p,q) \leq D_n(e^1,e^2) \quad (= D_n(e^j,e^i) \quad \forall\, i,j),$$

*where $e^1 = (1,0,0,\ldots,0)$ and $e^2 = (0,1,0,\ldots,0)$.*

## Coarsening

Consider a state space of cardinality $n$, and two beliefs $p,q \in \Delta_n$. If we merge states $\theta_1,\theta_2$ into a unique event $\{\theta_1,\theta_2\}$, we obtain two beliefs $p',q' \in \Delta_{n-1}$ defined by:

$$p \to p' = (p_1+p_2,p_3,\ldots,p_n) \quad \text{and} \quad q \to (q_1+q_2,q_3,\ldots,q_n).$$

This transformation has the effect of *eliminating* any disagreement on the states $\theta_1, \theta_2$ so we will impose that disagreement cannot increase if we coarsen the state space.

**Axiom 4** (Coarsening). *For all* $p, q \in \Delta_n$*:*

$$D_{n-1}((p_1 + p_2, \ldots, p_n), (q_1 + q_2, \ldots, q_n)) \leq D_n((p_1, p_2, \ldots, p_n), (q_1, q_2, \ldots, q_n)).$$

By induction, Axiom 4 implies that for any partition[12] $\mathcal{A} = (A_j)_j$ of $\Theta$ we have that:

$$D_{|\mathcal{A}|}(p_{\mathcal{A}}, q_{\mathcal{A}}) \leq D_n(p, q),$$

where $p_{\mathcal{A}} \in \Delta(\mathcal{A})$, and $p_{\mathcal{A}}(A_j) := \sum_{\theta \in A_j} p(\theta_j), \forall A_j \in \mathcal{A}$.

Notice that Axiom 4 allows to bound disagreement on a state space of cardinality $n$ with disagreement on a state space of lower cardinality. This sets it apart from Axioms 1–3 which instead involved only $D_n$ for a fixed $n$. The next axiom, Axiom 5, also allows us compare disagreement in different dimensions.

## Independence

Consider any two state spaces $\Theta_1, \Theta_2$ with $n = |\Theta_1|$ and $m = |\Theta_2|$ and the corresponding simplexes $\Delta_n, \Delta_m$. For any $p^{(1)} \in \Delta_n$ and $q^{(1)} \in \Delta_m$ we can define $r^{(1)} := p^{(1)} * q^{(1)} \in$

[12] A partition is a set $(A_j)_j$ of subsets of $\Theta$ such that:

$$A_j \cap A_i = \emptyset, \qquad \bigcup_j A_j = \Theta.$$

$\Delta(\Theta_1 \times \Theta_2)$ by:

$$r^{(1)}(\theta_1, \theta_2) := p^{(1)}(\theta_1)q^{(1)}(\theta_2).$$

In words, $r^{(1)}$ is defined as the joint distribution with independent marginals $p^{(1)}$ and $q^{(1)}$. Whenever two agents both believe that issues on state spaces $\Theta_1$ and $\Theta_2$ are independent, we will assume that the disagreement can be summed across state spaces. Formally:

**Axiom 5** (Independence). *For all $n, m \in \mathbb{N}$, and all $p^{(1)}, p^{(2)} \in \Delta_n$ and $q^{(1)}, q^{(2)} \in \Delta_m$:*

$$D_{nm}(p^{(1)} * q^{(1)}, p^{(2)} * q^{(2)}) = D_n(p^{(1)}, p^{(2)}) + D_m(q^{(1)}, q^{(2)}).$$

Axiom 5 is equivalent to assuming that for all $p^1, q^1, p^2, q^2 \in \Delta_n$:

$$D_{nm}(p^2 * r, q^2 * s) - D_{nm}(p^1 * r, q^1 * s) = D_n(p^2, q^2) - D_n(p^1, q^1) \quad \forall r, s \in \Delta_m. \tag{1.6}$$

Let us illustrate why equation (1.6) is a desirable property of a disagreement function. Suppose an experimenter wants to measure the belief of two agents on Global warming, and takes a survey in two consecutive years, year 1 and 2. Let $p^1$ and $p^2$ (resp. $q^1$ and $q^2$) be the *beliefs on global warming* of agent $P$ (resp. $Q$) in the two years. The experimenter's goal is to measure $D(p^2, q^2) - D(p^1, q^1)$, but other beliefs will be implicitly measured when a survey is taken. Say agent $P$ and $Q$ have different beliefs on the meaning of a word, and denote these beliefs by $r$ and $s$ respectively. Then the experimenter effectively measures (a statistic of) $p^1 * r$ and $q^1 * s$ in the first year, and $p^2 * r$ and $q^2 * s$ in the second year. Axiom 5 implies that if the experimenter can measure $r$ and $s$, too, such disagreement on the meaning of words will not affect the change in beliefs from year 1 to 2.

Notice how Axiom 5 can be broken down into two properties: first, it requires disagreement to be separable in *independent issues*; second, it imposes an additive structure across independent state spaces. The separability is quite natural: if both agent believe the issues to be independent it is conceivable to assume away *complementarities* between state spaces. The *additive structure* instead sets it apart from the previous axioms, as Axioms 1, 2, 3, 4 are all ordinal.[13]

As it has been shown in Observation 1, the disagreement measures we identify are *ordinally* different. We show here that Axiom 5 can be relaxed to an ordinal axiom that yields a similar representation theorem.

**Property 1** (Ordinal Independence). *We say that $D$ satisfies Ordinal Independence if both of the following conditions are satisfied:*

1. *Let $\underline{p}, \underline{q} \in \Delta_n$, $\bar{p}, \bar{q} \in \Delta_m$. If $D_n(\underline{p}, \underline{q}) \leq D_m(\bar{p}, \bar{q})$ then:*

$$D_{n \cdot n'}(\underline{p} * r, \underline{q} * s) \leq D_{m \cdot n'}(\bar{p} * r, \bar{q} * s),$$

   *for all $r, s \in \Delta_{n'}$.*

2. *Let $p, q \in \Delta_n$ and $r \in \Delta_m$. We assume that:*

$$D_n(p, q) = D_{n \cdot m}(p * r, q * r).$$

[13] We say that an axiom is ordinal if whenever $D$ satisfies it, so does any increasing transformation of $D$.

It is plain to see that a disagreement function satisfying Axiom 5 also satisfies Property 1. Besides being an ordinal property, Ordinal Independence also differs from Axiom 5 in that it is satisfied by constant disagreement functions. We will rule out such disagreement function by assuming that $D$ is not locally constant:

**Definition 1.** We say that $D$ is *not* locally constant if for all $p, q \in \Delta_n^\circ$ and for all $U_q$ neighborhoods of $q$ and $U_p$ neighborhood of $p$ there exists $p' \in U_p$ and $q' \in U_q$ such that:

$$D(p', q) \neq D(p, q) \neq D(p, q').$$

The following proposition shows that if a disagreement function satisfies Property 1 and Definition 1, then it is a strictly increasing transformation of a function that satisfies Axiom 5.

**Proposition 5.** *Let $D$ be a smooth measure of disagreement satisfying Axioms 1–4. If $D$ satisfies Property 1 and is not locally constant, there exists a strictly increasing function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ such that the disagreement function $\tilde{D} := \phi(D)$ satisfies Axiom 5.*

Therefore, assuming Axiom 5 instead of Property 1 amounts to choosing a convenient "cardinalization" of the orders represented by the disagreement functions of Theorem 1. For this reason, we will assume Axiom 5 instead of the weaker ordinal version. Let us derive the consequences of Axiom 5.

**Lemma 4.** *If $D_n$ satisfies Axioms 1–5, then $\forall p, q \in \Delta_n$:*

$$D_n(p, q) = D_{n+1}((p_1, \ldots, p_n, 0), (q_1, \ldots, q_n, 0)).$$

This lemma implies that including states that both agents dim impossible has no effect on disagreement. A consequence of this Lemma (and Axioms 1–5) is that if two agents' beliefs have disjoint support, then their disagreement must be infinite.

**Lemma 5**. *Let $p, q \in \Delta_n$ and suppose that $Supp(p) \cap Supp(q) = \emptyset$, meaning $p_j q_j = 0$, for all $j$. Then if $D$ satisfies Axioms 1–5, $D(p, q) = +\infty$.*

The next result shows how the statistic of interest when measuring disagreement on a state $\theta_j$ is given by the likelihood ratio. This is a key result in our characterization and it will simplify significantly the proof of Theorem 1. We break down the result in a proposition and a corollary. Proposition 6 shows that if the likelihood ratio on two states is the same, then the two states are undistinguishable in terms of disagreement; Corollary 3 states a direct consequence of the proposition.

**Proposition 6**. *Let $p, q \in \Delta_n^\circ$ be two beliefs, and suppose that $\frac{p_1}{q_1} = \frac{p_2}{q_2}$. Then:*

$$D_n(p, q) = D_{n-1}((p_1 + p_2, p_3, \ldots, p_n), (q_1 + q_2, q_3, \ldots, q_n)).$$

**Corollary 3**. *Fix a belief $p \in \Delta_n^\circ$ and consider the segment joining $q^1 = (q_1 + q_2, 0, q_3, \ldots, q_n)$ and $q^2 = (0, q_1 + q_2, q_3, \ldots, q_n)$:*

$$[q^1, q^2] = \{r \in \Delta_n \mid r = \lambda q^1 + (1 - \lambda) q^2, \lambda \in [0, 1]\}.$$

*The minimum of the distance between $p$ and $[q^1, q^2]$ is reached at the $q^* \in [q^1, q^2]$ satis-*

*fying:*

$$\frac{q_1^*}{q_2^*} = \frac{p_1}{p_2},$$

*And for all* $r, r' \in [q^1, q^2]$ *we have that if* $r \in [r', q^*]$ *then*

$$D_n(p, r) \leq D_n(p, r').$$

Figure 1.3 shows graphically the result of Corollary 3. Given any two vectors $p, q$, the dashed line passing by $q$ is the segment $[q^1, q^2]$ and the dashed line passing by $p$ that represents the set of beliefs $r$ such that $\frac{r_1}{r_2} = \frac{p_1}{p_2}$. The point at which they meet, $q^*$, represents the point on the segment with minimal distance to $p$, and for this reason we drew the ball of radius $D(p, q^*)$ around $p$ tangent to $[q^1, q^2]$.

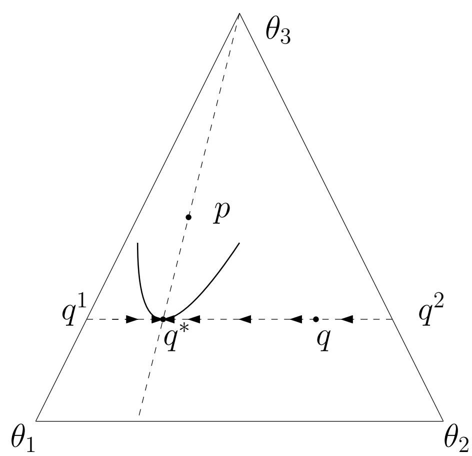



Figure 1.3: Graphical interpretation of Corollary 3 for $n = 3$. The curve represents the ball of radius $D(p, q^*)$ centered in $p$. The arrows imply that disagreement is decreasing as a belief approaches $q^*$ on the segment $[q^1, q^2]$.

### Marginal Rate of Substitution of Disagreement

This axiom constraints the local complementarities among states, by imposing a condition on the Marginal Rate of Substitution (MRS) of disagreement. In order to define a local condition on the disagreement functions, we define the derivatives along the simplex for $p, q \in \Delta_n^\circ$:

$$\partial_i D_n(p,q) := \lim_{\epsilon \to 0} \frac{D_n(p, (q_1, \dots, q_i - \epsilon, q_{i+1} + \epsilon, \dots, q_n)) - D_n(p,q)}{\epsilon} \quad \forall i = 1, \dots, n{-}1.$$

$\partial_i D_n(p,q)$ is the marginal change in $D_n$ as $q$ changes by reducing the likelihood of state $i$ and increasing that on state $i+1$. For all $(p,q)$ the derivatives $(\partial_i D_n(p,q))_{i=1}^{n-1}$ are a base for the differential of the function $D_n(p,\cdot) : \Delta_n^\circ \to \mathbb{R}^+$, $q \mapsto D_n(p,q)$. In other words, the derivative in any two directions $i, j$,

$$\lim_{\epsilon \to 0} \frac{D_n(p, (q_1, \dots, q_i - \epsilon, \dots, q_j + \epsilon, \dots, q_n)) - D_n(p,q)}{\epsilon}$$

is a linear combinations of the derivatives $(\partial_i D(p,q))_{i=1}^{n-1}$. Our last axiom constraints the relative change in disagreement when beliefs change on states $i, i+1, j, j+1$:

**Axiom 6** (MRS of Disagreement). *Consider any pair of states $i, j \in \{1, \dots, n-1\}$. If $\partial_j D_n(p,q) \neq 0$, then:*

$$\frac{\partial_i D_n(p,q)}{\partial_j D_n(p,q)} = g(p_i, p_{i+1}, p_j, p_{j+1}; q_i, q_{i+1}, q_j, q_{j+1}), \tag{1.7}$$

*that is, such ratio does not depend on the belief on any other state.*

Axiom 6 means that the *relative change* in disagreement when beliefs change on states $i$ and $i+1$ versus when they change on states $j$ and $j+1$ locally depends only on the beliefs on those states. We interpret this axiom as limiting the complementarity among states. A stronger version of Axiom 6 is given by the following property:

**Property 2** (Separability). *For all $n$, and $\forall\, 1 \leq j \leq n-1$:*

- *if for some $\bar{p}, \bar{q}$:*

$$D((p_1, \ldots, p_j, \bar{p}_{j+1}, \ldots, \bar{p}_n), (q_1, \ldots, q_j, \bar{q}_{j+1}, \ldots, \bar{q}_n))$$
$$\geq D((p'_1, \ldots, p'_j, \bar{p}_{j+1}, \ldots, \bar{p}_n), (q'_1, \ldots, q'_j, \bar{q}_{j+1}, \ldots, \bar{q}_n)),$$

- *then:*

$$D((p_1, \ldots, p_j, \tilde{p}_{j+1}, \ldots, \tilde{p}_n), (q_1, \ldots, q_j, \tilde{q}_{j+1}, \ldots, \tilde{q}_n))$$
$$\geq D((p'_1, \ldots, p'_j, \tilde{p}_{j+1}, \ldots, \tilde{p}_n), (q'_1, \ldots, q'_j, \tilde{q}_{j+1}, \ldots, \tilde{q}_n)),$$

  *for all $\tilde{p}, \tilde{q}$.*

Separability implies Axiom 6. As a matter of fact, if for some $p, q \in \Delta_n$, we consider the function $h$ implicitly defined by:

$$D_n(p, q) = D_n(p, (\ldots, q_i + \epsilon, q_i - \epsilon, \ldots, q_j + h(\epsilon), q_{j+1} - h(\epsilon))),$$

we have that Property 2 implies that $h$ does not depend on states other than $i, i+1, j, j+1$.

On the other hand, Axiom 6 only implies that $-\frac{\partial_j D(p,q)}{\partial_i D(p,q)} = h'(0)$ depends only on beliefs on $i, i+1, j, j+1$. In this sense, Axiom 6 can be interpreted as the local version of Property 2. As it follows from our representation theorem (Theorem 1), all the disagreement measures satisfy Property 2, so we could have assumed such property instead of Axiom 6 without affecting our main result.

## 1.4 Applications

In this section we analyze two applications of our measures of disagreement. Whenever referring to disagreement measures or functions, we mean functions satisfying Axioms 1–6.

The first application shows how disagreement among Bayesian agents changes when they observe a piece of public information. We first show that there exists, generically, a signal realization that increases disagreement. On the other hand, we show in Theorem 2 that averaging on signal realizations, a perfectly informed agent expects to disagree less with any other agent after observing public information (for any disagreement measure). Finally, we compare the expected disagreement between any two agents, finding that it is decreasing in the informativeness of the experiment if and only if the disagreement measure satisfies a continuity axiom, Axiom 7.

In the second application, we consider a model of trade of contingent assets. We show that if agents have Constant Relative Risk Aversion (CRRA) utility functions (with coefficient $1/z_j$) then: i) the equilibrium price of the contingent assets $(\Pi_i)_i$ is the belief that minimizes a weighted sum of disagreement, as measured by

$D^{z_j+0.5}(\cdot,\cdot)$; ii) the volume of trade expected by an agent with beliefs $p^j$ is proportional to $D^{z_j+0.5}(p^j,\Pi)+D^{0.5-z_j}(\Pi,p^j)$; iii) the expected gains from trade are an increasing function of $D^{z_j-0.5}(\Pi,p^j)$. We then show that similar results hold for any utility function, if disagreement is small.

## Disagreement and Public Information

Rational agents update their beliefs upon observing signals that are correlated with the state of the world. Formally, let $\pi=(S,f(s|\theta))$ be an experiment given by a set of signals $S$ (finite, for simplicity) and a family of conditional distributions $f(\cdot|\theta)\in\Delta(S)$. For all $\theta\in\Theta$, $f(s|\theta)$ represents the probability of observing signal $s$ when the state of the world is $\theta$. We assume that agents update their beliefs using Bayes rule, therefore, indicating by $p(s)=(p_1(s),\dots,p_n(s))$ the posterior beliefs, we have that:

$$p(\theta_j|s)=p_j(s)=\frac{f(s|\theta_j)p_j}{\sum_i f(s|\theta_i)p_i},$$

where the denominator represents the *ex-ante* probability of observing signal $s$, which we also denote by: $\mathbb{P}_p(s):=\sum_i f(s|\theta_i)p_i$.

The first proposition we state shows that in general there exist signal realizations that increase or decrease disagreement:[14]

**Proposition 7.** *For any $p,q\in\Delta_n^\circ$ with $p\neq q$ and for all $D$ measures of disagreement there*

[14] This result holds for a large class of distances on the space of beliefs, not only of our measures of disagreement. On the other hand, the positive result of this subsection (in particular, Theorem 2 and Theorem 3) do not generically hold for other metrics on the space of beliefs (norms, Euclidean distance, etc.).

*exists an experiment* $\pi = (S, f(s|\theta))$ *and a signal* $s' \in S$ *such that:*

$$D(p(s'), q(s')) < D(p, q).$$

*For any* $D$ *measure of disagreement, there exist* $p, q \in \Delta_n$*, an experiment* $\pi = (S, f(s|\theta))$*, and a signal* $s \in S$ *such that:*

$$D(p(s), q(s)) > D(p, q).$$

This proposition involves single signal realizations, and does not take into account the fact that some signals are ex-ante more likely than others. In the rest of this subsection, we will analyze expected disagreement, that is: we will weight the disagreement in the posteriors by the ex-ante probability of the signal. Such ex-ante probability is defined only in terms of a certain belief on the state of the world, and we will take this belief to be $p$, the first of the two agents whose disagreement we are analyzing. Formally, for any experiment $\pi$, expected disagreement is defined as:

$$\mathbb{E}_p^\pi(D(p(s), q(s))) := \sum_s \mathbb{P}_p(s) D(p(s), q(s)).$$

Instead of comparing the expected posterior disagreement $\mathbb{E}_p^\pi(D(p(s), q(s)))$ with the disagreement in the priors $D(p, q)$, we will compare the expected posterior disagreement after observing two experiments ranked by Blackwell's notion of sufficiency. We will show under which conditions a more informative experiment induces lower expected disagreement, i.e. what assumptions imply that:

$$\mathbb{E}_p^{\pi}[D(p(s), q(s))] \leq \mathbb{E}_p^{\tilde{\pi}}[D(p(\tilde{s}), q(\tilde{s}))],$$

for all pairs of experiments such that $\pi$ is sufficient for $\tilde{\pi}$. Statistical sufficiency is defined as follows:

**Definition 2** (Sufficiency). Let $\Theta$ be a state space, and $\pi = (S, f(s|\theta))$ and $\tilde{\pi} = (\tilde{S}, g(\tilde{s}|\theta))$ be two experiments. We say that $\pi$ is sufficient for $\tilde{\pi}$ and write $\tilde{\pi} \preceq \pi$ if:

$$g(\tilde{s}|\theta) = \sum_s \lambda_{s,\tilde{s}} f(s|\theta),$$

for some set of positive $(\lambda_{s,\tilde{s}})_{s,\tilde{s}}$ such that $\sum_{\tilde{s}} \lambda_{s,\tilde{s}} = 1$.

In words, an experiment $\pi$ is sufficient for $\tilde{\pi}$ if $\tilde{\pi}$ can be obtained by *garbling* the experiment $\pi$. For example, the letter grade of an exam is a garbling of the percentage grade, because the former can be obtained by adding noise to the latter.

In the next theorem, we will consider $p$ to be the belief of a *perfectly informed agent*, an agent with degenerate belief on a state of the world $\theta$, that we interpret as the *true state of the world*. The following theorem states that more information will make $q$ disagree less with the correct state of the world. This result holds for all the measures of disagreement $D$.

**Theorem 2.** *Let $p$ be a degenerate distribution on the true state of the world, and let $q \in \Delta_n$.*

*Then for all measures of disagreement, and for all $\tilde{\pi} \preceq \pi$ we have that:*

$$\mathbb{E}_p^{\pi}[D(p(s), q(s))] \leq \mathbb{E}_p^{\tilde{\pi}}[D(p(\tilde{s}), q(\tilde{s}))].$$

This result says that, conditional on the true state of the world (i.e. belief $p$), more information will get any agent closer to the true state of the world, when we measure the distance via *any* of our disagreement measures. Notice how this result is *cardinal*, so Axiom 5, our only cardinal axiom, plays a key role in the proof of Theorem 2 (and similarly for the other results of this subsection). Francetich and Kreps (2014) analyze a property analogous to the result of Theorem 2, namely, they ask whether "Bayesian inference can lead us astray". They find a similar negative result, i.e. on average Bayesian inference lead us closer to the truth. Differently from us, though, they do not consider different measures of disagreement, and they do not compare different experiments.

If instead we consider any two agents with beliefs $p, q \in \Delta_n$, it is in general not true that more information decreases disagreement *for any measure of disagreement*, i.e. it will not be the case that:

$$\mathbb{E}_p^{\pi}[D(p(s), q(s))] \leq \mathbb{E}_p^{\tilde{\pi}}[D(p(\tilde{s}), q(\tilde{s}))], \tag{1.8}$$

for all $D$. A necessary and sufficient condition for this to happen is given by the following axiom:[15]

**Axiom 7** (Absolute Continuity). *We say that $D$ is absolutely continuous in the second vari-*

[15]Notice that this axiom is asymmetric in $p, q$ (unlike all the previous ones) and this is related to the fact that equation (1.8) is also asymmetric in $p$ and $q$.

*able if*

$$D(p,q) = +\infty \Leftrightarrow Supp(p) \setminus Supp(q) \neq \emptyset.$$

Let us illustrate with an example why Axiom 7 is necessary for information to decrease disagreement (for all $p, q$).

**Example 1**. *Axiom 7 can be violated in two ways: either because $D(p,q) = \infty$ and $Supp(p) \subseteq Supp(q)$; or because $D(p,q) < +\infty$ and $Supp(p) \setminus Supp(q) \neq \emptyset$. Let us analyze the latter case in detail, the reason why also the former case is necessary is analogous. Let $p, q \in \Delta_2$ with $p = (p_1, 1-p_1)$ , $q = (0,1)$, and $D(p,q) < \infty$. Consider a sequence of beliefs $q^n = \left(\frac{1}{n}, 1-\frac{1}{n}\right)$ and notice that, for large $n$, $D(p,q^n) < D(p,q) < \infty$, because $q^n \in [q,p]$. Therefore $D(p,q^n)$ is finite and bounded uniformly in $n$. We can construct an experiment with signal $s_0$ such that $p(s_0)$ is arbitrary close to $(1,0)$ and such that $q^n(s_0) \approx (0,1)$, for $n$ big enough. In other words, for all $\epsilon$, we can pick an experiment and a prior $q^n$ such that:*

$$D(p(s_0), q^n(s_0)) > D((1-\epsilon,\epsilon),(\epsilon,1-\epsilon)).^{16}$$

*It is easy to see that for all our measures of disagreement $D((1-\epsilon,\epsilon),(\epsilon,1-\epsilon)) \to \infty$, as $\epsilon \to 0$, and therefore we get that in this limit $\mathbb{E}_p^\pi(D(p(s),q^n(s))) \to +\infty$ while $D(p,q^n)$ is bounded, and hence:*

$$D(p,q^n) < \mathbb{E}_p^\pi(D(p(s),q^n(s))).$$

[16] Choosing the right $n$ and the right experiment precision requires some fine tuning, for all fixed $\epsilon$. We leave these details to the proof of Theorem 3, in Appendix A.3.

The problem with disagreement functions that do not satisfy Axiom 7 is that, whenever $Supp(p) \setminus Supp(q) \neq \emptyset$, the disagreement between $p$ and $q$ can be made arbitrarily large by using the fact that $q(\theta|s) = 0$ for all $\theta \in Supp(p) \setminus Supp(q)$. Because of Bayesian updating, an agent who assigns zero probability to a state $\theta$ will assign zero probability after any signal. Therefore, unless $D(p, q) = \infty$, it is possible to find two priors and and an experiment such that $D(p, q) < \mathbb{E}_p^\pi[D(p(s), q(s))]$.

The next theorem, Theorem 3, shows that Axiom 7 is necessary and sufficient for more information to decrease expected disagreement. The proof uses the following characterization of the measures of disagreement that satisfy Axiom 7:

**Lemma 6.** *$D$ satisfies Axioms 1–7 if and only if $D(p, q) = aD^z(p, q)$ for some $z \geq 0.5$, and $a > 0$.*

**Theorem 3.** *Let $D$ be a measure of disagreement. Then the following statements are equivalent:*

1. *$D$ satisfies Axiom 7;*
2. *$D(p, q) = aD^z(p, q)$ for some $z \geq 0.5$ (and $a > 0$);*
3. *for all experiments $\pi \succeq \tilde{\pi}$ and priors $p, q \in \Delta_n$:*

$$\mathbb{E}_p^\pi[D(p(s), q(s))] \leq \mathbb{E}_p^{\tilde{\pi}}[D(p(\tilde{s}), q(\tilde{s}))].$$

## Disagreement and Trade of contingent Assets

Consider an economy with incomplete information on a state space $\Theta$ (with $|\Theta| = n < +\infty$). Let $J$ be a finite set of agents,[17] and with an abuse of notation suppose that $J \in \mathbb{N}$ denotes also the number of agents. Let there be one commodity $x \in \mathbb{R}^+$. Agent $j \in J$ has beliefs $p^j \in \Delta^\circ(\Theta)$ and von Neumann-Morgerstern (vNM) utility function $u_j : \mathbb{R}^+ \to \mathbb{R}$. For simplicity, we assume that for all $j$, $u_j$ is strictly increasing, strictly concave and satisfies the Inada conditions $\lim_{x\to 0} u'(x) = \infty$, $\lim_{x\to\infty} u'(x) = 0$. These assumptions will imply that the equilibrium of our model exists, is unique, and is pinned down by the first order conditions.[18] The ex-ante utility of agent $j$ is given by:

$$U_j(\mathbf{x}) := \mathbb{E}_{p^j}(u_j) := \sum_i p_i^j u_j(x_i^j),$$

where $x_i^j$ is the amount of commodity consumed in state $i$, and we denote by $\mathbf{x} = (x_1^j, \ldots, x_n^j)$ the vector of contingent commodity to be consumed in states $\theta_1, \ldots, \theta_n$.

We assume that there is a market for Arrow Debreu (AD) securities for states $i \in I \subseteq \Theta$. I.e. there is a set of risky assets that pay $1$ unit of the commodity $x$ in state $i$ and $0$ in all the other states (for all $i \in I$). Let $I \geq 2$ be also the cardinality of the set $I$. The case of $I = \Theta$ corresponds to a complete market in which agents can insure themselves against any contingency. The prices of the Arrow Debreu securities will be denoted by $(\Pi_i)_{i\in I}$, and we normalize this vector to have $\sum_{i\in I} \Pi_i = 1$, as we will later interpret such price

[17]Studying a model with countably many agents, or a continuum of them, would yield analogous results.

[18]In particular, notice that strict concavity implies that for all set of beliefs $(p^j)_j$ with $p^j \in \Delta^\circ(\Theta)$ the equilibrium amount of trade will be finite. This makes sure that infinite bets such as those described in Eliaz and Spiegler (2007) are never an equilibrium.

vector as the "market belief".

We assume that each agent is endowed with a unit of commodity in each state, and can trade it for units of commodity in other states, at prices $\Pi_i$. Equal endowments across agents and states implies that this is a purely speculative economy: agents trade if and only if they disagree with the market. Formally, an equilibrium of this economy is defined as follows:

**Definition 3** (Equilibrium). An equilibrium is a price vector $(\Pi_i)_{i\in I}$ and a set of allocations $(x_i^j)_{i\in I}^{j\in J}$ such that:

- for all $j \in J$ the vector $(x_i^j)_{i\in I}$ solves:

$$\max_{\mathbf{x}\in\mathbb{R}^n} \quad \sum_i p_i^j u_j(x_i^j) \qquad \text{s.t.} \sum_{i\in I} \Pi_i x_i \leq 1, \tag{1.9}$$

- market clears, i.e. for all states $i$:

$$\sum_{j\in J} x_i^j = J.$$

As noted by Sebenius and Geanakoplos (1983), if two agents with common prior and asymmetric information agree to trade, that information should induce both parties to update their beliefs on the state of the world. In particular, if agents have common knowledge of the information partitions (i.e. they agree to disagree, see Aumann (1976)) then discussion between the agents will reveal enough information to make the trade unappealing. In our model, we abstract from these considerations, as we do not model the source of heterogeneity in beliefs.

The quantities we will analyze are the expected volume of trade and gains from trade. The expected volume of trade of agent $j$ is given by:

$$V_j(\mathbf{x}) := \sum_i p_i^j (x_i^j - 1),$$

i.e. the net amount of commodity that agent $j$ expects to receive after uncertainty is resolved.[19] The gains from trade are defined as:

$$G_j(\mathbf{x}) := U_j(\mathbf{x}) - U_j(\mathbf{1}),$$

i.e. the difference of expected utility between the equilibrium of the model and the no-trade outcome, $\mathbf{1} = (1, \ldots, 1)$.

It is plain to see (from the first order conditions) that amount of asset $x_i^j$ traded in state $i$ is and increasing function of the likelihood ratio $p_i^j/\Pi_i$. Therefore, in order to measure the equilibrium vector $\mathbf{x}$ we expect to find a function of the vectors of likelihood ratios $(p_1^j/\Pi_i, \ldots, p_n^j/\Pi_n)$. This implies that norms, or metrics that are not based on likelihood ratios, are not a good proxy for the amount of trade. In other words, it is generically not the case that $G_j(\mathbf{x})$ is increasing in $\|p - \Pi\|$. The next subsection shows that, on the other hand, $G_j(\mathbf{x})$ is increasing in an appropriate function of disagreement $D(p, \Pi)$ whenever an agent has constant relative risk aversion (CRRA).

[19]For each state $i$, agent $j$ is endowed with 1 unit of commodity in state $i$ and in equilibrium she will consume $x_i^j$, therefore the net amount of trade in state $i$ is $x_i^j - 1$.

**CRRA utility functions**

Suppose that the utilities $(u_j)_{j\in J}$ exhibit Constant Relative Risk Aversion, i.e. utility functions parametrized by:

$$u_j(x) = \frac{x^{1-\frac{1}{z_j}}}{1-\frac{1}{z_j}} \qquad z_j > 0, \quad x > 0.$$

The risk tolerance of an agent with utility $u_j$ is $-\frac{u_j'(x)}{u_j''(x)} = z_j x$, and empirical estimation of CRRA utility functions typically yield $z_j \in (0,1)$.[20] For this reason, the next theorem will impose $z_j \in (0,1)$ and we defer to Remark 2 a discussion of the differences with the case $z_j > 1$.

**Theorem 4.** *Let $p^1, \ldots, p^J \in \Delta_n^\circ$, and let $(u_j)_j$ be CRRA utility functions with parameters $z_j \in (0,1)$. The equilibrium of the economy exists and is unique, and it can be characterized as follows:*

- $(\Pi_i)_i$ *is the unique solution to the problem:*

$$\min_{q\in\Delta(I)} \sum_j \frac{1}{1-z_j} D^{z_j+0.5}(q, p^j(\cdot|I)), \tag{1.10}$$

  *i.e. the prices of the Arrow Debreu securities are the beliefs that minimize weighted disagreement with the agents;*

[20]See Table 1 in Neilson and Winter (2002). The values they find are positive and smaller than 1, with the exception of Hansen and Singleton (1983) who find $-\frac{xu''(x)}{u'(x)} = \frac{1}{z_j} \in [0.07, 0.62]$)).

- *the expected volume of trade is given by:*

$$V_j((x_i^j)_i) = \exp(D^{z_j+0.5}(p^j(\cdot|I), \Pi) + D^{-z_j+0.5}(\Pi, p^j(\cdot|I))) - 1 \quad \forall j \in J; \tag{1.11}$$

- *the gains from trade are given by:*

$$G_j((x_i^j)_i) = \frac{1}{\frac{1}{z_j} - 1}\left(1 - \exp\left(-\frac{D^{z_j-0.5}(\Pi, p^j(\cdot|I))}{z_j}\right)\right), \quad \forall j \in J. \tag{1.12}$$

The first point of the theorem says that the equilibrium price of the assets can be interpreted as the belief that minimizes the sum of disagreement among agents. Therefore, the market aggregates beliefs to a price that minimizes total disagreement. Secondly, the volume of trade and the gains from trade are both increasing in the disagreement between one agents' belief and the market belief.

*Remark* 2. We stated the theorem for $z_j \in (0, 1)$ as this corresponds to the empirically relevant case. When $z_j > 1$ the formula for the gains from trade becomes:

$$G_j((x_i^j)_i) = \frac{1}{\frac{1}{z_j} - 1}\left(1 - \exp\left(\frac{D^{z_j-0.5}(\Pi, p^j(\cdot|I))}{z_j}\right)\right),$$

and therefore gains from trade are still increasing in disagreement $D^{z_j-0.5}$. On the other hand, the same does not hold for the volume of trade, as the equilibrium formula becomes:

$$V_j((x_i^j)_i) = \exp(D^{z_j+0.5}(p^j(\cdot|I), \Pi) - D^{-z_j+0.5}(\Pi, p^j(\cdot|I))) - 1.$$

This is due to the fact that if an agents' risk tolerance increases fast with wealth (i.e. $z_j \gg$

0) the shadow price of her income become increasing in disagreement with market belief. Therefore the equilibrium effect on gains from trade is not unambiguously increasing in disagreement.

### Trade for Moderate Disagreement

We conclude this section by extending the results of Theorem 4 to agents with *generic* utility functions, under the additional assumption that agents' disagreement is small. Small disagreement, in our setting, means that the disagreement between agents is small compared to the rate at which their relative risk aversion changes, and then it can be approximated with a constant. Formally, we will model small disagreement as the limit model as agents' beliefs all converge to the same belief $p^*$.

This case is particularly interesting for two reasons: firstly, one rarely observes very large differences in beliefs in financial markets (with the exception of economic crisis or periods of political turmoil); secondly, most agents invest a small portion of their wealth so even when agents' preferences are not CRRA, approximating them locally with CRRA utility function provides a useful benchmark.

We formalize the idea of "moderate disagreement" by taking a limit of beliefs. Let the belief of agent $j$ depend on an index $m \in \mathbb{N}$, let us denote them by $p^{j,(m)}$.

**Definition 4** (Merging Beliefs). We say that beliefs are merging if for some norm $\|\cdot\|$ on $\Delta_n$:

$$\lim_{m \to +\infty} \left( \sup_{\substack{j_1, j_2 \in J \\ m_1, m_2 > m}} \|p^{j_1,(m_1)} - p^{j_2,(m_2)}\| \right) = 0. \tag{1.13}$$

In words, beliefs are merging if for all $\epsilon$, we can find $\bar{m}$ such that all beliefs $p^{j,(m)}$

are in a ball of radius $\epsilon$ for all $j \in J$ and $m > \bar{m}$. Notice furthermore that Definition 4 implies that if beliefs are merging, then for all $j$, $(p^{j,(m)})_m$ has a limit in $\Delta_n$ as $m \to +\infty$, and such limit is the same for different agents $j$.[21] Since in general such limit could not belong to the interior of $\Delta_n$ we will assume that agents' beliefs are bounded away from the boundary of $\Delta_n$, to avoid complications related to zero probability states.

**Definition 5** (Uniformly Mixed Beliefs). We say that the sequence of family of beliefs $(p^{j,(m)})_{m\in\mathbb{N}}^{j\in J}$ is uniformly mixed if there exists $\varepsilon > 0$ such that $p_i^{j,(m)} > \varepsilon > 0$ for all $i = 1, \ldots, n$, $j \in J$, $m \in \mathbb{N}$.

In order to highlight the dependence of trade on beliefs, we will fix the preferences of the agents (i.e. they will not depend on $m$). For each $j$, let $u_j$ be any smooth strictly increasing and concave utility function defined in a compact neighborhood of $1$. As agents' beliefs merge, the allocation $x_i^j$ converges to $1$, the no trade outcome. Therefore, in the next theorem we approximate the utility functions $u_j$ with the best CRRA approximation at $1$. Without loss of generality we assume that $u_j'(1) = 1$ for all $j$, and then defining $z_j = -\frac{1}{u''(1)}$ we have that the function:

$$\tilde{u}_j(x) := \frac{x^{1-\frac{1}{z_j}}}{1-\frac{1}{z_j}}$$

is the only CRRA function such that $u'(1) = \tilde{u}_j'(1)$ and $u''(1) = \tilde{u}_j''(1)$. Theorem 5 implies that as agents' beliefs merge, the equilibrium of the model is asymptotic to the solution to the CRRA approximation. To simplify the notation, we assume that $I = \Theta$, i.e. the

[21] This follows directly from the compactness of $\Delta_n$.

market for contingent securities is complete. It is easy to see that all the results extend to the case of $I \subset \Theta$.

**Theorem 5.** *Let $(u_j)_{j\in J}$ be a family of strictly concave, twice continuously differentiable utility functions defined in a compact neighborhood of $1$. Suppose (without loss of generality) that $u_j'(1) = 1$, and assume that $-1/u_j''(1) =: z_j \in (0,1)$. Let $(p^{j,(m)})_{m\in\mathbb{N}}^{j\in J}$ be any family of merging and uniformly mixed beliefs.*

*For all $m$ the equilibrium exists and is unique, denote it by $((x_i^{j,(m)})_{i=1,\dots,n}^{j\in J}, (\Pi_i^m)_i)$. For all $m$, let $\tilde{\Pi}^m$ be the solution of the problem $\min_{q\in\Delta_n^\circ} \sum_j \frac{1}{1-z_j} D^{z_j+0.5}(q, p^{j,(m)})$. We have that:*

- *for all $j$, $\lim_{m\to\infty} \frac{D^{z_j+0.5}(p^{j,(m)},\Pi^m)}{D^{z_j+0.5}(p^{j,(m)},\tilde{\Pi}^m)} = 1$, so the disagreement between any belief $p^j$ and the approximate equilibrium $\tilde{\Pi}^m$ is asymptotically equivalent to the disagreement with the market belief $\Pi^m$.*
- *The volume of trades is asymptotic to the volume of trades in the economy with CRRA utility functions:*

$$\lim_{m\to\infty} \frac{\sum_i p_i^{j,(m)}(x_i^{j,(m)} - 1)}{D^{z_j+0.5}(p^{j,(m)}, \tilde{\Pi}^m) + D^{-z_j+0.5}(\tilde{\Pi}^m, p^{j,(m)})} = 1.$$

- *The gains from trade are asymptotic to the volume of trades in the economy with CRRA utility functions:*

$$\lim_{m\to\infty} \frac{\sum_i p_i^{j,(m)} u_i(x_i^{j,(m)})}{\frac{1}{1-z_j} D^{z_j-0.5}(\tilde{\Pi}^m, p^{j,(m)})} = 1.$$

This theorem shows that whenever agents disagree moderately, their disagreement is a sufficient statistic for the volume and gains from trade, regardless of their utility functions. The parameter of the disagreement measures captures the local coefficient of relative risk aversion of the agents, around the no trade outcome.

Chapter 2

# *A belief-independent cost of information*[22]

## 2.1 Introduction

In order to model agents' attitude toward information, it is necessary to specify both the value of information, and its cost. Economists have studied the value of information in many instances,[23] whereas the cost of information has received considerably less attention. The prevailing method to measure the cost of information has been pioneered by Sims (2003), who employed the expected change in entropy[24] between prior and posteriors as a measure for the cost of information. In this chapter, we argue that this cost function has some problematic properties and we propose an alternative function, which we characterize axiomatically.

The first reason why the expected change in entropy is a problematic measure of the cost of an experiment[25] is that this cost function depends on the experimenter's belief. In particular, the more the experimenter is sure of a state of the world (i.e. the lower the

[22]I am grateful to Yeon-Koo Che, Alexander Frankel, Emir Kamenica, Navin Kartik, Pietro Ortoleva, Daniel Rappopport, Rajiv Sethi, Paolo Siconolfi, and Michael Woodford for the useful discussions.

[23]See for example Blackwell (1953), Persico (2000), Ganuza and Penalva (2010), and Athey and Levin (2017).

[24]Sims' cost of information is defined below in Corollary 4, with $H$ being entropy as defined thereafter. For a thorough analysis of entropy-related measures see Cover and Thomas (2006).

[25]We denote by "experiment", or "information structure", any observable signal that is correlated with the state of the world. A precise mathematical definition is postponed to Section 2.2.

entropy of the prior), the less costly the experiment. Let us illustrate with an example why this is an undesirable property, for a function measuring the cost of an experiment. The reader of this chapter must hold a prior belief on the quality of this chapter. Reading the chapter, she is incurring in some cost of acquiring information about the chapter's quality, the cost being the time spent on these paragraphs, the time spent learning English, etc. How costly the "experiment" of reading the chapter is depends on how focused the reader is, whether she reads the details of the proofs, etc. In other words, the cost is clearly related to the *precision* of the experiment. Whether the reader believes this chapter to be good or bad should not affect the the cost of reading it. The change of entropy from prior to posterior (i.e. the relative entropy between them) is *correlated* with the experiment's precision and cost, but it is a spurious measure of them.

The second reason why we argue that the change of entropy is not a good measure of the cost of an information structure is that the cost of running the same experiment twice is not equal to twice the cost of the experiment. More generally, the cost of running two independent experiments has no general relation with the sum of the costs of each experiment.

There are other properties of the expected change in entropy that instead are desirable for a cost measure. For example, according to such measure, observing the garbling of an experiment is less costly than observing the experiment itself. Another useful property of the change of entropy is that it is additively separable in the signal realizations of the experiments. In this chapter, the alternative cost function we characterize satisfies these properties, while simultaneously not satisfying the two problematic ones mentioned above.

The primitive of our analysis is an experiment, that is a set of signals that an agent can observe, and the likelihood of each signal, given the state of the world. Since the experiment, as we define it, is unrelated to the agents' beliefs on the state of the world, our cost function is prior independent, and this solves the first problem of the relative change in entropy. Our cost function will only depend on the *precision* of the experiment, as defined by our postulates.

In our Postulate 7, we will impose that that the cost of observing two independent experiments equals the sum of the cost of observing each of them, therefore removing the second problematic property mentioned above. Two other postulates (Postulate 3 and 4) imply that the garbling of an experiment is less costly than the experiment (see Proposition 8). Postulates 5 and 6 imply that our cost function is additively separable in the signal realization, as is the expected change in entropy.

This chapter is related to Chapter 1 where a different question motivates a set of axioms that describe some properties of a disagreement measure. Some of those axioms are introduced here as they can be used to characterize the precision of an experiment. The two chapters differ in that the primitive of Chapter 1 is given by the beliefs of two agents; whereas here we consider as primitive an experiment on the state of nature.

## Related Literature

The importance of correctly measuring the cost of information has been highlighted by many authors, in particular Woodford (2012) documents how a large body of experimental evidence is inconsistent with the measure proposed by Sims (2003). More than that, Woodford (2012) also advocates for a cost function that is independent of the belief of the agent, like the one we propose in this chapter. Hébert and Woodford (2016) define a class of prior-independent information cost that generalize relative entropy. Importantly, as we stress in the discussion following the main body of the paper, one of the conditions imposed in Hébert and Woodford (2016) is not satisfied by our cost function (continuity at the non-informative experiment, see Proposition 10).

On the other hand, our cost function satisfies the conditions imposed by Caplin and Dean (2015) and Denti et al. (2016) . Caplin and Dean (2015) show that a convexity property (strengthening Postulate 3), monotonicity in Blackwell informativeness (Proposition 8) and a normalization are necessary and sufficient conditions for rationalizing a data set. Denti et al. (2016) define a cost function satisfying those conditions as *canonical*, and they show that in a model of rationally inattentive preferences it is possible to identify a canonical cost function. Since our cost function is canonical, it is possible to test empirically the predictions of our cost function against different functions.

Many papers describe experiments designed to empirically measure the cost of acquiring information in different contexts, e.g. Gabaix et al. (2006), Caplin et al. (2011), Dewan and Neligh (2017). While our chapter does not provide testable implications, in our future research we would like to study if our cost of information allows to better rationalize the experimental evidence.

## 2.2 Model and Postulates

Let $\Theta = \{\theta_1, \dots, \theta_n\}$ be a finite set of unknown states of the world. We consider *information structures*, or *experiments*, $\pi = (S, (f(s|\theta))_{s \in S, \theta \in \Theta})$ defined by a set of signal realizations $S$ and a family of conditional probability distributions $f(\cdot|\theta) \in \Delta(S)$,[26] which represent the likelihood of observing a certain signal when $\theta$ is the true state of the world.

We say that a signal $s$ is possible if for some $\theta$ we have that $f(s|\theta) > 0$. For simplicity, we assume that only finitely many signals are possible, i.e. that the support of $f(\cdot|\theta)$:

$$S(f(\cdot|\theta)) := \{s \in S \mid f(s|\theta) > 0\},$$

is finite, $|S(f(\cdot|\theta))| < \infty$ for all $\theta$. To make notation lighter, we also assume that $S$ is countably infinite.[27] Formally, the domain of our cost function is the following space of experiments:

**Definition 6** (Space of experiments). For any $\Theta$ finite, let $\mathscr{E}_{\Theta,S}$ be the set of experiments $\pi = (S, (f(s|\theta))_{s \in S, \theta \in \Theta})$ with $S$ countably infinite, and such that $|S(f(\cdot|\theta))|$, the support of $f(\cdot|\theta)$, is finite for all $\theta$.

An experiment $\pi \in \mathscr{E}_{\Theta,S}$ can be represented as a matrix with $|\Theta|$ columns, and in-

---

[26] For any set $A$, we denote $\Delta(A)$ to be the set of probability distributions on $A$. Since we will be dealing with countable sets, we do not need to specify a $\sigma$-algebra (it is always possible to take the partition set to be such $\sigma$-algebra).

[27] I.e. we embed all the *finite* possible signal realization into a countably infinite set, which therefore contains infinitely many impossible signals. This simplifies the notation, because it allows not to bother defining how many different signal realizations an experiment can produce.

finitely many rows:

$$\pi = \begin{pmatrix} f(s_1|\theta_1) & \dots & f(s_1|\theta_n) \\ f(s_2|\theta_1) & \dots & f(s_2|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix}. \tag{2.1}$$

The condition that $f(\cdot|\theta)$ be a probability distributions imply that $f(s|\theta) \geq 0$ for all $s, \theta$, and that $\sum_s f(s|\theta) = 1$, for all $\theta \in \Theta$. In terms of the matrix notation, $\pi$ is positive and column stochastic. For all $\pi \in \mathscr{E}_{\Theta,S}$, furthermore, only finitely many rows are non-null, as only finitely many signals are possible.

Among the set of experiments, it is useful to single out some of them, as we will often refer to them in the rest of the chapter.

**Definition 7.** 1. We say that an experiment $\pi$ is non-informative if all signals are equally likely under different states of the world. I.e. if for all $s \in S$:

$$f(s|\theta) = f(s|\theta'), \quad \forall \theta, \theta' \in \Theta.$$

2. We say that an experiment does not preclude any state if there does not exist any *possible* signal $s$ such that $f(s|\theta) = 0$ for some $\theta$. Equivalently an experiment does not preclude any state if:

$$S(f(\cdot|\theta)) = S(f(\cdot|\theta')) \quad \forall \theta, \theta' \in \Theta,$$

i.e. the set of possible signals does not depend on the state.

3. We say that an experiment is fully informative if for all $\theta$ there exists a unique signal $s_\theta \in S$ such that $f(s_\theta|\theta) > 0$.

It is useful to comment on how an experiment affects a Bayesian agent's beliefs. Let $\mu \in \Delta(\Theta)$ be the agent's prior on the state of the world, and assume that $\mu(\theta) > 0$ for all $\theta \in \Theta$, i.e. without loss of generality we assume that each state is deemed possible. Only signals such that $f(s|\theta) > 0$ for some $\theta$ are possible, and for such signals we can define the posterior belief:

$$\mu(\theta|s) = \frac{\mu(\theta) f(s|\theta)}{\sum_{\theta'} \mu(\theta') f(s|\theta')}.$$

A non informative experiment can then be defined as one for which $\mu(s) = \mu$ for all $s$. An experiment does not preclude any state if $\mu(\theta) > 0$ implies that $\mu(\theta|s) > 0$ for all possible $s$. Finally, a fully informative experiment can be defined as one for which $\mu(\theta|s)$ is a degenerate belief[28] for all possible signal realizations.

The goal of this chapter is characterizing, for all state space $\Theta$ and signal space $S$, the functions

$$c_{\Theta,S} : \mathscr{E}_{\Theta,S} \to [0, +\infty]$$

that satisfy the postulates of the following subsection. We assume $c_{\Theta,S}$ to be three time differentiable.[29]

[28]A degenerate belief is a belief concentrated on one state of the world, $\mu = (1, 0, \ldots, 0)$; $(0, 1, 0, \ldots, 0)$; etc.

[29]More precisely, we assume that for any $\pi, \pi' \in \mathscr{E}_{\Theta,S}$ the limit $\lim_{\epsilon \to 0^+} \frac{c_{\Theta,S}((1-\epsilon)\pi + \epsilon\pi') - c_{\Theta,S}(\pi)}{\epsilon}$ exists, and so do the second and third order derivatives. Since $\mathscr{E}_{\Theta,S}$ is convex, $(1-\epsilon)\pi + \epsilon\pi' \in \mathscr{E}_{\Theta,S}$ and therefore the limit is well-defined.

## Postulates

### Zero and Finite Cost

**Postulate 1** (Zero and Finite Cost). $c_{\Theta,S}(\pi) = 0$ *if and only if* $\pi$ *is uninformative. If* $\pi$ *does not preclude any state, then* $c_{\Theta,S}(\pi) < \infty$.

This postulate implies that any informative experiment is (strictly) more costly than the non-informative experiment. Assuming that any experiment $\pi$ that does not preclude any state has finite cost implies that it is possible for the perfectly informative experiment to be *strictly* more costly. Removing this part of Postulate 1 would imply that:

$$c_{\Theta,S}(\pi) = \begin{cases} 0 & \text{if } \pi \text{ is uninformative} \\ +\infty & \text{otherwise} \end{cases}$$

satisfies all our postulates. We want to exclude this cost function, while allowing $c_{\Theta,S}(\pi)$ to be infinite for some experiment.

### Anonymity

In this postulate, we impose that $c_{\Theta,S}$ does not depend on the state space $\Theta$ and on the set of signal $S$, but only on how signals and states are correlated (i.e. on $f(s|\theta)$). In our treatment, both the states and the signals are labels that do not carry any intrinsic value, and therefore we assume that $c_{\Theta,S}$ be invariant from relabeling.

To formally assume this property, let us firstly introduce *isomorphic experiments*:

**Definition 8** (Isomorphic Experiments). For any $\Theta, \Theta'$ state spaces and $S, S'$ signal spaces, we say that two experiments $\pi = (S, (f(s|\theta))_{s\in S, \theta\in\Theta})$ and $\pi' = (S', (g(s'|\theta'))_{s'\in S', \theta'\in\Theta'})$ are isomorphic if there exist $\gamma : \Theta \to \Theta'$ and $\lambda : S \to S'$ bijections such that:

$$f(s|\theta) = g(\lambda(s)|\gamma(\theta)), \quad \forall s \in S, \theta \in \Theta.$$

Representing experiments as matrices (as in equation (2.1)) we have that $\pi$ and $\pi'$ are isomorphic if and only if there exists a permutation of the rows and the columns of $\pi$ that yields $\pi'$.

**Postulate 2** (Anonymity). *If* $\pi = (S, (f(s|\theta))_{s\in S, \theta\in\Theta})$ *and* $\pi' = (S', (g(s'|\theta'))_{s'\in S', \theta'\in\Theta'})$ *are isomorphic, then* $c_{\Theta,S}(\pi) = c_{\Theta',S'}(\pi')$.

Thanks to Postulate 2, the dependence of $c_{\Theta,S}$ on $S$ is irrelevant, because any two countably infinite signal sets yield the same cost function. Therefore, we can write $c_\Theta$ without specifying what the set of signals is. On the other hand, notice that we cannot remove the dependence on $\Theta$, because Postulate 2 does not allow to compare cost functions $c_\Theta$ and $c_{\Theta'}$ such that $|\Theta| \neq |\Theta'|$ (i.e. such that there exists no bijection between $\Theta$ and $\Theta'$). Nonetheless, we can simplify the notation by writing $c_n$ to be the cost of information structures $\pi = (S, (f(s|\theta))_{s\in S, \theta\in\Theta})$, where $\Theta$ is *any* set of cardinality $n \in \mathbb{N}$. Similarly, we will denote by $\mathscr{E}_n$ the set of experiments on any state space of cardinality $n$.

**Mixture of experiments**

Observe that $\mathscr{E}_n$ is a convex set, therefore the experiment $\lambda\pi + (1 - \lambda)\pi'$ belongs to $\mathscr{E}_n$, for any $\pi, \pi' \in \mathscr{E}_n$ and any $\lambda \in [0, 1]$. The following postulate relates the cost of such experiment to the costs of $\pi$ and $\pi'$.

**Postulate 3** (Quasi-Convexity). *For any $n$, and for any $\pi, \pi' \in \mathscr{E}_n$:*

$$c_n(\lambda\pi + (1-\lambda)\pi') \leq \max\{c_n(\pi), c_n(\pi')\} \qquad \forall \lambda \in [0,1].$$

In other words, supposing without loss of generality that $\pi$ is more costly than $\pi'$, we have that the cost of $\lambda\pi + (1-\lambda)\pi'$ cannot be strictly larger than $c_n(\pi)$. Mixing a costly experiment with a less costly one cannot increase the cost. Formally, Postulate 3 amounts to assume that $c$ be quasi-convex.

A similar postulate is often assumed in axiomatizations of cost of information functions. For example, Caplin and Dean (2015) consider convex cost functions $c$, which are a subset of the cost functions that satisfy Postulate 3 (because convexity implies quasi-convexity).

**Simple Garbling**

Consider an experiment $\pi = (S, (f(s|\theta))_{s\in S, \theta\in\Theta})$, and let $\mathscr{A} = (A_i)_i^\infty$ be an infinite partition of $S$.[30]. If instead of observing the signal realization $s \in S$, an agent observes the set $A_i \subseteq \mathscr{A}$ that contains $s$, the agent has access to less refined information. Therefore, in the next postulate, we assume that the experiment $\pi' = (\mathscr{A}, f(A|\theta)_{A\in\mathscr{A},\theta\in\Theta})$ is less costly than $\pi$.

A simple example captures this idea. Suppose that $\theta$ represents the quality of a student, and that $\pi$ is the percentage grade in an exam. A simple garbling of $\pi$ would be the

[30]A partition is a set of $A_j \subset S$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $\bigcup_i A_i = S$.

letter grade, because a subset of percentage grades are mapped into the same letter grade. Naturally, measuring the letter grade is less costly than measuring the percentage grade.

To formally present this postulate, consider two infinite set of signals $S$ and $S'$ and a surjective function $\gamma : S \to S'$. For any $\pi = (S, (f(s|\theta))_{s\in S, \theta\in\Theta})$ define the experiment $\pi' := \gamma \circ \pi = (S', g(s'|\theta)_{s'\in S', \theta\in\Theta})$ to be defined by: $g(s'|\theta) = \sum_{s\in\gamma^{-1}(s')} f(s|\theta)$.[31]

To represent $\gamma\circ\pi$ in terms of the matrix representation of $\pi$, we have that the experiment $\gamma\circ\pi$ can be obtained by summing rows of $\pi$. For example, if the simple garbling $\gamma\circ\pi$ *merges* signals $s_1$ and $s_2$ into the same $s'_1$ (and leaves the other signals unvaried) we have that:

$$\pi = \begin{pmatrix} f(s_1|\theta_1) & \dots & f(s_1|\theta_n) \\ f(s_2|\theta_1) & \dots & f(s_2|\theta_n) \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix} \quad \to \quad \gamma\circ\pi \left( = \begin{matrix} f(s_1|\theta_1)+f(s_2|\theta_1) & \dots & f(s_1|\theta_n)+f(s_2|\theta_n) \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{matrix} \right)$$

Observe that if $\pi \in \mathscr{E}_n$, also $\gamma\circ\pi \in \mathscr{E}_n$ for any $\gamma : S \to S'$. We briefly say that $\pi' = \gamma\circ\pi$ is a simple garbling of $\pi$, because the experiment $\pi'$ can be obtained by merging a subset of signals in $S$ into the same $s'$. Postulate 4 says that $\gamma\circ\pi$ cannot be more costly than $\pi$.

**Postulate 4** (Simple Garbling). *For all $\gamma : S \to S'$ we impose tha: $c_n(\gamma\circ\pi) \leq c_n(\pi)$.*

Postulates 1–4 have some important implications. First, we show that if two signals $s_1, s_2 \in S$ have the same likelihood ratios $f(s|\theta)/f(s|\theta')$ for all $\theta, \theta'$, then considering the simple garbling $\pi'$ that merges $s_1$ and $s_2$ into the same event $\{s_1, s_2\}$ does not change the cost of the information structure.

[31] $\gamma^{-1}(s)$ is defined as the pre-image of $s'$, or subset of signals $s \in S$ such that $\gamma(s) = s'$.

**Lemma 7** (Redundant signals). *Suppose that for two signals $s_1, s_2$ we have that:*

$$f(s_1|\theta)f(s_2|\theta') = f(s_1|\theta')f(s_2|\theta)$$

*for all $\theta, \theta'$. Then considering $\gamma : S \to S'$ such that $\gamma(s_1) = \gamma(s_2)$ and $\gamma(s) \neq \gamma(s')$ for any $s \neq s'$ ($s, s' \neq s_1, s_2$) we have that:*

$$c_n(\gamma \circ \pi) = c(\pi).$$

*Proof.* All proofs are in the appendix. $\square$

To rephrase the result of Lemma 8 in terms of the matrix representation of an experiment $\pi$ suppose that one row is proportional to another one:

$$\pi = \begin{pmatrix} f(s_1|\theta_1) & ... & f(s_1|\theta_n) \\ f(s_2|\theta_1) & ... & f(s_2|\theta_n) \\ ... & ... & ... \\ f(s_m|\theta_1) & ... & f(s_m|\theta_n) \\ kf(s_m|\theta_1) & ... & kf(s_m|\theta_n) \\ ... & ... & ... \end{pmatrix}, \quad \exists\, k > 0,$$

then summing the rows we find another experiment:

$$\tilde{\pi} = \begin{pmatrix} f(s_1|\theta_1) & ... & f(s_1|\theta_n) \\ f(s_2|\theta_1) & ... & f(s_2|\theta_n) \\ ... & ... & ... \\ (1+k)f(s_m|\theta_1) & ... & (1+k)f(s_m|\theta_n) \\ ... & ... & ... \end{pmatrix}.$$

Postulates 4 implies that $c_n(\tilde{\pi}) \leq c_n(\pi)$, thanks to Lemma 8 we also know that $c_n(\tilde{\pi}) = c_n(\pi)$. Lemma 8 also implies that adding signals that are not possible (i.e. such that $f(s|\theta) = 0$ for all $\theta$) does not change the cost of the experiment. Therefore, even though our theory formally applies only to experiments with finitely many possible signals embedded in an infinite signal spaces, the same theory applies to experiments with finitely many signals. More importantly, Lemma 7 implies that what determines the cost of a signal realization is the set of likelihood ratios.

Another important consequence of the first four postulates (Postulates 1–4) is that if $\pi$ is sufficient for $\tilde{\pi}$, then it must be more costly, $c(\pi) \geq c(\tilde{\pi})$. This is a very natural property and it is met also by the cost function typically used in the rational inattention literature.

**Definition 9** (Sufficiency). $\pi = (S, (f(s|\theta))_{s \in S, \theta \in \Theta})$ is sufficient for $\pi' = (S', (g(s'|\theta))_{s' \in S', \theta \in \Theta})$ if

$$g(s'|\theta) = \sum_{s} \lambda_{s,s'} f(s|\theta),$$

for some state independent $\lambda_{s,s'} \geq 0$ with $\sum_{s' \in S'} \lambda_{s,s'} = 1$.[32] Equivalently, we say that $\pi'$ is a *garbling* of $\pi$.

In other words, $\pi$ is sufficient for $\pi'$ if the latter can be obtained by adding noise to the first one. Sufficiency can also be characterized in terms of a rational agent's decision problem,[33] in the sense that any rational agent solving any decision problem would achieve a higher expected utility were she to observe the experiment $\pi$ instead of $\pi'$. In this sense, sufficiency is a very strong order on experiments, so it is desirable to have that if $\pi$ is sufficient for $\pi'$ it is also more costly, as the next proposition proves:

**Proposition 8.** *Let $\pi$ be sufficient for $\tilde{\pi}$. If $c_n$ satisfies Postulates 1–4 then $c_n(\pi) \geq c_n(\tilde{\pi})$.*

[32]Observe that the two experiments can have different signal sets (while they have the same state space).

[33]See Blackwell and Girshick (1954).

### Separability on the Signals

Consider an experiment $\pi = (S, (f(s|\theta))_{s\in S,\theta\in\Theta})$ and let $S_1, S_2 \subset S$ be two disjoint and exhaustive subsets of signals, $S_1 \cap S_2 = \emptyset$ and $S_1 \cup S_2 = S$. Let $\pi_1 = (S_1, (f(s|\theta))_{s\in S_1,\theta\in\Theta})$ and $\pi_2 = (S_2, (f(s|\theta))_{s_2\in S,\theta\in\Theta})$ be the "sub-experiments".[34] To indicate this we resort to our matrix notation and write $\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}$.[35]

The following postulate limits the way in which the signals in $\pi_1$ can affect the cost of the signals in $\pi_2$ (and vice versa, because of Postulate 2). In other words, Postulate 5 rules out any complementarity among signals, by imposing that the cost of an information structure is separable.

**Postulate 5** (Separability (Signal)). *If for some* $\pi_1$*:*

$$c_n \begin{pmatrix} \pi \\ \pi_1 \end{pmatrix} \geq c_n \begin{pmatrix} \pi' \\ \pi_1 \end{pmatrix},$$

*then:*

$$c_n \begin{pmatrix} \pi \\ \pi_2 \end{pmatrix} \geq c_n \begin{pmatrix} \pi' \\ \pi_2 \end{pmatrix},$$

*for all* $\pi_2$*.*

---

[34] Notice that in general $\pi_1$ and $\pi_2$ are *not* experiments, because $\sum_{s\in S_1} f(s|\theta) \leq 1$, and for some $\pi$ such inequality might be strict.

[35] Whenever we write an $\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}$ we implicitly assume that $\pi$ is an experiment, without explicitly imposing that:

$$\sum_{s_1\in S_1} f(s_1|\theta) + \sum_{s_2\in S_2} f(s_2|\theta) = 1,$$

for all $\theta \in \Theta$.

Notice that the separability assumed here only involves the signals (i.e. the *rows* of the matrix), while nothing is assumed on the states. As an example of a family of functions that satisfy Postulate 5 we have:

$$c_n(\pi) = h\left(\sum_{s\in S} g(f(s|\theta_1),\dots,f(s|\theta_n))\right),$$

for some $g : [0,1]^n \to [0,\infty]$ and $h : [0,+\infty] \to [0,+\infty]$.

In the next postulate, instead, we restrict the way in which the cost of information depends on the state space $\Theta$. Observe that Postulates 4 and 5 both involve transformations on the space of signals. Postulate 6, in the next subsection, will impose restrictions when the space $\Theta$ is transformed.

**Additive Separability of The States**

Consider an experiment $\pi \in \mathscr{E}_n$

$$\pi = \begin{pmatrix} f(s_1|\theta_1) & \dots & f(s_1|\theta_n) \\ f(s_2|\theta_1) & \dots & f(s_2|\theta_n) \\ \dots & \dots & \dots \end{pmatrix},$$

and define its modification $\pi(\theta_1,\dots,\theta_{n-1}) \in \mathscr{E}_{n-1}$ by:

$$\pi(\theta_1,\dots,\theta_{n-1}) = \begin{pmatrix} f(s_1|\theta_1) & \dots & f(s_1|\theta_{n-1}) \\ f(s_2|\theta_1) & \dots & f(s_2|\theta_{n-1}) \\ \dots & \dots & \dots \end{pmatrix}.$$

Thanks to Postulate 2 we can redefine the latter experiment as one on the state space $\Theta' = \{\theta_1, \dots, \theta_{n-2}, \{\theta_{n-1}, \theta_n\}\}$, where the last element, $\{\theta_{n-1}, \theta_n\}$, is the set containing both $\theta_n$ and $\theta_{n-1}$ (which we identify with $\theta_{n-1}$ for simplicity).

The information obtained by $\pi$ differs from that obtained by $\pi(\theta_1, \dots, \theta_{n-1})$ in that the signals of the first experiment allow to compare the likelihood between state $\theta_n$ states $\theta_1, \dots, \theta_{n-1}$. Therefore, in the following postulate, we impose that the cost of $\pi$ is equal to the cost of $\pi(\theta_1, \dots, \theta_{n-1})$ plus the cost of learning about states $(\theta_1, \theta_n)$, $(\theta_2, \theta_n)$, ..., $(\theta_{n-1}, \theta_n)$.

To formalize this, write (consistently with the notation above):

$$\pi(\theta_i, \theta_j) := \begin{pmatrix} f(s_1|\theta_i) & f(s_1|\theta_j) \\ f(s_2|\theta_i) & f(s_2|\theta_j) \\ \dots & \dots \end{pmatrix}, \tag{2.2}$$

which is the experiment on $\{\theta_i, \theta_j\}$ *induced* by $\pi$.

**Postulate 6** (State Separability). *For all $n > 2$ and any $\pi \in \mathscr{E}_n$:*

$$c_n(\pi) = c_{n-1}(\pi(\theta_1, \dots, \theta_{n-1})) + h_{n-1}(c_2(\pi(\theta_1, \theta_n)), \dots, c_2(\pi(\theta_{n-1}, \theta_n))),$$

*where $h_n : [0, \infty]^{n-1} \to [0, \infty]$ is a smooth function.*[36]

This postulate will play a key role in the proof of the main result, as it allows to reduce the cost in dimension $n$ to cost function in lower dimensions. As it will be clear from that

[36]Even though it would be reasonable to assume some properties for $h_n$ – for example, $h_n$ increasing in all argumens – we prefer to state the postulate in the more general version and derive such properties.

proof, some results are obtained by drawing an analogy between measures of disagreement (introduced in Chapter 1) and cost functions of information structures. Postulate 6 does not have a counter part in our chapter on disagreement, but to illustrate it further, we briefly comment on how it would translate into our *disagreement* vernacular. Identifying the conditional distributions $f(\cdot|\theta)$ with the beliefs of different agents, the cost function $c_n$ can be thought of measuring *disagreement in groups*. With this analogy in mind, Postulate 6 says that disagreement in a group of $n$ agents is the sum of disagreement in a subset of $n-1$ of them, plus a function of the disagreement of the $n-th$ person, with each member of the subgroup.

**Independence**

To introduce our last postulate, let us first define the product of two experiments.

**Definition 10** (Product of Experiments). Let $\pi_1 = (S_1, (f(s|\theta))_{s\in S_1,\theta\in\Theta})$ and $\pi_2 = (S_2, (f(s|\theta))_{s\in S_2,\theta\in\Theta})$ be two experiments defined on the same state space. The product of the experiments $\pi := \pi_1 \otimes \pi_2$ is defined as

$$\pi = (S_1 \times S_2, (f(s_1|\theta)f(s_2|\theta))_{(s_1,s_2)\in S_1\times S_2,\theta\in\Theta}).^{37}$$

In other words, the experiment $\pi_1 \otimes \pi_2$ is equivalent to drawing a signal from $\pi_1$ and one from $\pi_2$, independently. Given that the experiments are independent we assume that the cost of running both $\pi_1$ and $\pi_2$ (and then observing a signal from $\pi_1 \otimes \pi_2$) is equal to the sum of the cost of $\pi_1$ and $\pi_2$.

[37] Since it was implicitly assumed that $S_1$ and $S_2$ are countably infinite, so is $S_1 \times S_2$.

**Postulate 7** (Independence). *We assume that:*

$$c_n(\pi \otimes \tilde{\pi}) = c_n(\pi) + c_n(\tilde{\pi}).$$

Notice that Postulate 6 and Postulate 7 are cardinal, i.e. if a family of functions $(c_n)_n$ satisfies them, it is not true that any increasing function of $(c_n)_n$ also does. In Chapter 1, we discuss how to relax Postulate 7 to an ordinal version, but in the case of measuring the cost of information, we think that a cardinal measure suits better the purpose of the chapter, because statements about the cost of an experiment are often cardinal in nature.

As observed in the introduction, cost functions currently used in the literature[38] do not satisfy Postulate 7. This will be a consequence of the next proposition, which for the sake of generality we state in terms of function of experiments:

**Proposition 9**. *Let $\phi : \mathscr{E}_n \to \mathbb{R} \cup \{+\infty\}$ be any function. If $\phi$ satisfies Postulate 7,[39] and $\pi$ is the fully informative experiment, then either $\phi(\pi) = 0$ or $\phi(\pi) = \infty$.*

A direct corollary of Proposition 9 involves the family of cost functions commonly used in economics. In the following Corollary, we consider the set $\Delta^\circ(\Theta)$ as the set of distributions $\mu$ with $\mu(\theta) > 0$ for all $\theta \in \Theta$. Furthermore, we will refer to $\delta_j$ as the distribution concentrated on state $\theta_j$: $\delta_j = (0, \dots, \underset{j-th}{1}, \dots, 0)$.

---

[38]See, for example, Sims (2003) for the literature in rational inattention; and Gentzkow and Kamenica (2014) for the literature in Bayesian persuasion.

[39]I.e. if:

$$\phi(\pi_1 \otimes \pi_2) = \phi(\pi_1) + \phi(\pi_2), \quad \forall \pi_1, \pi_2 \in \mathscr{E}_n.$$

**Corollary 4.** *Let $H : \Delta(\Theta) \to \mathbb{R}$ be any bounded function such that $H(\mu) > H(\delta_j)$ for all $\mu \in \Delta^\circ(\Theta)$ and $j = 1, \dots, |\Theta|$. For a fixed $\mu \in \Delta^\circ(\Theta)$, define the function $\phi_\mu : \mathscr{E}_n \to \mathbb{R}$ by:*

$$\phi_\mu(\pi) := H(\mu) - \sum_{s \in S} \mathbb{P}_\mu(s) H(\mu(s)).$$[40]

*$\phi_\mu(\pi)$ does not satisfy Postulate 7.*

Picking $H$ to be entropy:

$$H(\mu) := -\sum_\theta \mu(\theta) \log(\mu(\theta)), \tag{2.3}$$

we have that $\phi_\mu$ is the cost function typically employed in the rational inattention literature. Since $H$ is bounded, and $H(\mu) > 0 = H(\delta_j)$, for all $\mu \in \Delta^\circ(\Theta)$, Corollary 4 implies that $\phi_\mu$ does not satisfies our Postulate 7. More generally, other papers consider different functions $H$, but they also typically do not satisfy Postulate 7 (see, e.g. Ely et al. (2015)).

## 2.3 Main Result

**Theorem 6.** *For any $\Theta$ with $|\Theta| > 3$ and $S$, $c_{\Theta,S}$ satisfies Postulates 1–7 if and only if it is proportional to:*

$$\sum_s \sum_{i,j} (f(s|\theta_i) - f(s|\theta_j)) \log\left(\frac{f(s|\theta_i)}{f(s|\theta_j)}\right) = \sum_{i,j} D(f(\cdot|\theta_i), f(\cdot|\theta_j)),$$

[40]We write $\mathbb{P}_\mu(s)$ as a short-cut for $\sum_\theta \mu(\theta) f(s|\theta)$, i.e. the ex-ante probability that an agent with belief $\mu$ assigns to the signal $s$ being realized.

*where* $D$ *is the symmetric Kullback-Leibler divergence, defined by* $D(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) + \sum_i q_i \log\left(\frac{q_i}{p_i}\right)$.

For any $\theta, \theta' \in \Theta$, the function $D(f(\cdot|\theta), f(\cdot|\theta'))$ measures how divergent the conditional distributions on the space of signals are. The more divergent these distribution, the more precise the experiment, because the more clearly a signal $s$ *distinguishes* between state $\theta$ and $\theta'$. Therefore, the larger $D(f(\cdot|\theta), f(\cdot|\theta'))$, the more costly the experiment.

The divergence measure that appears in our main result

$$D(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) + \sum_i q_i \log\left(\frac{q_i}{p_i}\right),$$

can be obtained by summing the relative entropy, or Kullback-Leibler divergence, between $p$ and $q$ and that between $q$ and $p$. The Kullback-Leibler divergence is one example of convex statistical measures, in that it measure the dispersion of the vector of likelihood ratios.[41] The dispersion of likelihood ratios is measured by considering the expectation of a convex function, and this functional form ensures that Postulates 3 and 4 are satisfied. The convex function involved in the Kullback-Leibler divergence is $-\log(\cdot)$, and this is a consequence Postulate 7 – which broadly implies that products of statistical distributions are mapped into sums of costs (analogously, the logarithm maps products of positive number into sums of real numbers).

Finally, our unique cost function sums across signals (because of Postulate 5), and across states (Postulate 6). Even though this might suggest that signals and states are treated in a similar way, observe that $c_{\Theta,S}$ is separable in the signals and it is *not* separable

[41]For more information on these statistical measures, see Liese and Vajda (1987).

in the states. As a matter of fact, the cost function is tightly related to the precision of the experiment, which is measured by the relation between $f(s|\theta)$ and $f(s|\theta')$ for $\theta \neq \theta'$. Therefore some degree of complementarity between states is needed to capture the precision (and hence the cost) of the experiment.

We present, as a Corollary, a characterization of the experiments with infinite cost:

**Corollary 5.** $c_{\Theta,S}(\pi) = \infty$ *if and only if* $\pi$ *precludes a state, i.e. if* $\exists s \in S, \theta, \theta' \in \Theta$ *such that* $f(s|\theta) > 0 = f(s|\theta')$.

This implies that an experiment that precludes a state is weakly more costly than any other experiment, including the fully informative experiment. As discussed in Proposition 9, the fact that the fully informative experiment has infinite cost is related to Postulate 7 alone. On the other hand, having that an experiment that preclude a state has infinite cost is the joint effect of Postulates 6 and 7. This will be clear from the next subsection, where we comment on the proof of the main result, separating the case of $|\Theta| = 2$ (case in which Postulate 6 has no bite, and Corollary 5 does not hold); and $|\Theta| > 2$.

### Proof of the Main Result

The proof of the main result is done by induction on the dimension of the state space $n = |\Theta|$. The base case, $|\Theta| = 2$, is treated first. For this case, Postulate 6 does not apply, and we will find that there are two cost functions that satisfy Postulate 1–5 and 7. This result for $|\Theta| = 2$ is particularly useful because binary state spaces are often employed to describe uncertainty in economic models. Therefore we state it separately, as Theorem 7.

To obtain the inductive step of the proof, we rely mostly on Postulate 6, which relates cost functions $c_n$ to cost functions in lower dimensions ($c_{n-1}$ and $c_2$). In the first part of the proof, we characterize the functions $h_n$ introduced in Postulate 6; then we show that Postulate 6 constraints the cost functions for $n = 2$, and then implies that the solution to our problem is unique.

The proof of the base case relies on the axiomatization in Chapter 1, and we illustrate the analogies between the axioms and the postulates in the next subsection. The proof of the inductive step is novel, and based mostly on Postulate 6.

**Two states of the world,** $|\Theta| = 2$

**Theorem 7**. *The only cost functions $c_2$ that satisfy postulates 1–5 and 7 are:*[42]

1. $c_2^{KL}(\pi) = \sum_s [f(s|\theta_1) - f(s|\theta_2)][\log(f(s|\theta_1)) - \log(f(s|\theta_2))]$;
2. $c_2^{B}(\pi) = -\log\left(\sum_s \sqrt{f(s|\theta_1)f(s|\theta_2)}\right)$.

Observe that these two measures are *quantitatively* different, as explained in the following remark.

*Remark* 3. $c_2^B(\pi) = +\infty$ if and only if $\pi$ is fully informative. $c_2^{KL}(\pi) = +\infty$ if and only if $\pi$ precludes some state.[43]

The proof of Theorem 7 is based on the results in Chapter 1, where we propose an axiomatic definition of disagreement. The primitive of that chapters where two opinions

[42]The subscript 2 refers to the fact that $|\Theta| = 2$; the superscript $KL$ and $B$ instead stand for Kullback-Leibler and Bhattacharyya, because these cost functions are based on the Kullback-Leibler divergence and the Bhattacharyya distance, respectively. For more information, see Bhattacharyya (1946) and Kullback and Leibler (1951).

[43]An experiment $\pi$ precludes a state $\theta_1$ if there exists a signal realization $s$ with $f(s|\theta_1) = 0 < f(s|\theta_2)$. We did not provide this definition formally because it is the logical negation of defintion 7, point 2.

on a finite state space, that we call here $\Omega$ to avoid confusion with the state space of *this* chapter. Chapter 1 characterizes the set of functions $D : \Delta(\Omega) \times \Delta(\Omega) \rightarrow [0, +\infty]$ that satisfy some axioms aimed at describing *disagreement*.

The mapping that allows to use the results of Chapter 1 here maps the space of state $\Omega$ to the space of signals $S$. With this analogy, beliefs in $\Delta(\Omega)$ can be related to probability distributions in $\Delta(S)$, and this can be done by thinking of conditional distributions $f(\cdot|\theta)$ for different $\theta$'s as the beliefs of different agents. Since Chapter 1 analyzes disagreement between *two* agents, the mapping can be made only for two states of the world, $\theta_1, \theta_2$.

Even in this case, though, there is a difference between Chapter 1 and this chapter, in that when analyzing disagreement we did not constraint it to be symmetric, i.e. we did not impose that $D(p, q) = D(q, p)$ for all $p, q \in \Delta(\Omega)$. On the other hand, notice that Postulate 2 implies that the cost function does not depend on the order of states $\theta_1$ and $\theta_2$. Therefore, the cost functions satisfying our Postulates will be symmetric in $\theta_1$ and $\theta_2$.

Proposition 1 in Chapter 1 characterized all the symmetric disagreement functions satisfying the axioms of Chapter 1. Since the Postulates 1–5 and 7 imply to the axioms in Chapter 1 *and* symmetry, Theorem 7 above is analogous to Proposition 1. The exact mapping between the axioms of Chapter 1 and this chapter is postponed to the proof of Theorem 7.

**More than two states of the world,** $2 < n < \infty$

The proof of the inductive step is divided in two steps. First, in Lemma 8, we characterize the possible functions $h_n$ that define how $c_n$ depends on the cost of the experiments $c_2(\pi(\theta_1, \theta_n)), \ldots$ (see Postulate 6).

**Lemma 8.** *If a cost function $c_n$ satisfies Postulates 1,2 and 7 then $f_n$ (as defined in Postulate 6) must be the sum:*

$$f_n(x_1,\dots,x_n)=\sum_{i=1}^{n}x_i,\quad \forall (x_i)_i\in[0,\infty]^n.$$

Therefore, applying Lemma 8 inductively, we have that:

$$c_{\Theta,S}=\sum_{\theta,\theta'}c_2(\pi(\theta,\theta')).$$ [44]

Then, using the base step of Theorem 7 we have that there are two possible cost functions for general state spaces $\Theta$:

$$c_{\Theta,S}(\pi)=\sum_{\theta,\theta'}c_2^{KL}(\pi(\theta,\theta')),$$

or:

$$c_{\Theta,S}(\pi)=\sum_{\theta,\theta'}c_2^{B}(\pi(\theta,\theta')).$$

The last part of the proof, Lemma 9, shows that in the latter case (in which divergence is measured via the Bhattacharyya distance) $c_{\Theta,S}$ does not satisfy Postulate 5 for any $\Theta$ with $|\Theta|>2$. This is due to the fact that $c_2^B(\pi)$ is a logarithm of a sum, therefore summing across pairs of states we obtain a function that is not separable in the signals.

[44] Recall that $\pi(\theta,\theta')$ was defined as the experiment induced on states $\theta$ and $\theta'$, see equation (2.2).

**Lemma 9.** *For any $n = |\Theta| > 2$, the function*

$$c_{\Theta,S}(\pi) = \sum_{\theta,\theta'} c_2^B(\pi(\theta,\theta')).$$

*does not satisfy Postulate 5.*

This Lemma allows to conclude that there is a unique set of cost functions compatible with our postulates (Theorem 7) and therefore concludes the proof.

## 2.4 Analysis of the cost function

In this section, we analyze some properties of the cost function characterized in the main body of this chapter:

$$c(\pi) = \sum_s \sum_{i,j} (f(s|\theta_i) - f(s|\theta_j)) \log\left(\frac{f(s|\theta_i)}{f(s|\theta_j)}\right) = \sum_{i,j} D(f(\cdot|\theta_i), f(\cdot|\theta_j)),$$ [45]

and we write $c(\cdot)$ instead of $c_n(\cdot)$ or $c_{\Theta,S}(\cdot)$ for brevity, whenever unambiguous.

### Discontinuity at the non-informative experiment

The first result shows that $c(\cdot)$ is discontinuous with respect to the usual norm on the space of matrices. Let $\pi_n \in \mathscr{E}_{\Theta,S}$ be a sequence of experiments with conditional distributions $f^n(s|\theta)$, and let $\pi \in \mathscr{E}_{\Theta,S}$ have conditional distributions $f(s|\theta)$. We say that $\pi^n$ converges

[45]We will assume that $c$ takes this functional form even when $|\Theta| = 2$, even though in this case there are two cost functions that satisfy our postulates. As a matter of fact, any example we consider with two states of the world can be changed to allow for $|\Theta| > 2$.

in norm to $\pi$ (and write $\pi^n \to \pi$) if:

$$\sup_{s \in S, \theta \in \Theta} |f^n(s|\theta) - f(s|\theta)| \to 0. \tag{2.4}$$

**Proposition 10.** *Let $\pi$ be the non-informative experiment. There exists a sequence of experiments $(\pi^n)_n$ such that $\pi^n \to \pi$ and $c(\pi^n) \to \infty$. On the other hand, if for some sequence $(\pi^n)_n$ we have that $c(\pi^n) \to 0$, then $\pi^n \to \pi$ non-informative.*

This proposition implies that $c(\cdot)$ is not continuous around the non-informative experiment, and it then does not satisfy one of the conditions in Hébert and Woodford (2016). The reason for this discontinuity is related to the fact that our cost function consists in averaging the logarithm of the likelihood ratio $f(s|\theta)/f(s|\theta')$ for different states of the world. The fact that the logarithm is discontinuous at $0$ implies that our cost function is discontinuous at the non-informative experiment. Notice also how $c_B^2$, as defined in Theorem 7, is instead continuous with respect to any norm (as it follows from Proposition 4, in Chapter 1).

## Cost of information as a function of the posterior beliefs

As we argued in this chapter, a cost function based on the experiment primitives (the conditional distributions $(f(s|\theta))_{s,\theta}$) is more desirable than one based on the prior and posterior beliefs of the experimenter. Nonetheless, it is useful to rewrite our cost function (equation (2.4)) in terms of the posterior belief of an agent with prior $\mu$. This rewriting allows for a simpler comparison of the cost function we propose with others used in the literature. Also, rewriting our cost function in terms of posteriors allows to apply directly some results on the strategic acquisition of information.

In the next lemma, for a given belief $\mu \in \Delta(\Theta)$, we will used the notation introduced at the beginning of this chapter. $\mathbb{P}_\mu(s) := \sum_\theta \mu(\theta) f(s|\theta)$ denotes the ex-ante probability of observing signal $s$, and $\mu(\theta|s)$ is the posterior belief on $\theta$, after observing $s$. It will be useful to denote, for any $\mu \in \Delta(\Theta)$, $g(\mu) := \sqrt[|\Theta|]{\mu(\theta_1) \cdots \mu(\theta_n)}$, the geometric mean of $\mu(\theta_1), \ldots, \mu(\theta_n)$. Without loss of generality we also assume that $\mu(\theta) > 0$ for all $\theta \in \Theta$.

**Lemma 10.** *For any $\mu \in \Delta(\Theta)$ with $\mu(\theta) > 0$ for all $\theta \in \Theta$, the cost of an experiment $\pi$ that does not preclude any state can be written as:*

$$c(\pi) = \sum_s \mathbb{P}_\mu(s) \sum_\theta \frac{\mu(\theta|s)n}{\mu(\theta)} \left(\log(\mu(\theta|s) - \log[a(\mu(s))]\right)) =: \sum_s \mathbb{P}_\mu(s) \phi_I(\mu(s); \mu),$$

*where we defined $\phi_I(\mu(s); \mu) := \sum_\theta \frac{\mu(\theta|s)n}{\mu(\theta)} \left(\log(\mu(\theta|s) - \log[a(\mu(s))]\right))$.*

Rewriting the cost of information as in Lemma 10 allows to apply our cost function to problems such as costly Bayesian persuasion (Gentzkow and Kamenica (2014)) and in general to the typical framework of the rational inattention literature. Indicating with $u : A \times \Theta \to \mathbb{R}$ any utility function, we can define by $V(\mu) := \max_a \sum_\theta u(a,\theta)\mu(\theta)$ the value function of an agent with belief $\mu$. When choosing how much information to acquire, then, a rationally inattentive agent solves the problem

$$\max_\pi \sum_s \mathbb{P}_\mu(s) V(\mu(s)) - c(\pi),$$

where $c$ represents the cost of information. In our case, then, we find that the problem of finding the optimal experiment is equivalent to choosing the optimal set of posteriors to maximize:

$$\max_{(\mu(s))_s} \sum_s \mathbb{P}_\mu(s)(V(\mu(s)) - \phi_I(\mu(s);\mu)), \quad \text{subject to } \sum_s \mathbb{P}_\mu(s)\mu(s) = \mu, \tag{2.5}$$

where the latter condition ensures that all posteriors are Bayes plausible, i.e. they correspond to the signal realizations of an experiment $\pi$. Analogously, we can write a general persuasion problem in terms of the posteriors only, thus being able to use the concavification results of Kamenica and Gentzkow (2011).

In the following lemma, we compare our cost function with the one typically employed in the literature:

$$c_H(\pi;\mu) := \sum_s \mathbb{P}_\mu(s)(H(\mu) - H(\mu(s))) =: \sum_s \mathbb{P}_\mu(s)\phi_H(\mu(s);\mu),$$

with $H(\mu) = -\sum_\theta \mu(\theta)\log(\mu(\theta))$ entropy. Since both our cost function and this are written as averages of functions of posteriors $\mu(s)$, we only compare the two functions $\phi_I$ and $\phi_H$.

**Lemma 11**. *For all $\mu \in \Delta(\Theta)$, we have that:*

- $\phi_H(\mu;\mu) = \phi_I(\mu;\mu) = 0$*;*
- $\phi_H(\cdot;\mu)$ *and* $\phi_I(\cdot;\mu)$ *are convex function;*
- $\lim_{x\to\partial\Delta(\Theta)} \phi_H(x;\mu) < \infty$ *whereas* $\lim_{x\to\partial\Delta(\Theta)} \phi_I(x;\mu) = \infty$*.*[46]

[46]The set $\partial\Delta(\Theta)$ denotes the boundary of $\Delta(\Theta)$, formally:

$$\partial\Delta(\Theta) = \{\mu \in \Delta(\Theta) \mid \mu(\theta), \exists\theta\}$$

The first statement of Lemma 11 reflects the property that it is always costless to produce a non-informative signal, i.e. a signal for which $\mu(s) = \mu$. The second property is a consequence of the fact that both functions are increasing in the Blackwell informativeness of the experiment. This shows that our cost function (similarly to the one typically used in the literature) satisfies the conditions imposed in Gentzkow and Kamenica (2014). Therefore our cost function is well-suited for studying costly Bayesian persuasion. Finally, the last statement reflects the main difference between the two cost functions. While under our cost functions ruling out one state (i.e. generating a posterior in the boundary of $\Delta(\Theta)$) is infinitely costly, for the cost function that considers the average change in entropy such cost is always finite. This, in particular, shows that in problems such as those of equation (2.5) the solution will always be interior (using our cost function).

Chapter 3

# *Matching in the Smallest Large Market*[47]

## 3.1 Introduction

This paper analyzes a marriage market with countably infinite agents. This chapter is motivated by a mathematical curiosity of analyzing one-to-one two sided matching markets when agents are countably infinite, as to our knowledge no other author has studied this case. The economic literature on large matching markets has focused on the case of a continuum of agents, or on the analysis of finite markets with number of agents growing to infinity. Therefore this chapter provides the missing intermediate case.

Another reason for studying these matching markets comes from the study of dynamic matching markets. If one considers a finite market opening every year, in which agents can prefer being matched with agents entering the market in the future (or that entered in the past), then a complete description of the preferences must involve all the years in which the market opens. If the market is supposed to open forever, then the "smallest" model of such market is one with countably infinite agents, or, more precisely, a countably infinite union of finitely many agents entering the market in a given year. In this sense our model provides a benchmark for the study of dynamic markets, and the results obtained

[47]I would like to thank Francesco Caravenna, Alessandra Casella, Yeon-Koo Che, Navin Kartik, Jacob Leshno, Lorenzo Rocco, Paolo Siconolfi, and Olivier Tercieux for their advice and feedback.

here can be used for the design of matching in a dynamic environment (in particular the lattice structure of the set of stable matching, and the failure of the Rural Hospital Theorem).

This paper shows that when agents are countably infinite, the set of stable matchings is non-empty and satisfies the lattice structure observed in the finite case, provided agents' preferences admit a maximum on each subset. This is not a very stringent assumption, and in particular it is met in matching environments in which all agent finds only finitely many agent of the opposite set *better than remaining unmatched.* Contrary to the finite case, we show that in our setting the set of "matched" men or women is not constant across stable matching: in other words there can be agents who are single in some stable matching, and matched in other. Hence the so-called Rural Hospital Theorem fails.

We show that if all agents' preferences satisfy our assumption on the maximum, there exists a stable matching in the market, and we show that besides being sufficient this assumption is necessary, too. In other words, if one agent's preferences do not admit maximum on some subset of the agents on the opposite side, then there might not exist any stable matching. The existence of a stable matching cannot be obtained directly by applying the Deferred Acceptance (henceforth DA) algorithm proposed by Gale and Shapley (1962), because such algorithm relied on the finiteness of the number of agents to be well-defined. For this reason, we provide a modification of it (the "generalized Deferred Acceptance algorithm") that is well-defined with countably many agents, and yields a stable matching.

After proving that the set of stable matching is non-empty, we show that it is a complete lattice under the partial order of "men-preferred" matching.[48] A matching is said to be "men-preferred" to another matching whenever all men are matched to a (weakly) better woman under the former. The generalized DA algorithm in which men propose to women is easily proved to be the men optimal stable matching, and analogously one can see that the matching obtained by women proposing is men pessimal. Given any two stable matching, we can define the operation of meet and join, and show that they yield a well-defined stable matching, and thus proving that the set of stable matching constitutes a complete lattice.[49]

Contrarily to what happens for finite markets, we find that the set of matched agents is not constant across stable matchings. We provide a simple example in which a man is matched under the "man-optimal" stable matching, and he is single under the "woman-optimal" stable matching. We further provide necessary and sufficient conditions for the Rural Hospital to hold and discuss the implications that this has on the manipulability result of the finite case.[50] This failure of the Rural Hospital Theorem distinguishes our model from the finite one, and in fact from the other extensions to infinitely many agents modeled with a continuum set. As a matter of fact, in the models with a continuum of agents the measure of agents is finite, and this ensures that the measure of matched agents

[48]Given the symmetry in the problem, the same holds under the "women-preferred" order. We present our results only with respect of men for the sake of brevity.

[49]Alternatively, the same result could have been obtained by framing the problem as a fixed point problem as firstly noted by Adachi (2000) and then invoking the Tarski fixed point theorem (see Tarski (1955)). This approach would have been more abstract, so we prefer to obtain the results directly, without relying on stronger mathematical theorems.

[50]In the study of the finite marriage market, one obtained that whenever the set of stable matching is not a singleton, at least one agent has incentive to misreport his/her preference under any stable mechanism. This result was proved using the Rural Hospital theorem, and we show that whenever the Rural Hospital theorem fails, the result on manipulability might fail too. In some sense, this shows that the Rural Hospital theorem is necessary and sufficient for manipulability.

is constant across stable matchings. In our setting, the same argument would not provide the desired result, in that the *measure* of matched agents can be constantly infinite across stable matchings, without the *set* of matched agents being constant.

This result has consequences for the related literature in large matching markets, and dynamic matching markets. Firstly, if one consider the market with countably many agents as the limit of finite markets, then we showed a discontinuity in the limit. Secondly, when modeling finite markets that open infinitely many times if agents include in their preferences agents who enter the market in different years,[51] then we might have that the Rural Hospital theorem does not hold in the complete market (even though it does in any finite market with "truncated" preferences). We elaborate on this in Section 3.5, where we show how the failure of the Rural Hospital theorem yields a criterion for choosing between matching mechanisms in a dynamic matching market (Theorem 12).

The rest of the paper is structured as follows: in Section 3.2 we introduce the model, and in Section 3.3 we analyze the properties of the set of stable matching that hold for both finite and infinite marriage models. In Section 3.4, we show the main difference between the finite model and the infinite one, that is the failure of the Rural Hospital Theorem. In Section 3.5 we summarize how the results obtained apply to a dynamic market, and in Section 3.6 we conclude.

[51]An example of this would be that of a high-school graduate who considers taking a year off before going to college. A complete description of her preferences would then include a college assignment this year and an assignment one year in the future.

## Related Literature

The model we analyze was initially proposed with finitely many agents by Gale and Shapley (1962), who also introduced the Deferred Acceptance algorithm. In this paper, we show that a modification of the same algorithm yields a stable matching also when there are countably many agents. Some of the further extensions and modifications of that model include Abdulkadiroglu et al. (2005), Abdulkadiroglu and Sonmez (2005), Alkan (1988), Ostrovsky (2008), Roth (1984), Roth (1986), Peranson and Roth (1999), Roth et al. (2004), Shapley and Shubik (1971), etc. A more detailed review of such applications can be found in Roth (2008).

Many of the proofs of the results that we obtain in Section 3.3 are adapted from Roth and Sotomayor (1992), who provide an extensive analysis of the marriage model with finitely many agents. A different approach to the problem of many to one matching is given in Hatfield and Milgrom (2005). While our notation follows that of Roth and Sotomayor (1992), it is worth noting that our results of Section 3.3 could be obtained by adapting Hatfield and Milgrom (2005) to our case of countably many agents.

Our results are related to the extensions of the classical model to large markets. The approaches adopted in the literature are essentially two-fold: Che et al. (2015) and Azevedo and Leshno (2016) model the set of agents as a continuum space, whereas Immorlica and Mahdian (2005) and Kojima and Pathak (2009) analyze finite markets with a number of agents increasingly large. Our theory is also related to the theory of dynamic matching markets, i.e. finite matching markets opening every year, possibly infinitely many times. Kurino (2008) defines dynamic matching as a sequence of one-period matching, and the same approach is taken by Kurino (2014) and Kadam and Kotowski (2016). Our approach is markedly different, in that the object of interest is a unique matching in the infinite market. Other papers in the literature in dynamic matching markets include Monte and Tumennasan (2015), who analyze allocations in multiple markets (for which a special case is intertemporal allocation).

## 3.2 The Model

The model is analogous to the finite model proposed by Gale and Shapley (1962), with the extension that now the set of men and women is countably infinite, $|M| = |W| = +\infty$.

Let $M = \{m_0, m_1, m_2, \dots\}$ be the set of men and $W = \{w_0, w_1, w_2, \dots\}$ the set of women. Each man $m_j$ (resp. woman $w_n$) has a *strict total preference* $>_{m_j}$ on the set $W \cup \{m_j\}$ (resp. $M \cup \{w_n\}$), i.e. an irreflexive, asymmetric, and transitive relation on the set $W \cup \{m_j\}$ (resp. $M \cup \{w_n\}$). By writing:

$$w_1 >_{m_j} w_2 >_{m_j} m_j >_{m_j} w_3 \tag{3.1}$$

we mean that $m_j$ prefers $w_1$ to $w_2$, and would rather be single than being matched to $w_3$.

We assume that all agents have a preferred element in each subset, formally:

**Assumption 1**. *For each $m_j$, (resp. $w_n$) and for every $A \subseteq W \cup \{m_j\}$ (resp. $A \subseteq M \cup \{w_n\}$), $\exists x \in A \cup \{m_j\}$ (resp. $A \cup \{w_n\}$) such that:*

$$x >_{m_j} y, \quad \forall y \in A \setminus \{x\} \quad \textit{(resp. } >_{w_n}).$$

Notice that we assume that the maximum for agent $m_j$ exists on any subset $A \subset W \cup \{m_j\}$. This assumption is stronger than just requiring the existence of a "global maximum" on $W \cup \{m_j\}$. For example, we could have a man preferring $w_0$ over all other women, and preferring $w_j$ to $w_{j'}$ whenever $j > j'$, $j' \neq 0$. While such preferences admit a global maximum ($w_0$), there exists no maximum on the set $\{w_1, w_2, \ldots, w_n, \ldots\}$. As Lemma 12 below shows, assuming the existence of global maximum is not sufficient for the existence of a stable matching.

Observe also that Assumption 1 is trivially met when the set $A$ is finite - since we assume the preferences to be complete. On the other hand when $A$ is infinite a strict preference might not admit a maximum on $A$. We will show that this assumption is both necessary and sufficient for the existence of a stable matching.

If an agent's preferences satisfy Assumption 1, we can represent them by an infinite sequence. Then, besides the notation in (3.1), we will also write:

$$P(m_j) : w_{j_1}, w_{j_2}, \ldots, w_{j_l}, m_j, w_{j_{l+1}}, \ldots, w_{j_m}, \ldots; \tag{3.2}$$

to indicate that $m_j'$s most preferred agent is $w_{j_1}$, and that in general $w_{j_k}$ is preferred to $w_{j_{k'}}$ for all $k < k'$. Each woman ranked *after* $m_j$ is worse than being single, and we will truncate such list to $m_j$, because the ranking of women thereafter is irrelevant for the scope of our paper (as it will be clear from the definition of stable matching below, Definition 13). Then (3.2) will be written as:

$$P(m_j) : w_{j_1}, w_{j_2}, \ldots, w_{j_l},$$

with the understanding that a woman not listed in $P(m_j)$ is *worse than being single*, or *unacceptable*.

*Remark* 4. Observe that if a man $m$ has finitely many acceptable women, then his preferences satisfy Assumption 1. While our model allows also for infinitely many acceptable partners, all the results we will obtain equally hold whenever agents find only finitely many agents acceptable (a very natural assumption in many applications, in particular in dynamic matching markets).

Formally, we can define a marriage market as follows:

**Definition 11** (Marriage Market). A Marriage Market is defined by a triple $(M, W, \mathbf{P})$ where:

- $M = \{m_0, m_1, \dots\}$ is the set of men;
- $W = \{w_0, w_1, \dots\}$ is the set of women;
- $\mathbf{P} = (P(x))_{x \in M \cup W}$ where for each $x \in W$ (resp. $x \in M$):
  - $P(x)$ is an irreflexive, asymmetric, and transitive relation on $W \cup \{x\}$ (resp. $M \cup \{x\}$);
  - $P(x)$ satisfies Assumption 1;

A matching is a set of pairs of agents, each pair consisting of one man and one woman, with unpaired agents remaining single. Mathematically, a matching is defined as:

**Definition 12** (Matching). A matching $\mu$ is a function:

$$\mu : M \cup W \to M \cup W,$$

such that:

1. for all $x \in M \cup W$, $\mu(\mu(x)) = x$;
2. if $x \neq \mu(x)$, then $\mu(x) \in M \Leftrightarrow x \in W$;

We interpret $\mu$ as mapping an agent $x$ to itself if and only if the agent is left single. The second property then requires that if an agent is not single, then s/he is matched to an agent in the other set. We will study the subset of matching that satisfy stability:

**Definition 13** (Stable Matching). A matching $\mu$ is stable in the marriage market $(M, W, \mathbf{P})$ if:

- *(individual rationality)*: there exists no $x$ such that:

$$x >_x \mu(x);$$

- *(non-existence of blocking pairs)*: there does not exist any pair $(m, w)$ such that:

$$m >_w \mu(w) \quad \text{and} \quad w >_m \mu(m).$$

Individual rationality means that no agent is matched to an agent that s/he likes worse than being single. Non-existence of blocking pairs means that there does not exist any pair that would rather be matched to each other than being matched according to $\mu$ (to put it in other words: there does not exist any blocking pair if anytime an agent $x$ prefers $y$ to his/her match, then $y$ prefers her/his match to $x$).

## 3.3 Analysis of the set of Stable Matching

For each marriage market $(M, W, \mathbf{P})$, we will be interested in the set of stable matching. Before passing to this analysis, let us show by means of an example that Assumption 1 is necessary for the existence of a stable matching.

**Lemma 12.** *If the preferences $P(x)$ of an agent $x$ do not satisfy Assumption 1 then there exist a market $(M, W, \boldsymbol{P})$ in which there are no stable matching.*

*Proof.* The proofs are in Section C.2. □

If preferences satisfy Assumption 1, instead, there exists at least a stable matching. In the following subsection, we prove existence by showing how the algorithm introduced by Gale and Shapley (1962) extends to our case of infinitely many agents.

### Generalized Deferred Acceptance algorithm

The man proposing DA algorithm, as defined for finite markets, consists in the following procedure:

1. All men propose to the best acceptable woman in their preference list (if any);
2. All women who got at least an acceptable proposal, retain the best one, and reject the remaining ones. If a woman receives no acceptable proposal she rejects them all;
3. If a man has been rejected, in the second step of the algorithm he proposes to the second best acceptable woman (if any);

4. All women with multiple proposals (of which at most one can be the *retained one* from the previous period) pick the best one;
5. and so on...

*Remark* 5. The Deferred Acceptance algorithm can be defined with men proposing, and with women proposing. We will give the definition for men, but the one for women is analogous.

A key property of the DA algorithm in markets with finitely many agents is that it ends in finitely many steps - as no man can propose to the same woman more than once (see Gale and Shapley (1962)). Hence the matching provided by the algorithm is well-defined, and a simple argument proves it must also be stable. Obviously, when dealing with countably many agents the DA algorithm might not end in finitely many steps (even if we assume that each agent finds only finitely many agents acceptable):

**Example 2.** *Consider the following preference scheme:*

$$
\begin{array}{ll}
P(m_1) : w_1 & P(w_1) : m_1, m_2 \\
P(m_2) : w_1, w_2 & P(w_2) : m_2, m_3 \\
P(m_3) : w_2, w_3 & P(w_3) : m_3, m_4 \\
\quad \ldots\ldots & \quad \ldots\ldots \\
P(m_i) : w_{i-1}, w_i & P(w_i) : m_i, m_{i+1} \\
\quad \ldots\ldots & \quad \ldots\ldots
\end{array}
$$

*Following the algorithm described above, in the first step $w_1$ gets proposed to by $m_1$ and*

$m_2$*, and rejects* $m_2$ *who then is left unmatched. In the second step* $m_2$ *proposes to* $w_2$*, who was withholding the proposal of* $m_3$*, but prefers* $m_2$ *and thus rejects* $m_3$*, who is temporarily left unmatched. In general, at the* $i$*-th step, we have that the tentative matching leaves unmatched* $m_{i+1}$*, who was rejected by* $w_i$ *and the proposes to* $w_{i+1}$*, who then rejects* $m_{i+2}$*. Then in this market the DA algorithm does not end in finitely many iterations.*

Even though the original algorithm is infinite, it is plain to see from Example 2 that we can tweak the original algorithm to find a generalized version that is well-defined also with countably many agents. The idea is to define a "limit" matching that as we will see in Theorem 8 below is stable for any set of preferences.

**Definition 14** ((Man Proposing) - Generalized DA algorithm). Consider the DA algorithm as defined in the finite market. If the algorithm finishes in a finite number of steps, call $\mu_M$ the matching obtained when in the first step in which there are no rejections.

If the algorithm is infinite, define $\mu^{(j)}(\cdot)$ to be the *tentative* matching arising in the $j$-th step of the algorithm and then define the limit function $\mu_M$ as follows:[52]

$$\mu_M(m) := \begin{cases} \lim_{j\to\infty} \mu^{(j)}(m) & \text{if} \quad (\mu^{(j)}(m))_{j=1}^{+\infty} \quad \text{is eventually constant;} \\ m & \text{otherwise.} \end{cases}$$

And complete the matching in the natural way: if $\mu_M(m) = w$ for some $m$ and $w$, then let $\mu_M(w) = m$. Otherwise let $\mu_M(w) = w$.

*Remark* 6. Observe that we implicitly used Assumption 1 in the definition of the general-

[52] We consider the limit $\lim_{j\to\infty} \mu^{(j)}(m)$ only for sequences $(\mu^{(j)}(m))_{j=1}^{+\infty}$ that are eventually constant and therefore its meaning is independent of the topology one considers on the discrete $W$.

ized DA algorithm in two parts. First, once a man's proposal is rejected, the man passes to the next best woman and if his preferences do not satisfy Assumption 1, then there might not exist a *next* woman. Secondly, we assumed that a woman receiving multiple proposals picks the best one - and hence we implicitly assumed that according to her preferences the *best* proposal exists.

**Theorem 8**. *The $\mu_M$ so defined is a stable matching.*

As a corollary, we get that the set of stable matching is never empty, as $\mu_M$ is well defined for any set of preferences. Similarly, if we instead used the generalized woman proposing DA algorithm we would have found a stable matching that we denote $\mu_W$.

## Ranking of Stable Matching

In general, the set of stable matching is not a singleton - and it has been shown that in the model with finitely many agents it is possible to define an order on the set of matching under which the structure of stable matching is a lattice.

Similarly to what is done in the literature, we define the following (partial) order, that formalizes what it means to say that all men prefer matching $\mu$ to $\mu'$:

**Definition 15** ($>_M$ order). Given any two stable matching $\mu$ and $\mu'$ we say that:

$$\mu >_M \mu' \Leftrightarrow \begin{cases} \mu(m) \geq_m \mu'(m) & \forall m \in M; \\ \\ \mu(m) >_m \mu'(m) & \exists m \in M; \end{cases}$$

In the symmetric way we can define the women order $>_W$. Whenever $\mu >_M \mu'$ we will say that $\mu$ *is preferred to* $\mu'$ *by men* or that $\mu$ is men-preferred to $\mu'$. It is clear that such strict order relation is irreflexive, antisymmetric, and transitive. Also, this order is, in general, not a complete order.[53]

A result that carries through in our study of countably many agents is that under $>_M$, the matching $\mu_M$ is the maximal element within the set of stable matching, that is:

**Theorem 9** (Optimality of $\mu_M$ (under $>_M$)). *For any $\mu$ stable, we have that $\mu_M \geq_M \mu$.*

And obviously, the symmetric statement for women also holds:

$$\mu_W \geq_W \mu, \quad \text{for all } \mu \text{ stable.}$$

Another well-known result that we can prove in our setting is that the ranking $>_M$ and $>_W$ are opposite orders, in the sense specified in the following lemma:

**Lemma 13.** *Let $\mu$ and $\mu'$ be stable matching of the marriage market $(M, W, \boldsymbol{P})$, we have that:*

$$\mu >_M \mu' \Leftrightarrow \mu' >_W \mu.$$

The hypothesis that both matching be stable is key for the statement of Lemma 13. This lemma implies a simple corollary:

[53] That is to say: there may exist $\mu$ and $\mu'$ such that neither $\mu >_M \mu'$ nor $\mu' >_M \mu$, because some men are better matched under $\mu$ and some other under $\mu'$.

**Corollary 6.** *For all $\mu$ stable we have that:*

$$\mu_M \geq_M \mu \geq_M \mu_W,$$

*so in particular $\mu_W$ is man pessimal.*

*Remark* 7. The properties proved thus far (and the ones of the next subsection) carry through from the finite case to our setting with countably many agents essentially because their proofs rely on the definition of stability only. Whenever the result can be obtained by assuming the counter positive (i.e. the logical negation of the thesis) and finding a blocking pair, the same kind of argument can be replicated here – and this is essentially what we do in the proofs of the appendix. The same does not hold for the Lattice Theorem, and the Rural Hospital Theorem (because their proos also rely on the finiteness of the set of agents).

### Lattice Theorem

In this subsection, we will show that the set of stable matching endowed with the men-preferred order has the algebraic structure of a complete lattice. Corollary 6 showed that there exists a max and a min in the set of stable matching ordered with the man-preferred relation. In what follows, we will show that *any* family of stable matching admits a sup and an inf according to the order just defined.

A lattice is couple $(A, \succeq)$ where $A$ is a set, and $\succeq$ is a partial order on $A$. We call $(A, \succeq)$ a lattice if it is closed under the operation of meet ($\vee$) and join ($\wedge$).[54] Given any

[54]It is convenient to think of $\wedge$ and $\vee$ analogously to the inf and sup on the reals.

two elements $a, b \in A$, the element $c := a \vee b$ is defined by the following properties: (i) $c \succeq a$ and $c \succeq b$; (ii) if $c' \succeq a$ and $c' \succeq b$, then $c' \succeq c$. In other words, $a \vee b$ is the least of the elements larger than $a$ and $b$. Symmetrically, we can define the join of $a$ and $b$, $c := a \wedge b$, as the element such that: (i) $c \preceq a$ and $c \preceq b$; (ii) if $c' \preceq a$ and $c' \preceq b$ then $c' \preceq c$.

In the context of stable matching, the order we use is the men-preferred order (Definition 15), and the operations of $\vee$ and $\wedge$ can be explicitly defined as follows:

$$\mu \vee \nu(x) := \begin{cases} \max_{\geq_x}\{\mu(x), \nu(x)\} & \text{if } x \in M \\ \\ \min_{\geq_x}\{\mu(x), \nu(x)\} & \text{if } x \in W \end{cases} \tag{3.3}$$

and:

$$\mu \wedge \nu(x) := \begin{cases} \min_{\geq_x}\{\mu(x), \nu(x)\} & \text{if } x \in M \\ \\ \max_{\geq_x}\{\mu(x), \nu(x)\} & \text{if } x \in W \end{cases}. \tag{3.4}$$

For any man $m$, $\mu \vee \nu(m)$ is defined as the best of the matching of $m$, according to his preferences. For all women, instead, the matching $\mu \vee \nu(w)$ is the *worst* of the two. It is then clear that for all man $m$:

$$(\mu \wedge \nu)(m) \leq_m \mu(m), \nu(m) \leq_m (\mu \vee \nu)(m).$$

What is not obvious is the fact that both $\mu \vee \nu$ and $\mu \wedge \nu$ are stable matching - i.e. the set of stable matching is closed under the operation of $\vee$ and $\wedge$:

**Theorem 10.** *Let $\mu$ and $\nu$ be two stable matching on $(M, W, \boldsymbol{P})$, the functions $\vee$ and $\wedge$*

*defined in* (3.3) *and* (3.4) *yield stable matching.*

Furthermore, if a set endowed with a partial order admits meet and join of arbitrary families of elements (instead of just two), then it is called a *complete lattice.* As a matter of fact, the proof of Theorem 10 defines the sup $\lambda = \mu \vee \nu$, but a similar argument can be done with the generalized sup $\lambda := \bigvee_{i \in I} \nu_i$, hence we have that the set of stable matching with the operations $\vee$ and $\wedge$ is a complete latice.[55]

*Remark* 8. All the results obtained so far could alternatively be obtained by framing the problem in a more abstract space, and reducing the problem of finding the set of stable matching to a fixed-point problem. This is the approach adopted in Adachi (2000) and Hatfield and Milgrom (2005), where they study many-to-one matching in a finite setting. They introduce a notation under which a stable matching is a fixed point of an operator that matches couple of sets to couple of sets. Once this characterization is obtained, Tarski Fixed Point Theorem (see Tarski (1955)) implies non-emptiness of the set of stable matching, lattice structure of such set, and its completeness.

The same approach would work in our setting, essentially because Tarski Fixed Point Theorem works in very general settings, including one with countably infinite sets.

Such approach, though, would make the results less clear - in our opinion - and so we decided to provide a direct proof of all those results, without relying on stronger mathematical theorems.

*Remark* 9. In the finite setting, a *characterization* of the lattice structure was obtained by

[55]More precisely the proof of Theorem 10 assumed that the operation $\mu \vee \nu$ did not define a stable matching, and obtained that $\mu$ or $\nu$ then could not be stable. If instead one considers $\lambda := \bigvee_{i \in I} \nu_i$ (i.e. the sup of an arbitrary family) one can use the same argument to prove that if $\lambda$ is not a stable matching, then $\exists\, i \in I$ such that $\nu_i$ is not stable.

Gusfield et al. (1987), who proved that not only is the set of stable matching a complete distributive lattice, but actually also the opposite holds: every finite distributive lattice is a set of stable matching for some matching market. This shows that set of stable matching and distributive lattices are effectively the same mathematical object.

We do not investigate the same statement in this paper, but we doubt that the same happens when the set of agents is countably infinite. In fact we conjecture that when agents are countably infinite the set of stable matching is either finite or uncountable. If this conjecture was true, then quite clearly it would not be true that every distributive lattice is a set of stable matching – as no countably infinite distributive lattices could be found as set of stable matching.

## 3.4 Rural Hospital Theorem

A well-known result in the theory of stable matching with finitely many agents is that the set of matched agents is constant across stable matching. Precisely, define the set of non-single men and women in a given matching $\mu$ as:

$$M(\mu) := \{m \in M \mid \mu(m) \neq m\},$$

$$W(\mu) := \{w \in W \mid \mu(w) \neq w\}.$$

**Definition 16** (Rural Hospital Theorem). We say that the rural hospital theorem holds if for all stable matching $\mu$ and $\mu'$:

$$M(\mu) = M(\mu'), \text{ and } W(\mu) = W(\mu'),$$

Whenever the set of agents is finite, it was proved that the Rural Hospital Theorem holds, see Roth (1986). In our infinite framework, this need not be the case. Before analyzing why this anomaly arises, let us show an example of a marriage market $(M, W, \mathbf{P})$ in which there exist $\mu$ and $\mu'$ stable such that $M(\mu) \neq M(\mu')$.[56]

**Example 3** (Failure of the Rural Hospital Theorem). *Consider the following preferences:*

$$
\begin{array}{ll}
P(m_1) : w_1 & P(w_1) : m_2, m_1 \\
P(m_2) : w_2, w_1 & P(w_2) : m_3, m_2 \\
P(m_3) : w_3, w_2 & P(w_3) : m_4, m_3 \\
\\
\ldots\ldots & \ldots\ldots \\
P(m_i) : w_i, w_{i-1} & P(w_i) : m_{i+1}, m_i \\
\ldots\ldots & \ldots\ldots
\end{array}
$$

*In this example, there are two stable matching, the one obtained by man proposing DA algorithm, $\mu_M$, and the one obtained by woman proposing, $\mu_W$. It is plain to see that under the man-proposing algorithm we get $\mu_M(w_i) = m_i$ for all $i \in \mathbb{N}$; whereas when women*

[56]Jagadeesan (2016) independently found the same example we illustrate here. He goes on to prove that despite the Rural Hospital theorem fails, the matching mechanism remain group strategy-proof.

*propose we obtain* $\mu_W(w_i) = m_{i+1}$ *for all* $i \in \mathbb{N}$*, and* $\mu_M(m_1) = m_1$*. Therefore we have that:*

$$M(\mu_M) \neq M(\mu_W),$$

*because* $m_1 \in M(\mu_M) \setminus M(\mu_W)$.

The failure of the Rural Hospital Theorem implies another difference between the model with finitely many agents and the extension to infinitely many agents analyzed here. In the finite model, it has been proved that whenever the stable matching is not unique, under *any stable mechanism* at least one agent has incentive to misreport his/her preferences - assuming everybody else reports truthfully (see Theorem 4.6 in Roth and Sotomayor (1992)). When there are countably many agents, instead, it might be that *no agent has incentives to misreport his/her preferences* (even when there are multiple stable matchings). We show this in Appendix C.1.

Even though the Rural Hospital Theorem does not hold in our setting, the following result is still true:

**Lemma 14**. *Let* $\mu$ *and* $\mu'$ *be stable matching. If* $\mu \leq_M \mu'$ *then:*

$$M(\mu) \subseteq M(\mu').$$

*In particular, for any* $\mu$*, we have that:*

$$M(\mu_W) \subseteq M(\mu) \subseteq M(\mu_M);$$

$$W(\mu_M) \subseteq W(\mu) \subseteq W(\mu_W).$$

In particular, this lemma explains why in the finite case we could then conclude that the set of matched men is the same across stable matching, as if the set of men was finite, then the inclusions in Lemma 14 would imply that:

$$|M(\mu_W)| \leq |M(\mu)| \leq |M(\mu_M)|;$$

and

$$|W(\mu_M)| \leq |W(\mu)| \leq |W(\mu_W)|.$$

But trivially for any $\mu$ stable, $|W(\mu)| = |M(\mu)|$ – as each couple is made of one man and one woman – and then:

$$M(\mu) = M(\mu')$$

for all $\mu$ and $\mu'$ stable. The same is not true in the infinite case, because having two sets $A \subseteq B$ with $|A| = |B|$ does not imply that $A = B$ when $A$ is countably infinite.[57]

To relate this result to the literature in large markets, it is useful to observe that in Azevedo and Leshno (2016) the Rural Hospital theorem was obtained by imposing an equilibrium condition of supply meeting demand (in measure). The finiteness of the measure of the continuum of agents, in their model, plays the same role of the finiteness of the set of matched men and women. Hence in their setting they obtain that the Rural Hospital Theorem holds - differently from what found in our setting.

[57]As a trivial example, if $A$ is the set of even numbers and $B$ the set of natural numbers then $A \subsetneq B$, but they have the same counting measure, $|A| = |B|$.

In our setting, the measure of matched agents is still constant across stable matching, if we measure the number of matched agents with the "counting" measure, which is natural in our setting. In other words, if in a matching infinitely many agents are matched, then in all matching the set of matched agents is infinite (because of Lemma 14). Nonetheless, the Rural Hospital theorem can fail, because infinite sets can have the same cardinality as proper subsets, so even though there are always infinitely many matched agents, in some matching the set of matched men (or women) can be strictly larger.

We can characterize the instances in which the Rural Hospital theorem fails:

**Proposition 11**. *The Rural Hospital Theorem holds if and only if:*[58]

$$M = \mu_W \circ \mu_M(M).$$

*Remark* 10. Observe that the same statement could have been given for women - that is to say:

$$M = \mu_W \circ \mu_M(M) \Leftrightarrow W = \mu_M \circ \mu_W(W).$$

To see this, if by contradictions $M \supset \mu_W \circ \mu_M(M)$, pick $m \in M \setminus \mu_W \circ \mu_M(M)$, it is easy to check that $\mu_M(m) \in W$, and $\mu_M(m) \in W \setminus \mu_M \circ \mu_W(W)$.

The proposition provides necessary and sufficient conditions for the failure of the RH theorem using the fact that the two matching $\mu_M$ and $\mu_W$ are "extremal points" in the set of stable matching. In other words, say there exist a man $m$ such that $\mu(m) \in W$ and $\mu'(m) = m$ for some $\mu, \mu'$ stable matching. But then it must be that $\mu_M(m) \in W$ and

[58]We write $f \circ g$ to mean the composition map: $x \mapsto f(g(x))$, and as always for any set $A$, $f(A) = \{y \mid y = f(x), \exists x \in A\}$.

$\mu_W(m) = m$, and hence it is possible to find a characterization based solely on $\mu_M$ and $\mu_W$.

*Remark* 11. The characterization we found is not given in terms of the exogenous variables of the model, i.e. the preferences, but in terms of $\mu_M$ and $\mu_W$ which are endogenous. The link between the preferences and the matching $\mu_M$ and $\mu_W$ is given by the generalized DA algorithm and using that jointly with Proposition 11 it is possible to obtain a characterization of the failure of the Rural Hospital theorem in terms of the exogenous variables. A clearer connection between preferences and the failure of the theorem is hindered by the fact that different sets of preferences yield the same set of stable matching (and in particular the same $\mu_M$ and $\mu_W$). Then the characterizations of the result will depend, implicitly or explicitly, on agents' preferences through the matching $\mu_M$ and $\mu_W$.

As a corollary, we can provide some sufficient conditions for the Rural Hospital theorem to hold:

**Corollary 7.** *The Rural Hospital holds if one of the following conditions holds:*

- $|M(\mu)| < +\infty$ *for some* $\mu$ *stable;*
- *there exists a partition of* $M = \bigcup_j M_j$ *and* $W = \bigcup_j W_j$*, such that for all* $j$ $|M_j|, |W_j| < +\infty$*, and for all* $m \in M_j$*,* $m$ *finds acceptable only the women in* $W_j$*, mathematically:*

$$m >_m w, \quad \forall m \in M_j, w \in W \setminus W_j.$$

Notice how both of the conditions are sufficient but not necessary. To see this, consider the preferences obtained by changing those of Example 3. Let $M$ and $W$ be indexed on $\mathbb{Z}$

- that is a bi-infinite sequence.[59] Then let $P(m_i) : w_i, w_{i-1}$ and $P(w_i) : m_{i+1}, m_i$, for all $i \in \mathbb{Z}$.

It is easy to see how in this example there are two stable matching ($\mu_M$ matches $m_i$ with $w_i$ and $\mu_W$ matches $m_i$ with $w_{i-1}$), and the Rural Hospital Theorem is met as $M(\mu_W) = M$ and $W(\mu_M) = W$. Nonetheless neither of the conditions of Corollary 7 are met.

## Maximal stable matching

Whenever the Rural Hospital Theorem fails it is convenient to select a stable matching that maximizes the set of matched agents (in the sense of set inclusion). Formally we can define:

**Definition 17** (Maximal Stable Matching). A matching $\mu$ is maximal if for all $\mu'$ stable:

$$M(\mu) \cup W(\mu) \supseteq M(\mu') \cup W(\mu').$$

The following examples show two negative results: there might exist no stable matching that maximizes the set of matched agents (Example 4); and if there exists it might be different from the optimal matching $\mu_M$ and $\mu_W$ (Example 5). In Theorem 11 we will give a further description of the set of maximal matching, whenever it is not empty.

[59]Explicitly, let

$$M = \{\dots, m_{-i}, \dots, m_{-1}, m_0, m_1, \dots, m_i, \dots\}$$

and the same for $W$.

**Example 4** (Non-Existence of maximal matching). *Consider the following preferences:*

$$P(m_1) : w_3 \qquad P(w_1) : m_3$$

$$P(m_2) : w_4 \qquad P(w_2) : m_4$$

$$P(m_3) : w_5, w_1 \qquad P(w_3) : m_5, m_1$$

$$\dots\dots \qquad \dots\dots$$

$$P(m_i) : w_{i+2}, w_{i-2} \qquad P(w_i) : m_{i+2}, m_{i-2}$$

$$\dots\dots \qquad \dots\dots$$

*It is easy to see that there are only two stable matching, $\mu_M$ and $\mu_W$, and:*

$$\{m_3, \dots, m_i, \dots\} = M(\mu_W) \subsetneq M(\mu_M) = \{m_1, \dots, m_i, \dots\},$$

*and identically for women.*

**Example 5** (Existence of $\mu$ maximal, $\mu \neq \mu_M, \mu_W$). *Consider the following preferences:*

$$P(m_1) : w_3, w_1 \qquad P(w_1) : m_3, m_1$$

$$P(m_2) : w_4, w_2 \qquad P(w_2) : m_4, m_2$$

$$P(m_3) : w_5, w_3, w_1 \qquad P(w_3) : m_5, m_3, m_1$$

$$\ldots\ldots \qquad\qquad \ldots\ldots$$

$$P(m_i) : w_{i+2}, w_i, w_{i-2} \qquad\qquad P(w_i) : m_{i+2}, m_i, m_{i-2}$$

$$\ldots\ldots \qquad\qquad \ldots\ldots$$

*Similarly to the previous case, we have that:*

$$\{m_3, \ldots, m_i, \ldots\} = M(\mu_W) \subsetneq M(\mu_M) = \{m_1, \ldots, m_i, \ldots\},$$

*and identically for women.*

*In this case, though, there is another stable matching $\mu$, defined by $\mu(m_i) = w_i$. Clearly, it is maximal in the sense of set inclusion of matched agents, as the set of matched agents is $M \cup W$.*

To analyze the set of maximal matchings, let us define two subsidiary sets in terms of which we can provide a description of the set of maximal matchings.

**Definition 18** (Man maximal and woman maximal matchings). A stable matching $\mu$ is said to be man maximal if:

$$M(\mu) \supseteq M(\mu'),$$

for all $\mu'$ stable. The definition for women is analogous.

The following theorem describes the set of man and woman maximal matchings and how they relate to the set of maximal matching.

**Theorem 11.** *The set of man maximal matching is a non-empty sublattice of the set of stable matching (under the meet and join operations defined in 3.3 and 3.4, and the man-preferred or woman-preferred order) .*

This theorem then yields the following corollary

**Corollary 8.** *The set of maximal matchings is a (possibly empty) sublattice of the set of stable matching.*

## 3.5 Application to Dynamic Matching Markets

An application which requires a model with countably infinite agents is given by dynamic matching markets. Even though any real market is finite, agents often have preferences for matches in different periods. When we then model the matching market of this year, we have to consider the preferences of this year's agents to possibly include agents entering next year. But then a complete description of the market would include next year's agents, who in turn might find acceptable agents entering the market two years from now. Proceeding in this fashion we get that the *smallest model* describing the preferences for this matching market must include infinitely many agents, from this year's till indefinitely in the future.

More formally, let $M^t$ and $W^t$ be the set of men and women entering the market at time $t$. Agents in $M^t$ or $W^t$ can find agents who entered before or who are going to enter in the future acceptable. We assume that the preferences of each agent are exogenously given and summarized in the preference vector **P**. The object of interest is a matching $\mu$ that is stable in the market given by $(\bigcup_{t\in\mathbb{N}} M^t, \bigcup_{t\in\mathbb{N}} W^t, \mathbf{P})$. It is worth underlining that differently from other approaches in dynamic matching markets, the matching does not change with time. This assumption capture matching markets in which re-matching is either legally impossible or very costly, such as matching of kindergarteners to primary schools.

Depending on the application, only certain preferences **P** might be considered. To illustrate this with an example, consider the matching high-school graduates to colleges.[60] Interpret period $t$ as a given graduation year, $M^t$ as the high-school students graduating in year $t$, and $W^t$ as the college positions rendered available at time $t$. This application induces a structure on the set of preferences that each agents can hold. Students might find it acceptable to be admitted by colleges in further years, i.e. for $t' \geq t$; because they can either decide to apply the year they graduate, or postpone to applications to a year in the future. Conversely, a college which admits student in a given year $t$, will be able to admit only students who already hold a high-school diploma, and to rephrase this in terms of preferences, a college in year $t$ will find acceptable a subset of the students who graduated in $t' \leq t$.

To formalize the idea of the example just described, and in general any application to dynamic matching markets, we define *preferences for the future/past.*

**Definition 19.** We say that a man $m \in M^t$ has *preferences for the future*, if he finds acceptable (a subset of the) women entering the market in $t' \geq t$. Mathematically:

$$m >_m w, \quad \forall w \in W^s, s < t,$$

[60]We abstract from the fact that this would be a many-to-one matching problem, to focus on the dynamic nature of the model.

i.e. he finds *unacceptable* any woman who entered the market in the past.

On the other hand, he has *preferences for the past* if:

$$m >_m w \quad \forall w \in W^s, s > t,$$

and we use the same definitions for women.

When all agents on one side of the market have preferences that satisfy the properties just defined, we can give a description of the set of stable matching by using the theory developed in the previous section.

**Theorem 12.** *If all men and women have preferences for the future (or the past) then the dynamic matching market is isomorphic to a sequence of "disjoint" finite matching markets in which all agents find only women in their year acceptable.*[61]

*If all men have preferences for the past, and women have preferences for the future, then the market is in general not isomorphic to a sequence of finite markets.*

We can rephrase the statement of the theorem as follows. If the dynamic market one is considering induces preferences for men and women that have the same structure, (say, both for the future) then analyzing the dynamic market is equivalent to analyzing a sequence of finite markets. This isomorphism can be shown explicitly by modifying the preferences of all agents: let $\mathbf{P}$ be the original preferences, and $\mathbf{P}'$ be the preferences where all agents find unacceptable any agent entering the market in any other period

[61]In this context isomorphic refers to the set of stable matchings. Formally, two markets $(M, W, \mathbf{P}_1)$ and $(M, W, \mathbf{P}_2)$ with different preferences are isomorphic if the set of stable matchings is the same.

(whereas the ranking of agents in the same period are not changed). Then $(M, W, \mathbf{P})$ and $(M, W, \mathbf{P}')$ have the same set of stable matching.

On the other hand, if the structure of preferences for men is opposite to that of women, then the set of stable matching in the dynamic market is (in general) different from the set of stable matching obtained by considering each period's market separately. In this second case, then, the analysis carried out in the previous sections gives a description of the set of stable matching in the infinite market.

## 3.6 Conclusions

We studied one-to-one two sided matching with countably infinite agents, showing that the set of stable matching constitutes a non-empty lattice, if agents' preferences satisfy our assumption on the maximum. We showed that the Rural Hospital Theorem might fail in our setting and we described the implications that this has for dynamic matching markets.

As a direction for future research, it would be interesting to find conditions on the preferences of agents that are necessary and sufficient for the Rural Hospital Theorem does hold. The characterization we provide in 11 is given in terms of the maximal matchings $\mu_M$ and $\mu_W$, thus effectively *endogenous variables.* Also, we could describe the set of maximal stable matchings, but it would be interesting to provide a mechanism that maximizes the set of matched agent in any matching market, whenever such maximizer exists.

As pointed out in the introduction, the failure of the Rural Hospital Theorem implies that the manipulability result observed in the finite result may fail to hold in our case. Precisely, we show that in a market with multiple stable matchings, all agents might have no incentives to misreport their preferences. We provide a simple example in which this happens, but it would be interesting for future developments to understand if this happens all the time the Rural Hospital Theorem does not hold.

Finally, it would be interesting to study dynamic matching markets with strategic agents. The inefficiency due to the failure of the Rural Hospital Theorem might be worsened by agents' strategically misreporting their preferences to secure a matching in their period, instead of waiting. Furthermore, if one introduced uncertainty about the preferences of agents entering the market in the future, then agents' attitude toward risk would also need to be modeled. This extensions of the model studied in this paper would provide further insights on the strategic issues of matching markets opening repeatedly.

# *Bibliography*


Abdulkadiroglu, A., P. A. Pathak, and A. E. Roth
2005. The New York City High School Match. *American Economic Review*, 95(2):364–367.

Abdulkadiroglu, A. and T. Sonmez
2005. House Allocation with Existing Tenants. *Journal of Economic Theory*, 2005(December):1–62.

Aczel, J. and Z. Daroczy
1975. *On measures of information and their characterizations.* Academic Press.

Adachi, H.
2000. On a characterization of stable matchings. *Economics Letters*, 68:43–49.

Alkan, A.
1988. Nonexistence of stable threesome matchings. *Mathematical Social Sciences*, 16(2):207–209.

Alonso, R. and O. Camara
2016. Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory*, 165:672–706.

Athey, S. and J. Levin
2017. The Value of Information in Monotone Decision Problems. *Research in Economics.*

Aumann, R. J.
1976. Agreeing to Disagree. *The Annals of Statistics*, 4(6):1236–1239.

Azevedo, E. M. and J. D. Leshno
2016. A Supply and Demand Framework for Two-Sided Matching Markets. *Journal of Political Economy*, 124(5):1235–1268.

Baliga, S., E. Hanany, and P. Klibanoff
2013. Polarization and Ambiguity. *American Economic Review*, 103(7):3071–3083.

Bhattacharyya, A.
1946. On a Measure of Divergence between Two Multinomial Populations. *The Indian Journal of Statistics*, 7(4):401–406.

Blackwell, D.
1951. Comparison of experiments. *Proceedings of the second Berkeley symposium on on Mathematical Statistics and Probability*, Pp. 93–102.

Blackwell, D.
1953. Equivalent Comparison of Experiment. *Annals of Mathematical Statistics.*

Blackwell, D. H. and M. A. Girshick
1954. Theory of games and statistical decisions.

Campbell, L. L.
1972. Characterization of entropy of probability distributions on the real line. *Information and Control*, 21(4):329–338.

Caplin, A. and M. Dean
2015. Revealed Preference, Rational Inattention, and Costly Information Acquisition. *American Economic Review*, 105(7):2183–2203.

Caplin, A., M. Dean, D. Martin, D. Bernheim, I. Brocas, J. Carrillo, V. Crawford, S. Dellavigna, D. Laibson, A. Lizzeri, U. Malmendier, R. Nagel, E. Ok, A. Rangel, A. Rubinstein, Y. Salant, and A. Schotter
2011. Search and Satisficing. *American Economic Review*, 101:2899–2922.

Chambers, C. P. and P. J. Healy
2010. Updating toward the signal. *Economic Theory*, 50(3):765–786.

Che, Y. and N. Kartik
2009. Opinions as Incentives. *Journal of Political Economy*, 117(5):815–860.

Che, Y.-K., J. Kim, and F. Kojima
2015. Stable Matching in Large Economies *.

Chernoff, H.
1952. A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations. *The Annals of Mathematical Statistics*, 23(4):493–507.

Cookson, J. A. and M. Niessner
2016. Why Don't We Agree ? Evidence from a Social Network of Investors. *Working Paper.*

Cover, T. M. and J. A. Thomas
2006. *Elements of information theory.* Wiley-Interscience.

Cragg, J. G. and B. G. Malkiel
1982. *Expectations and the structure of share prices.* University of Chicago Press.

Csiszár, I.
2008. Axiomatic Characterizations of Information Measures. *Entropy*, 10:261–273.

Denti, T., M. Mihm, H. de Oliveira, and K. Ozbek
2016. Rationally inattentive preferences and hidden information costs. *Theoretical Economics.*

Dewan, A. and N. Neligh
2017. Estimating Information Cost Functions in Models of Rational Inattention.

Dixit, A. K. and J. W. Weibull
2007. Political polarization. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18):7351–6.

Eliaz, K. and R. Spiegler
2007. A mechanism-design approach to speculative trade. *Econometrica*, 75(3):875–884.

Eliaz, K. and R. Spiegler
2016. Search design and broad matching. *American Economic Review*, 106(3):563–586.

Ely, J., A. Frankel, and E. Kamenica
2015. Suspense and Surprise Alexander Frankel Emir Kamenica. *Journal of Political Economy*, 123(1):215–260.

Francetich, A. and D. Kreps
2014. Bayesian inference does not lead you astray… on average. *Economics Letters*, 125(3):444–446.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg
2006. Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4):1043–1068.

Gale, D. and L. Shapley
1962. College Admissions and the Stabilty of Marriage. *The American Mathematical Monthly*, 69(1):9–15.

Ganuza, J. and J. Penalva
2010. Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions. *Econometrica*, 78(3):1007–1030.

Gentzkow, M. and E. Kamenica
2014. Costly persuasion. In *American Economic Review*, volume 104, Pp. 457–462.

Gusfield, D., R. Irving, P. Leather, and M. Saks
1987. Every finite distributive lattice is a set of stable matchings for a small stable marriage instance. *Journal of Combinatorial Theory, Series A*, 44(2):304–309.

Hansen, L. P. and K. J. Singleton
1983. Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns. *Journal of Political Economy*, 91(2):249–265.

Hansen, S., M. McMahon, and A. Prat
2014. Transparency and Deliberation within the FOMC: A Computational Linguistics Approach. *CEP Discussion Papers*.

Harris, M. and A. Raviv
1993. Differences of Opinion Make a Horse Race. *Review of Financial Studies*, 6(3):473–506.

Harrison, J. M. and D. M. Kreps
1978. Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations. *The Quarterly Journal of Economics*, 92(2):323–336.

Hatfield, J. W. and P. R. Milgrom
2005. Matching with contracts.

Hébert, B. and M. Woodford
2016. Rational Inattention with Sequential Information Sampling.

Hong, H. and J. C. Stein
2007. Disagreement and the stock market. *Journal of Economic Perspectives*, 21(2):109–128.

Immorlica, N. and M. Mahdian
2005. Marriage, honesty, and stability. In *In Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA*, Pp. 53–62.

Jagadeesan, R.
2016. Lone wolves in infinite, discrete matching markets.

Jeffreys, H.
1946. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A: Mathematical and physical sciences*, 186(1007):453–461.

Kadam, S. V. and M. H. Kotowski
2016. Multi-period Matching.

Kailath, T.
1967. The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Transactions on Communications*, 15(1):52–60.

Kamenica, E. and M. Gentzkow
2011. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kandel, E. and N. D. Pearson
1995. Differential Interpretation of Public Signals and Trade in Speculative Markets. *Journal of Political Economy*, 103(4):831.

Kannappan, P. and P. Rathie
1972. A directed-divergence function of type $\beta$. *Information and Control*, 20(1):38–45.

Kannappan, P. and P. N. Rathie
1988. An Axiomatic Characterization of J-Divergence. In *Transactions of the Tenth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Pp. 29–36. Dordrecht: Springer Netherlands.

Kannappan, P. L. and P. N. Rathie
1973. On a characterization of directed divergence. *Information and Control*, 22(2):163–171.

Kartik, N., X. F. Lee, and W. Suen
2015. Information Validates the Prior and Applications to Signaling Games. *Working Paper, School of Economics and Finance, University of Hong Kong.*

Kim, O. and R. E. Verrecchia
1991. Market reaction to anticipated announcements. *Journal of Financial Economics*, 30(2):273–309.

Kim, O. and R. E. Verrecchia
1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics*, 17(1-2):41–67.

Kojima, F. and P. A. Pathak
2009. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–627.

Kullback, S. and M. Khairat
1966. A note on minimum discrimination information. *Annals of Mathematical Statistics*, 37(1):279–280.

Kullback, S. and R. A. Leibler
1951. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Kurino, M.
2008. Credibility , Efficiency and Stability : A Theory of Dynamic Matching Markets. (February).

Kurino, M.
2014. House allocation with overlapping generations. *American Economic Journal: Microeconomics*, 6(1 D):258–289.

Liese, F. and I. Vajda
1987. *Convex Statistical Distances.* Teubner.

Mankiw, N. G., R. Reis, and J. Wolfers
2003. Disagreement about Inflation Expectations. *NBER Macroeconomics Annual*, 18(2003):209–248.

Monte, D. and N. Tumennasan
2015. Centralized allocation in multiple markets. *Journal of Mathematical Economics*, 61:74–85.

Morris, S.
1994. Trade with Heterogeneous Prior Beliefs and Asymmetric Information. *Econometrica*, 62(6):1327–1347.

Neilson, W. S. and H. Winter
2002. A verification of the expected utility calibration theorem. *Economics Letters*, 74(3):347–351.

Ostrovsky, M.
2008. Stability in supply chain networks. *American Economic Review*, 98(3):897–923.

Peranson, E. and A. E. Roth
1999. The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, 89(4):748–780.

Persico, N.
2000. Information Acquisition in Auctions. *Econometrica*, 68(1):135–148.

Piketty, T.
1995. Social Mobility and Redistributive Politics. *The Quarterly Journal of Economics*, 110(3):551–584.

Polemarchakis, H.
2016. Rational Dialogs. *The Warwick Economics Research Paper Series (TWERPS) 1115, University of Warwick, Department of Economics.*

Rényi, A.
1961. On Measures of Entropy and Information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1.

Roth, A.
1984. The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *The Journal of Political Economy*, 92(6):991–1016.

Roth, A. and M. Sotomayor
1992. *Two-sided matching: A study in game-theoretic modeling and analysis*, volume 4. Cambridge University Press.

Roth, A. E.
1986. On the allocation of residents to rural hospitals: a general property of two-sided matching markets. *Econometrica: Journal of the Econometric Society*, 54(2):425–427.

Roth, A. E.
2008. What have we learned from market design? In *Economic Journal*, volume 118, Pp. 285–310. Blackwell Publishing Ltd.

Roth, A. E., T. Sonmez, and M. U. Unver
2004. Kidney Exchange. *The Quarterly Journal of Economics*, 119(2):457–488.

Sebenius, J. K. and J. Geanakoplos
1983. Don't Bet on It: Contingent Agreements with Asymmetric Information. *Journal of the American Statistical Association*, 78(382):424–426.

Sethi, R. and M. Yildiz
. Communication with Unknown Perspectives. *Econometrica*.

Sethi, R. and M. Yildiz
2012. Public Disagreement. *American Economic Journal: Microeconomics*, 4(3):57–95.

Shaked, M. and J. G. Shanthinkumar
2006. *Stochastic orders*, Springer Series in Statistics. New York, NY: Springer New York.

Shapley, L. S. and M. Shubik
1971. The assignment game I: The core. *International Journal of Game Theory*, 1(1):111–130.

Shores, T. S.
2007. *Applied linear algebra and matrix analysis*. Springer.

Sims, C. A.
2003. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Sims, C. A.
2010. Rational Inattention and Monetary Economics. *Handbook of Monetary Economics*, 3:155–181.

Sunstein, C. R.
2002. The law of group polarization. *Debating Deliberative Democracy*, 10(2):80–101.

Taneja, I. J.
1987. Statistical aspects of divergence measures. *Journal of Statistical Planning and Inference*, 16(C):137–145.

Tarski, A.
1955. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285–309.

Van Dan Steen, E.
2010. Disagreement and the Allocation of Control. *Journal of Law, Economics, and Organization*, 26(2):386–426.

Varian, H. R.
1985. Divergence of Opinion in Complete Markets: A Note. *The Journal of Finance*, 40(1):309–317.

Varian, H. R.
1989. Differences of Opinion in Financial Markets. In *Financial Risk: Theory, Evidence and Implications: Proceedings of the Eleventh Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis*, Pp. 3–37. Dordrecht: Springer Netherlands.

Woodford, M.
2012. Inattentive Valuation and Reference-Dependent. *Unpublished Manuscript, Columbia University*, P. 89.

Yildiz, M.
2004. Waiting to Persuade. *Quarterly Journal of Economics*, 119(1):223–248.

Chapter A

# *Appendix of Chapter 1*

## A.1 Proofs of Section 1.2

In this section we prove the main theorem, Theorem 1, and the results related to the analysis of the disagreement functions.

### Proof of the main theorem

The proof proceeds by steps, that we state and prove separately for clarity. We summarize the steps here, pointing out at which point each Axiom we assumed plays a role. Axioms 1 and 2 are used at various points. The starting point of the Theorem is Axiom 6.

- Result 1 shows that the differential of $D(p, q)$ at $p \neq q$ is never identically $0$, and it uses Axiom 5 and Axiom 1;
- Result 2 proves that $\partial_i D(p, q)$ can be written as the ratio of a function, $h$, that depends only on the beliefs in states $i$ and $i + 1$ and a function, $\alpha$, that depends on the whole beliefs $p, q$, and is independent of the state $i$:

$$\partial_i D(p, q) = \frac{h(p_i, p_{i+1}, q_i, q_{i+1})}{\alpha(p, q)}.$$

- Result 3 uses the properties of the derivatives to show that $h_p^n(p_i, p_{i+1}, q_i, q_{i+1})$ can be written as a difference:

$$h(p_1, p_2, q_1, q_2) = i(p_1, q_1) - i(p_2, q_2),$$

  for some function $i : [0,1]^2 \to \mathbb{R}$.

- Result 4 then builds on Corollary 3 to show that $i(x, y)$ is a function of the ratio:

$$i(x, y) = G(x/y),$$

  for some $G : (0, +\infty) \to \mathbb{R}$ Since we use Corollary 3, this is the step of the proof where we use Axioms 3 and 4, together with Lemma 4, which was a consequence of those axioms and Axiom 5.

- Results 5 and 6 then apply Axiom 5 to show that the derivative of $G$ must be a homogeneous function:

$$G'(x) = cx^{\alpha} \quad \exists \alpha, c \in \mathbb{R}.$$

  or $G'(x) = \frac{1}{x}\left(a + \frac{b}{x}\right)$. Result 6 is the main part of the proof.

- Result 7 then shows what the possible primitives of $G'(x)$ are, and Result 8 uses this to pin down the functional form of $\alpha(p, q)$.

- Finally we use the conclusions of Results 7 and 8 to wrap up the proof, after restating Theorem 1.

**Result 1.** *For all $p \neq q$ we have that there exists $j = 1, \ldots, n-1$ such that $\partial_j D(p, q) \neq 0$.*

*Proof of Result 1.* Let us prove that if $n = 2$, then for all $p \neq q$ we have that $\partial_1 D(p, q) \neq 0$. Suppose toward contradiction that instead $\partial_1 D(p, q) = 0$ for some $p, q$.

Now, for any $q' \in [p, q]$ we can define a function $\gamma(q, q') : \Delta_2 \times \Delta_2 \to \mathbb{R}$ such that:

$$D(p, q) = D(p, q') + D(p, p + \epsilon(q, q')z),$$

where $z = (1, -1)$. Therefore $\partial_1 D(p, q) = 0$ implies that:

$$\partial_1 D(p, p + \epsilon(q, q')z) = \partial_1 D(p, p + \epsilon(q, q')z) \frac{\partial \epsilon(q, q')}{\partial q} = 0, \tag{A.1}$$

and this holds for all $q' \in [p, q]$. Notice that as $q'$ approaches $q$ we have that $\epsilon(q, q') \to 0$, and we for a dense subset of $y \in (0, \delta) \subset \mathbb{R}$, $\partial_1 D(p, p + yz) \neq 0$.[62] This implies that $\frac{\partial \epsilon(q,q')}{\partial q} = 0$ for all $q'$ in a neighborhood of $q$. Therefore for $q'$ in a neighborhood of $q$ we have that $\epsilon(q, q') = h(q')$ and since at $q = q'$ we have that $\epsilon(q, q') = 0$, then $h(q') = 0$. Therefore, we have that for $q' \in [p, q]$, and in a neighborhood of $q$, $D(p, q) = D(p, q')$.

In order to complete the proof, we show that $D(p, q)$ must be constant on the segment $(p, q]$, which implies, by continuity, that $0 = D(p, p) = D(p, q)$. Since this contradicts Axiom 1, proving that $D(p, q)$ is constant on $(p, q]$ suffices to conclude the proof.

Suppose not, i.e. suppose there exists $\bar{q} \in (p, q)$ such that $D(p, q') < D(p, \bar{q}) = D(p, q)$ for all $q'' \in [p, \bar{q})$. Plainly, $\partial_1 D(p, \bar{q}) = 0$, because $D(p, \cdot)$ is constant on $[\bar{q}, q]$, and since it is differentiable its derivative must agree with the right derivative. Since $p \neq \bar{q}$

[62] If this was not the case, by continuity we have that $\partial_1 D(p, p + yz) = 0$ for all $y \in (0, \delta)$. But then $D(p, p + yz) = D(p, p) = 0$, contradicting Axiom 1.

and $\partial_1 D(p,\bar{q}) = 0$, we can reiterate the argument used earlier for $q$ to find for $q'' \in [p,\bar{q}]$ in a neighborhood of $\bar{q}$ we must have that $D(p,q'') = D(p,\bar{q})$, which contradicts the definition of $\bar{q}$. This concludes the proof for $n = 2$.

For general $n$, the proof is similar. If $\partial_j D(p,q) = 0$ for all $j$, then applying equation (A.1) for all $z_{i,j} = (0,\ldots,1,0,\ldots,-1,0,\ldots,0)$ we obtain that $D(p,q)$ must be locally constant for $q'$ in a neighborhood of $q$. Consider then the intersection of this neighborhood with the segment $[p,q]$. Denote by $\bar{q}$ the belief in $[p,q]$ such that $D(p,\bar{q}) = D(p,q)$ and $D(p,q') < D(p,\bar{q})$ such that $q' \in [p,\bar{q}]$. Take the derivative of $D$ in the direction of the segment $[p,q]$. Since $D(p,\cdot)$ is differentiable and it is constant for $q' \in [\bar{q},q]$, then we must have that $\partial_{q-p} D(p,\bar{q}) = 0$.[63] But then we can employ the argument used for $n = 2$ on the segment $[p,\bar{q}]$ obtaining the contradiction that $D(p,\cdot)$ must be constant in a neighborhood of $\bar{q}$. $\square$

**Result 2.** *If $D_n(p,q)$ satisfies Axiom 6, then there exists a function $h : [0,1]^4 \to \mathbb{R}$,[64] and a function $\alpha(p,q) : \Delta_n \times \Delta_n \to \mathbb{R}$ such that:*

$$\partial_i D(p,q) = \frac{h(p_i, p_{i+1}, q_i, q_{i+1})}{\alpha(p,q)}, \tag{A.2}$$

*for all $p,q \in \Delta_n$.*

---

[63] We denote by $\partial_{q-p} D(p,\bar{q})$ the derivative in the direction $q-p$, formally:

$$\partial_{q-p} D(p,\bar{q}) = \lim_{\epsilon \to 0} \frac{D(p,\bar{q}+\epsilon(q-p)) - D(p,\bar{q})}{\epsilon}.$$

[64] In practice the function $h$ will be used only for vectors $x = (x_1,x_2,x_3,x_4)$ such that $x_1 + x_2 \leq 1$ and $x_3, x_4 \leq 1$. We did not clarify this in the domain of $h$ to simplify the notation.

*Furthermore, if* $(\alpha, h)$ *and* $(\tilde{\alpha}, \tilde{h})$ *both satisfy equation* (A.2) *then:*

$$\alpha(p,q) = k\tilde{\alpha}(p,q) \quad \textit{and} \quad h(p_i, p_{i+1}, q_i, q_{i+1}) = k\tilde{h}(p_i, p_{i+1}, q_i, q_{i+1}),$$

*for some* $k \neq 0$.

*Proof of Result 2.* Take any $p \neq q \in \Delta_n^\circ$. By Result 1 there exists a $j$ such that $\partial_j D(p,q) \neq 0$. Apply then Axiom 6, and define:

$$h(p_i, p_{i+1}, q_i, q_{i+1}) := g(p_i, p_{i+1}, p_j, p_{j+1}, q_i, q_{i+1}, q_j, q_{j+1}),$$

where $g$ the function defined in Axiom 6. Notice that in the definition of $h$, the values of $p_j, p_{j+1}, q_j, q_{j+1}$ are fixed. Also, notice that $h(p_i, p_{i+1}, q_i, q_{i+1}) = 0$ if and only if $\partial_i D_n(p,q) = 0$ (by Axiom 6) so for all $i$ such that $\partial_i D(p,q)$ we have that:

$$\frac{\partial_i D(p,q)}{h(p_i, p_{i+1}, q_i, q_{i+1})} \quad \text{is independent of} \quad i,$$

because:

$$\frac{\partial_i D_n(p,q)}{h(p_i, p_{i+1}, q_i, q_{i+1})} = \partial_j D_n(p,p) = \frac{\partial_k D_n(p,q)}{h(p_k, p_{k+1}, q_k, q_{k+1})}$$

as it again follows from Axiom 6. Define $\alpha(p,q)$ to be such ratio:

$$\alpha(p,q) := \frac{h(p_i, p_{i+1}, q_i, q_{i+1})}{\partial_i D(p,q)},$$

we only need to show that if $(\alpha, h_p^n)$ and $(\tilde{\alpha}, \tilde{h}_p^n)$ both satisfy equation (A.2) then they are multiples by a constant *independent of* $p, q$.

To see this, denote for brevity $h(i) := h(p_i, p_{i+1}, q_i, q_{i+1})$, $\partial_i D := \partial_i D(p, q)$, and $\alpha := \alpha(p, q)$. Now suppose that $(\alpha, h)$ and $(\tilde{\alpha}, \tilde{h})$ satisfy (A.2). This means that:

$$\frac{h(j)}{\partial_j D} = \alpha \text{ and } \tilde{\alpha} = \frac{\tilde{h}(i)}{\partial_i D},$$

but then using that $\frac{\partial_j D}{\partial_i D} = \frac{h(j)}{h(i)}$, we obtain:

$$\frac{\tilde{\alpha}}{\alpha} = \frac{\tilde{h}(i)}{h(i)}.$$

Since the right hand side depends only on $p_i, q_i, p_{i+1}, q_{i+1}$, and the left hand side depends on the whole set of beliefs $p, q$ (in general), we have that:

$$\frac{\tilde{\alpha}}{\alpha} = k = \frac{\tilde{h}(i)}{h(i)},$$

for some $k$ independent of $p, q$. □

**Result 3.** *We have that for some function* $i : [0, 1]^2 \to \mathbb{R}$*:*

$$h(p_1, p_2, q_1, q_2) = i(p_1, q_1) - i(p_2, q_2),$$

*for all* $p_1, p_2, q_1, q_2 \in [0, 1]^4$.

*Proof of Result 3.* Denote by $\partial_{i,j} D(p, q)$ in the direction $z_{ij} = e_i - e_j$. Simple properties of the derivative imply that:

$$\partial_{i,i+1}D(p,q) + \partial_{i+1,i+2}D(p,q) = \partial_{i,i+2}D(p,q),$$

and rewriting this in terms of $h$ we find:

$$\frac{h(p_i, p_{i+1}, q_i, q_{i+1})}{\alpha(p,q)} + \frac{h(p_{i+1}, p_{i+2}, q_{i+1}, q_{i+2})}{\alpha(p,q)} = \frac{h(p_i, p_{i+2}, q_i, q_{i+2})}{\alpha(p,q)}.$$

Simplifying we obtain:

$$h(p_i, p_{i+1}, q_i, q_{i+1}) + h(p_{i+1}, p_{i+2}, q_{i+1}, q_{i+2}) = h(p_i, p_{i+2}, q_i, q_{i+2}). \tag{A.3}$$

Since we have that $h$ is differentiable (because $D$ was assumed to be three times differentiable), we can take the derivative with respect to $p_i$ in equation (A.3). So, defining $h_j$ to be the derivative with respect to the $j-$th variable,[65] we get that:

$$h_1(p_i, p_{i+1}, q_i, q_{i+1}) = h_1(p_i, p_{i+2}, q_i, q_{i+2}),$$

and since $p_{i+1}, q_{i+1}, p_{i+2}, q_{i+2}$ can take any value, we have that $h_1$ depends only on $p_i$ and $q_i$. Analogously, we find that $h_3$ depends only on $p_i$ and $q_i$. Hence $h$ can be written as the sum of two functions $i^{(1)}$ and $i^{(2)}$ such that:

$$h(p_1, p_{i+1}, q_i, q_{i+1}) = i^{(1)}(p_i, q_i) - i^{(2)}(p_{i+1}, q_{i+1}). \tag{A.4}$$

[65]Formally, $h_1(x, y, w, z) := \frac{\partial h}{\partial x}$, and similarly for $h_2, h_3, h_4$.

The proof would be complete if we proved that the functions $i^{(1)}$ and $i^{(2)}$ are the same function.

To see this, consider equation (A.3) with

$$x := p_i = p_{i+1} = p_{i+2} \quad \text{and} \quad y := q_i = q_{i+1} = q_{i+2},$$

these values yield the equality:

$$2h(x, x, y, y) = h(x, x, y, y) \Rightarrow h(x, x, y, y) = 0, \quad \forall x, y \in [0, 1]^2.$$

So rewriting this in terms of the functions $i^{(1)}$ and $i^{(2)}$ introduced in equation (A.4), we get:

$$i^{(1)}(x, y) - i^{(2)}(x, y) = 0, \quad \forall x, y \in [0, 1]^2,$$

which means that $i^{(1)}(x, y) = i^{(2)}(x, y)$ so we call it $i$ and find the thesis. $\square$

**Result 4.** *We have that the function $i(x, y)$ introduced in Result 3 is a function of the ratio of $x/y$, i.e.:*

$$i(x, y) = G(x/y),$$

*for a function $G : \mathbb{R}_+ \to \mathbb{R}$.*

*Proof of Result 4.* Pick $p, q \in \Delta_n^\circ$. As in Corollary 3 define $q^1 = (q_1 + q_2, 0, q_3, \ldots, q_n)$ and $q^2 = (0, q_1 + q_2, q_3, \ldots, q_n)$ and consider the segment joining these two beliefs:

$$[q^1, q^2] = \{r \in \Delta_n \mid r = \lambda q^1 + (1 - \lambda) q^2, \lambda \in [0, 1]\}.$$

It was proved in the same Corollary that the minimum $\min_{r\in[q^1,q^2]} D(p,r)$ is achieved at the $r$ for which $\frac{r_1}{r_2} = \frac{p_1}{p_2}$. Since $D$ is differentiable, at such point the derivative in the direction $1,2$ must be $0$, i.e.:

$$\partial_1 D(p,r) = \frac{h(p_1,p_2,r_1,r_2)}{\alpha(p,r)} = 0,$$

but then using Result 3 we have that:

$$\frac{i(p_1,r_1) - i(p_2,r_2)}{\alpha(p,q)} = 0, \quad \text{if } \frac{p_1}{p_2} = \frac{r_1}{r_2},$$

or equivalently:

$$i(p_1,r_1) = i(p_2,r_2) \quad \text{if} \quad \frac{p_1}{r_1} = \frac{p_2}{r_2},$$

so $i$ is a function of the ratio only:

$$i(p_1,r_1) = G\left(\frac{p_1}{r_1}\right).$$

$\square$

In the following part of the proof, we will apply the Independence Axiom, Axiom 5, to obtain the functional form of $G$.

Consider $p = (p_1,\ldots,p_n) \in \Delta_n^\circ$ and $q = (q_1,\ldots q_n) \in \Delta_n^\circ$ and take $\lambda := (\lambda, 1-\lambda) \in$

$\Delta_2^\circ$ and $\gamma = (\gamma, 1-\gamma) \in \Delta_2^\circ$.[66] We have that:

$$p * \lambda = (\lambda p_1, (1-\lambda)p_1, \ldots, \lambda p_n, (1-\lambda)p_n),$$

$$q * \gamma = (\gamma q_1, (1-\gamma)q_1, \ldots, \gamma q_n, (1-\gamma)q_n),$$

and by the independence Axiom, Axiom 5:

$$D(p * \lambda, q * \gamma) = D(p, q) + D(\lambda, \gamma),$$

so that perturbing $p$ with $\epsilon z_i$ we get the derivative on the right hand side is $\frac{\left(G\left(\frac{p_i}{q_i}\right) - G\left(\frac{p_{i+1}}{q_{i+1}}\right)\right)}{\alpha(p,q)}$. For brevity, call it:

$$RHS := \frac{\left(G\left(\frac{p_i}{q_i}\right) - G\left(\frac{p_{i+1}}{q_{i+1}}\right)\right)}{\alpha(p, q)}.$$

Similarly, the derivative of the left hand side will be defined as:

$$LHS := \frac{\lambda\left(G\left(\frac{\lambda p_i}{\gamma q_i}\right) - G\left(\frac{\lambda p_{i+1}}{\gamma q_{i+1}}\right)\right) + (1-\lambda)\left(G\left(\frac{(1-\lambda)p_i}{(1-\gamma)q_i}\right) - G\left(\frac{(1-\lambda)p_{i+1}}{(1-\gamma)q_{i+1}}\right)\right)}{\alpha(p * \lambda, q * \gamma)}$$

Since we must have that $LHS = RHS$, then:

$$\frac{\alpha(p * \lambda, q * \gamma)}{\alpha(p, q)} = \frac{\lambda\left(G\left(\frac{\lambda p_i}{\gamma q_i}\right) - G\left(\frac{\lambda p_{i+1}}{\gamma q_{i+1}}\right)\right) + (1-\lambda)\left(G\left(\frac{(1-\lambda)p_i}{(1-\gamma)q_i}\right) - G\left(\frac{(1-\lambda)p_{i+1}}{(1-\gamma)q_{i+1}}\right)\right)}{\left(G\left(\frac{p_i}{q_i}\right) - G\left(\frac{p_{i+1}}{q_{i+1}}\right)\right)} \tag{A.5}$$

[66] Excuse the abuse of notation here.

and for the sake of brevity define $r_i = \frac{p_i}{q_i}$, $R_1 = \frac{\lambda}{\gamma}$, and $R_2 = \frac{1-\lambda}{1-\gamma}$, so that we can rewrite:

$$\frac{\alpha(p*\lambda, q*\gamma)}{\alpha(p,q)} = \frac{\lambda\left(G\left(R_1 r_i\right) - G\left(R_1 r_{i+1}\right)\right) + (1-\lambda)\left(G\left(R_2 r_i\right) - G\left(R_2 r_{i+1}\right)\right)}{\left(G\left(r_i\right) - G\left(r_{i+1}\right)\right)}$$

and notice that $\frac{\alpha(p*\lambda, q*\gamma)}{\alpha(p,q)}$ does not depend on $r_i$ or $r_{i+1}$ and then we obtain that for all $r'_i, r'_{i+1} \in \mathbb{R}^+$:

$$\begin{aligned}\frac{\lambda\left(G\left(R_1 r_i\right) - G\left(R_1 r_{i+1}\right)\right) + (1-\lambda)\left(G\left(R_2 r_i\right) - G\left(R_2 r_{i+1}\right)\right)}{\left(G\left(r_i\right) - G\left(r_{i+1}\right)\right)} \\ = \frac{\lambda\left(G\left(R_1 r'_i\right) - G\left(R_1 r'_{i+1}\right)\right) + (1-\lambda)\left(G\left(R_2 r'_i\right) - G\left(R_2 r'_{i+1}\right)\right)}{\left(G\left(r'_i\right) - G\left(r'_{i+1}\right)\right)}\end{aligned}$$

Therefore, we found that there exists a constant $K(\lambda, \gamma)$ such that:

$$\frac{\lambda\left(G\left(R_1 x\right) - G\left(R_1 y\right)\right) + (1-\lambda)\left(G\left(R_2 x\right) - G\left(R_2 y\right)\right)}{\left(G\left(x\right) - G\left(y\right)\right)} = K(\lambda, \gamma), \tag{A.6}$$

$\forall x, y \in [0, +\infty)$, where $R_1 := \frac{\lambda}{\gamma}$ and $R_2 := \frac{1-\lambda}{1-\gamma}$.

Next, we prove that equation (A.6) implies that $G(x)$ cannot be bounded, unless it is a constant. I.e. if $G$ is not constant, then either $\lim_{x\to 0} G(x) = -\infty$ or $\lim_{x\to+\infty} G(x) = \infty$ (or both). Notice that $\lim_{x\to 0} G(x)$ and $\lim_{x\to+\infty} G(x)$ both exist because $G$ is weakly monotone. As a matter of fact $G$ is continuous and if for some $x, x'$ we have that $G(x) = G(x')$ then we have that $G(y) = G(x)$ for all $y \in [x, x']$. This follows by Corollary 3, as disagreement is monotone on segment $[q^1, q^*]$ and segment $[q^*, q^2]$ (see also Figure 1.3).

Therefore the limits of $G(x)$ at $0$ and $\infty$ exist, and furthermore:

$$\forall R > 0 \qquad \lim_{x \to +\infty} f(x) = \lim_{x \to +\infty} f(Rx), \tag{A.7}$$

as it is easy to check. We will invoke equation (A.7) repeatedly to prove the following result.

**Result 5.** *Either $G$ is constant or $G$ is unbounded, meaning:*

$$\sup_{x \in (0,+\infty)} |G(x)| = +\infty.$$

*Proof.* Suppose that $G$ is bounded, $c_1 < G(x) < c_2$, for two $c_1, c_2 \in \mathbb{R}$. Then, take the limit for $x \to +\infty$, and define $I := \lim_{x\to\infty} G(x)$ (notice it is finite because we assumed $G$ bounded). Using equation (A.7), we can rewrite (A.6) as:

$$\frac{\lambda \left(I - G\left(R_1 y\right)\right) + (1-\lambda)\left(I - G\left(R_2 y\right)\right)}{\left(I - G\left(y\right)\right)} = K(\lambda, \gamma).$$

Doing the same thing for the limit at $y \to 0$ (call $J := \lim_{y\to 0} G(y)$),

$$K(\lambda, \gamma) = \frac{\lambda\left(I - J\right) + (1-\lambda)\left(I - J\right)}{\left(I - J\right)} = 1.$$

But if $K(\lambda, \gamma) = 1$ for all $\lambda$ and $\gamma$ we get:

$$\lambda\left(G\left(R_1 x\right) - G\left(R_1 y\right)\right) + (1-\lambda)\left(G\left(R_2 x\right) - G\left(R_2 y\right)\right) = \left(G\left(x\right) - G\left(y\right)\right).$$

Now taking the limit for $\lambda \to 0$ we have that:

$$\lambda \left(G\left(R_1 x\right) - G\left(R_1 y\right)\right) \to 0,$$

because $\left(G\left(R_1 x\right) - G\left(R_1 y\right)\right)$ is bounded. Then $R_2 \to \frac{1}{1-\gamma}$, and since $G$ is continuous in $(0, +\infty)$ we get:

$$G\left(\frac{x}{1-\gamma}\right) - G\left(\frac{y}{1-\gamma}\right) = G(x) - G(y).$$

Then, considering again the limit at, say, $y \to +\infty$ and equation (A.7), we get that (for all $x > 0$ and $\gamma \in (0,1)$): $G\left(\frac{x}{1-\gamma}\right) = G(x)$, which implies that $G$ is constant. $\square$

Next, we will prove another important consequence of equation (A.6).

**Result 6**. *Either the derivative of $G$ is homogeneous, i.e.:*

$$G'(x) = cx^{\alpha},$$

*for some $\alpha \in \mathbb{R}$, $c \in \mathbb{R}$. Or:*

$$G'(x) = \frac{a}{x^2} + \frac{b}{x},$$

*for some $a, b \in \mathbb{R}$.*

*Proof.* Clearly if $G$ is constant, we have that $G_p' = 0$, and so the thesis of the Theorem holds true. Suppose that $G$ is not constant. Then, by Result 5, we have that it is unbounded.

Clearly, since $G$ is continuous on $(0, +\infty)$, if it is unbounded then either the limit for $x \to +\infty$ or $x \to 0$ (or both) have to be infinite. Suppose that $\lim_{x\to 0} G(x) = \infty$ (the case in which $\lim_{x\to 0} G(x) \in (-\infty, \infty)$ and $\lim_{x\to +\infty} G(x) = \pm\infty$ is analogous).

Recall that by (A.6):

$$\frac{\lambda\left(G\left(R_{1}x\right)-G\left(R_{1}y\right)\right)+(1-\lambda)\left(G\left(R_{2}x\right)-G\left(R_{2}y\right)\right)}{\left(G\left(x\right)-G\left(y\right)\right)}=K(\lambda,\gamma),$$

so the limit for $x\to 0$ must be the same value:

$$K(\lambda,\gamma)=\lim_{x\to 0}\frac{\lambda\left(G\left(R_{1}x\right)-G\left(R_{1}y\right)\right)+(1-\lambda)\left(G\left(R_{2}x\right)-G\left(R_{2}y\right)\right)}{\left(G\left(x\right)-G\left(y\right)\right)},$$

so, in particular, it must exist. Now, since $G(x)\to\infty$ as $x\to 0$, and $y$,$R_1$ and $R_2$ are fixed, then we have that:

$$\begin{aligned}\lim_{x\to 0}&\frac{\lambda\left(G\left(R_{1}x\right)-G\left(R_{1}y\right)\right)+(1-\lambda)\left(G\left(R_{2}x\right)-G\left(R_{2}y\right)\right)}{\left(G\left(x\right)-G\left(y\right)\right)}\\&=\lim_{x\to 0}\frac{\lambda G\left(R_{1}x\right)+(1-\lambda)G\left(R_{2}x\right)}{G\left(x\right)},\end{aligned}$$

so in particular the latter exists and is finite, for all $\lambda$ and $\gamma$. This implies that

$$\lim_{x\to 0}\frac{G(Rx)}{G(x)}\quad\text{exists for all } R.$$

Notice that

$$\lim_{x\to 0}\frac{\lambda G\left(R_{1}x\right)+(1-\lambda)G\left(R_{2}x\right)}{G\left(x\right)}=\lim_{x\to 0}\frac{\lambda G\left(R_{1}x\right)+(1-\lambda)G\left(R_{2}x\right)}{G\left(Rx\right)}\cdot\frac{G(Rx)}{G(x)},$$

and the limit on the LHS exists, and so does $\lim_{x\to 0}\frac{\lambda G(R_1x)+(1-\lambda)G(R_2x)}{G(Rx)}$ (and it is generically not zero). In fact, we know that for all $R$, $\lim_{x\to 0}\frac{G(Rx)}{G(x)}$ must be finite because of equation (A.6). So we can define:

$$L(R) := \lim_{x \to 0} \frac{G(Rx)}{G(x)}, \tag{A.8}$$

and observe that $L(R) = R^{\alpha}$ for some $\alpha \in \mathbb{R}$. This is a classical result in the theory of *slowly varying function*, and it can be easily deduced by noticing that:

$$L(R) = \lim_{x \to 0} \frac{G(Rx)}{G(x)} = \lim_{x \to +\infty} \frac{G(Rx)}{G(R'x)} \frac{G(R'x)}{G(x)} = L\left(\frac{R}{R'}\right) L(R').$$

Using again equation (A.6), we then find that:

$$\lim_{x \to +\infty} \frac{\lambda \left(G\left(R_1 x\right) - G\left(R_1 y\right)\right) + (1-\lambda)\left(G\left(R_2 x\right) - G\left(R_2 y\right)\right)}{\left(G\left(x\right) - G\left(y\right)\right)} = \lambda \left(\frac{\lambda}{\gamma}\right)^{\alpha} + (1-\lambda)\left(\frac{1-\lambda}{1-\gamma}\right)^{\alpha}.$$

But since by equation (A.6), the function

$$\frac{\lambda \left(G\left(R_1 x\right) - G\left(R_1 y\right)\right) + (1-\lambda)\left(G\left(R_2 x\right) - G\left(R_2 y\right)\right)}{\left(G\left(x\right) - G\left(y\right)\right)}$$

does not depend on $x$, we have that *for all* $x \in (0, +\infty)$:

$$\frac{\lambda \left(G\left(R_1 x\right) - G\left(R_1 y\right)\right) + (1-\lambda)\left(G\left(R_2 x\right) - G\left(R_2 y\right)\right)}{\left(G\left(x\right) - G\left(y\right)\right)} = \lambda \left(\frac{\lambda}{\gamma}\right)^{\alpha} + (1-\lambda)\left(\frac{1-\lambda}{1-\gamma}\right)^{\alpha}, \tag{A.9}$$

and this equation must hold for all $x, y, \gamma, \lambda$.

Moreover, having assumed that $D$ is three times continuously differentiable, we can take the derivative with respect to $x$ finding:

$$G'(x)\left(\lambda\left(\frac{\lambda}{\gamma}\right)^{\alpha}+(1-\lambda)\left(\frac{1-\lambda}{1-\gamma}\right)^{\alpha}\right)=\left(\lambda R_1 G'(R_1x)+(1-\lambda)R_2G'(R_2x)\right), \tag{A.10}$$

for all $x\in[0,+\infty)$ and $\lambda,\gamma\in(0,1)$.

We now divide the study into the two cases of $\alpha=0,-1$ and $\alpha\neq 0,-1$.

1. if $\alpha=0$ or $\alpha=-1$ we get that equation (A.9) becomes:

   $$\lambda\left(G\left(R_1x\right)-G\left(R_1y\right)\right)+(1-\lambda)\left(G\left(R_2x\right)-G\left(R_2y\right)\right)=\left(G\left(x\right)-G\left(y\right)\right),$$

   and differentiating with respect to $x$ we get:

   $$G'(x)=\left(\lambda R_1G'(R_1x)+(1-\lambda)R_2G'(R_2x)\right),$$

   now multiply both sides by $x$ and define $\phi(x):=xG'(x)$, which yields:

   $$\phi(x)=\lambda\phi(R_1x)+(1-\lambda)\phi(R_2x),$$

   and at $\lambda=\frac{1}{2}$ we obtain:

   $$2\phi(x)=\phi\left(\frac{x}{2\gamma}\right)+\phi\left(\frac{x}{2(1-\gamma)}\right).$$

then differentiating with respect to $\gamma$ we find:[67]

$$0 = -\frac{2x}{(2\gamma)^2}\phi'\left(\frac{x}{(2\gamma)}\right) + \frac{2x}{(2(1-\gamma))^2}\phi'\left(\frac{x}{(2(1-\gamma))}\right).$$

Now define $z = \frac{x}{2\gamma}$, and simplify so that we get:

$$\phi'\left(z\frac{\gamma}{1-\gamma}\right) = \left(\frac{\gamma}{1-\gamma}\right)^{-2}\phi'(z),$$

which implies that $\phi'(z) = \frac{c}{z^2}$ for some $c \in \mathbb{R}$ (since $\gamma \in (0,1)$ implies that $\gamma/(1-\gamma)$ spans $\mathbb{R}^+$, and $z$ can be any positive number). But then $\phi(z) = \frac{a}{z} + b$ (by integrating $\phi'(z)$) and given the definition of $\phi(\cdot)$ we get that a further integration yields:

$$G'(x) = \frac{a}{x^2} + \frac{b}{x}, \tag{A.11}$$

as in the thesis.

2. If $\alpha \neq 0, -1$ let us consider equation (A.10) and once again define $\phi(x) := xG'(x)$ so that we get:

$$\phi(x)\left(\lambda\left(\frac{\lambda}{\gamma}\right)^{\alpha} + (1-\lambda)\left(\frac{1-\lambda}{1-\gamma}\right)^{\alpha}\right) = \lambda\phi(R_1 x) + (1-\lambda)\phi(R_2 x). \tag{A.12}$$

Observe that since the left hand side is continuous in $\lambda \in [0,1]$, so is the right hand

[67]Here, in order to differentiate with respect to $\gamma$, we need to have that $G'$ (which remember is the second derivative of $D$) is differentiable. This is the only step of the proof where we use $D$ three times differentiable.

side, and hence the limit for $\lambda \to 0$ is well-defined. Now, define the function:

$$c(r) := \lim_{\lambda \to 0} \lambda \phi\left(\lambda r\right). \tag{A.13}$$

Notice that $c(r) = \frac{d}{r}$, for some $d \in \mathbb{R}$. As a matter of fact, for all $r' \neq r$ we have that:

$$c(r') = \lim_{\lambda \to 0} \lambda \phi\left(\lambda r'\right) = \frac{r}{r'} \lim_{\lambda r' \to 0} \lambda \frac{r'}{r} \phi\left(\lambda \frac{r'}{r} r\right) = \frac{r}{r'} c(r).$$

Now, suppose $\alpha > 0$, and take the limit for $\lambda \to 0$ in equation (A.12). We get that:

$$\phi(x) \frac{1}{(1-\gamma)^\alpha} = c\left(\frac{x}{\gamma}\right) + \phi\left(\frac{1}{(1-\gamma)} x\right), \tag{A.14}$$

where we used the continuity of $\phi$ in $(0, +\infty)$, and the definition of $c$ given in (A.13). We will now prove that it must be that $c(r) = 0$. Rewriting (A.14) and substituting $c(r) = \frac{d}{r}$ we get that:

$$d = \frac{x\phi(x)}{\gamma} \frac{1}{(1-\gamma)^\alpha} - \frac{x}{\gamma} \phi\left(\frac{1}{(1-\gamma)} x\right),$$

and letting $x \to 0$ on the right hand side we get that $x\phi(x) \to d$ and $\frac{x}{\gamma}\phi\left(\frac{1}{(1-\gamma)}x\right) \to \frac{d(1-\gamma)}{\gamma}$, (because of equation (A.13)). Thus:

$$d = \lim_{x \to 0} \frac{x\phi(x)}{\gamma} \frac{1}{(1-\gamma)^\alpha} - \frac{x}{\gamma} \phi\left(\frac{1}{(1-\gamma)} x\right) = \frac{d}{\gamma(1-\gamma)^\alpha} - \frac{d(1-\gamma)}{\gamma},$$

and since this has to be true for all $\gamma$ and $\alpha \neq -1, 0$, then the only possibility is

$d = 0$, so the function $c(r)$ is constantly $0$.

But then going back to (A.14) and substituting $c(x/r) = 0$ we get that:

$$\phi\left(\frac{1}{(1-\gamma)}x\right) = \frac{1}{(1-\gamma)^{\alpha}}\phi(x).$$

For the generality of $x$ and $\gamma$ we find that $\phi(x)$ is homogeneous of degree $\alpha$:

$$\phi(x) = cx^{\alpha}, \Rightarrow G'(x) = cx^{\alpha-1}, \tag{A.15}$$

and hence we get the thesis.

Observe that as $\alpha$ varies in $\mathbb{R}$ we find that $G'(x)$ can be *any homogeneous function.* Equation (A.15) covers all homogeneous functions except those of degree $-1, -2$ (since $\alpha \neq 0, -1$), but those are the cases covered in the first case, taking $a = 0$ or $b = 0$ (respectively) in equation (A.11). $\square$

This implies that:

**Result** 7. *$G$ takes either of these functional forms:*

$$G(x) = ax^{\alpha} + b \quad \exists \alpha \neq 0, -1$$

*or*

$$G(x) = a\log(x) + \frac{b}{x} + c,$$

*where* $a, b, c, z \in \mathbb{R}$.

*Proof.* In Result 6 we showed that $G'(x) = ax^{\alpha}$ for some $\alpha \in \mathbb{R}$ or $G'(x) = \frac{a}{x} + \frac{b}{x^2}$. This result gives a family of primitives for these functions $G'(x)$. $\square$

The next part of the proof builds on Result 7 to pin down the function $\alpha(p, q)$:

**Result 8.** *If* $G(x) = a\log(x) + \frac{b}{x} + c$*, then* $\alpha(p, q) = K$*.*

*If* $G(x) = ax^z + b$*, then for all* $p, q \in \Delta_n^{\circ}$*:*

$$\alpha(p, q) = K\left(\sum_j p_j \left(\frac{p_j}{q_j}\right)^z\right).$$

*Proof.* Let us rewrite equation (A.5), for $\lambda, \gamma \in \Delta_m^{\circ}$ (for a general $m$, instead of $m = 2$ as for the proof of Result 6). We get:

$$\frac{\alpha(p * \lambda, q * \gamma)}{\alpha(p, q)} = \sum_{j=1}^{m} \lambda_j \left(\frac{\lambda_j}{\gamma_j}\right)^{\alpha},$$

The case of $\alpha = 0, -1$ is easier, and it yields $\frac{\alpha(p*\lambda, q*\gamma)}{\alpha(p,q)} = 1$, which implies that $\alpha(\cdot, \cdot)$ is constant, for the generality of $p, q, \lambda, \gamma$.

Now in the case $\alpha \neq 0, -1$ we can write:

$$\alpha(p * \lambda, q * \gamma) = \alpha(p, q) \sum_{j=1}^{m} \lambda_j \left(\frac{\lambda_j}{\gamma_j}\right)^{\alpha}.$$

Since $D(p * \lambda, q * \gamma) = D(\lambda * p, \gamma * q)$ (by Axiom 2) then its derivatives must also coincide and we get that flipping $p, q$ with $\lambda, \gamma$:

$$\alpha(p * \lambda, q * \gamma) = \alpha(\lambda, \gamma) \sum_j p_j \left(\frac{p_j}{q_j}\right)^{z},$$

and this gives that

$$\frac{\alpha(p,q)}{\sum_j p_j \left(\frac{p_j}{q_j}\right)^z} = \frac{\alpha(\lambda,\gamma)}{\sum_{j=1}^m \lambda_j \left(\frac{\lambda_j}{\gamma_j}\right)^z},$$

and then $\frac{\alpha(p,q)}{\sum_j p_j \left(\frac{p_j}{q_j}\right)^z}$ does not depend on $p, q$ and so it must be constant. $\square$

This final result, buys us the main theorem.

**Theorem 1**. *The only functions $D_\Theta$ that satisfy Axioms 1, 2, 3, 4, 5, 6 have the following functional form (for all $\Theta$ finite, and all $p, q \in \Delta(\Theta)$):*[68]

1. *either:*

$$D_\Theta(p,q) = a \sum_{\theta \in \Theta} p(\theta) \log\left(\frac{p(\theta)}{q(\theta)}\right) + b \sum_\theta q(\theta) \log\left(\frac{q(\theta)}{p(\theta)}\right), \tag{A.16}$$

   *for some $a, b \geq 0$ (not both zero);*

2. *or:*

$$D_\Theta(p,q) = a \log\left(\sum_\theta p(\theta) \left(\frac{p(\theta)}{q(\theta)}\right)^{z-0.5}\right) \quad \textit{where} \quad \begin{cases} a > 0 & \textit{if } |z| > 0.5 \\ a < 0 & \textit{if } |z| < 0.5 \end{cases} \tag{A.17}$$

   *for $z \in \mathbb{R} \setminus \{-0.5, 0.5\}$.*

*Proof of Theorem 1.* Let us show first the first case, $z \neq 0$. We showed above that under

[68] The expression is not well-defined for non fully-mixed beliefs. See Remark 1 after the Theorem for a clarification.

Axioms 1–6 the derivatives:

$$\partial_i D_z(p,q) = \frac{h(p_i,p_{i+1},q_i,q_{i+1})}{\alpha(p,q)} = a\frac{\left(\frac{p_i}{q_i}\right)^z - \left(\frac{p_{i+1}}{q_{i+1}}\right)^z}{\sum_j p_j \left(\frac{p_j}{q_j}\right)^z},$$

for some $a, z \in \mathbb{R}$.

Then integrating on the directions of $q + \epsilon z_{i,i+1}$ (for any fixed $p$) we get that all primitives are given by:

$$D(p,q) = a \log\left(\sum_j p_j \left(\frac{p_j}{q_j}\right)^z\right) + K,$$

for some $K$ constant. Since by definition we assumed that $D(q,q) = 0$ we have that $K = 0$. Also, since $D(p,q) \geq 0$ for all $p, q \in \Delta_n$ it must be that:

- if $z \in (-1, 0)$, then $a < 0$; because for $z \in (-1,0)$ notice that $x^{-z}$ is concave and then by Jensen inequality:

  $$\sum_j p_j \left(\frac{p_j}{q_j}\right)^z \leq \left(\sum_j q_j\right)^{-z} = 1,$$

  and so $a \log\left(\sum_j p_j \left(\frac{p_j}{q_j}\right)^z\right) \geq 0$ if and only if $a \leq 0$. Then because of Axiom 1 we must have that $a < 0$;

- if $z \in (-\infty, -1) \cup (0, +\infty)$, then $a > 0$. To see this notice that for $z \in (0, \infty)$ the function $x^{-z}$ is convex, and then Jensen inequality implies that:

  $$\sum_j p_j \left(\frac{p_j}{q_j}\right)^z \geq \left(\sum_j q_j\right)^{-z} = 1,$$

  and then $D(p,q) \geq 0$ if and only if $a \geq 0$. If instead $z \in (-\infty, -1)$ then notice that:

$$a \log \left( \sum_j p_j \left( \frac{p_j}{q_j} \right)^z \right) = a \log \left( \sum_j q_j \left( \frac{q_j}{p_j} \right)^{-z-1} \right),$$

and since $-z-1 \in (0, +\infty)$ we have that $z+1 \in (-\infty, 0)$ and then by Jensen inequality:

$$\sum_j q_j \left( \frac{q_j}{p_j} \right)^{-z-1} \geq \left( \sum_j p_j \right)^{z+1} = 1$$

Passing now to the case of $G(z) = a\log(x) + \frac{b}{x}$ we showed in Result 8 that $\alpha(p, q)$ is constant, which in turns implies that

$$\partial_i D(p,q) = a \left( \log \left( \frac{p_i}{q_i} \right) - \log \left( \frac{p_{i+1}}{q_{i+1}} \right) \right) + b \left( \frac{q_i}{p_i} - \frac{q_{i+1}}{p_{i+1}} \right), \tag{A.18}$$

for some constants $a, b \in \mathbb{R}$.

Now observe that for all $p$, the family of functions

$$a \sum_j q_j \log \left( \frac{q_j}{p_j} \right) + b \sum_j p_j \log \left( \frac{p_j}{q_j} \right) + K,$$

are a family of primitives, where $a$ and $b$ are the same constant of equation (A.18). Then the normalization $D(p, p) = 0$ implies that $K = 0$, and let us prove that $a$ and $b$ have to be positive in order for $D(p, q) \geq 0$.

If at least one of $a$ and $b$ is negative, say $a$, we have that taking $p = (\epsilon, 1-\epsilon, 0, \ldots, 0)$

and $q = (0.5, 0.5, 0, \ldots, 0)$ gives us:[69]

$$\lim_{\epsilon \to 0} b \sum_j p_j \log\left(\frac{p_j}{q_j}\right) = \lim_{\epsilon \to 0} b(\epsilon \log(2\epsilon) + (1-\epsilon)\log(2(1-\epsilon))) = b\log(2),$$

whereas:

$$\lim_{\epsilon \to 0} a \sum_j q_j \log\left(\frac{q_j}{p_j}\right) = -\infty, \qquad \text{if } a < 0$$

so we have that $\lim_\epsilon D(p,q) = -\infty$, and we have that $a, b$ have to be positive.

The proof above shows that if Axioms 1–6 are satisfied, then the disagreement functions must take the form of equations (A.16) and (A.17). We have not showed that those functions satisfy all the axioms (the *easiest* implication of the theorem). Let us prove it now:

1. Axiom 1 is trivial, just substitute $p = q$ and see that $D(p,p) = 0$;
2. Axiom 2 is also trivial, because changing the name of the labels the disagreement function does not change;
3. Axiom 3 is not trivial. Let us analyze first the functional form of equation (A.16). Notice that for all $x_1, x_2, y_1, y_2, \lambda \in [0,1]$ we have that:

$$(\lambda x_1 + (1-\lambda)x_2)\log\left(\frac{(\lambda x_1 + (1-\lambda)x_2)}{(\lambda y_1 + (1-\lambda)y_2)}\right) \\ \leq \lambda x_1 \log\left(\frac{x_1}{y_1}\right) + (1-\lambda)x_2\log\left(\frac{x_2}{y_2}\right), \quad \text{(A.19)}$$

[69]Here we use the continuity result that $\lim_x x\log(x) = 0$

because this equation it is equivalent to:

$$
\begin{aligned}
-\log\left(\frac{\lambda x_1}{\lambda x_1+(1-\lambda)x_2}\left(\frac{y_1}{x_1}\right)+\frac{(1-\lambda)x_2}{\lambda x_1+(1-\lambda)x_2}\left(\frac{y_2}{x_2}\right)\right)& \\
\leq -\frac{\lambda x_1}{\lambda x_1+(1-\lambda)x_2}\log\left(\frac{y_1}{x_1}\right)-\frac{(1-\lambda)x_2}{\lambda x_1+(1-\lambda)x_2}\log\left(\frac{y_2}{x_2}\right)&, \quad \text{(A.20)}
\end{aligned}
$$

which is true by the convexity of $-\log$. But then using equation (A.19) repeatedly we find that:

$$
D_{\Theta}(\lambda p^1+(1-\lambda)p^2, \lambda q^1+(1-\lambda)q^2) \leq \lambda D(p^1, q^1)+(1-\lambda)D(p^2, q^2),
$$

and since the RHS is smaller than the maximum we obtain Axiom 3. The proof for the functions of the form written in equation (A.17) is exactly analogous, after noting that $x^{z-0.5}$ is convex for $|z|>0.5$ and concave for $|z|<0.5$.

4. Axiom 4 also requires a small proof, and it resembles that just done for Axiom 3. Since we showed the proof of Axiom 6 for the functions of the form (A.16), let us show this axiom using the other, equation (A.17). Notice that proving that Axiom 4 is equivalent to proving that for all $p_1, p_2, q_1, q_2$ and for $|z|>0.5$ we have that:

$$
(p_1+p_2)\left(\frac{p_1+p_2}{q_1+q_2}\right)^{z-0.5} \leq p_1\left(\frac{p_1}{q_1}\right)^{z-0.5}+p_2\left(\frac{p_2}{q_2}\right)^{z-0.5}.
$$

To show this, again notice that it is equivalent to having:

$$
\left(\frac{p_1}{p_1+p_2}\frac{q_1}{p_1}+\frac{p_2}{p_1+p_2}\frac{q_2}{p_2}\right)^{0.5-z} \leq \frac{p_1}{p_1+p_2}\left(\frac{q_1}{p_1}\right)^{0.5-z}+\frac{p_2}{p_1+p_2}\left(\frac{q_2}{p_2}\right)^{0.5-z},
$$

and if $|z| > 0.5$ the function $x^{0.5-z}$ is convex so the result follows. If instead $|z| < 0.5$ we have to prove that the opposite inequality holds, i.e.

$$\left(\frac{p_1}{p_1+p_2}\left(\frac{q_1}{p_1}\right)+\frac{p_2}{p_1+p_2}\left(\frac{q_2}{p_2}\right)\right)^{0.5-z} \geq \frac{p_1}{p_1+p_2}\left(\frac{q_1}{p_1}\right)^{0.5-z}+\frac{p_2}{p_1+p_2}\left(\frac{q_2}{p_2}\right)^{0.5-z},$$

which is true because for $|z| < 0.5$, $0.5 - z \in (0, 1)$ and then $x^{0.5-z}$ is concave.

5. Axiom 5 can be verified directly by using the fact that $\log(ab) = \log(a) + \log(b)$ and $\sum_{i,j}(p_i p_j)^z = \sum_i p_i^z \sum_j p_j^z$.

6. Axiom 6 can be proved by taking the derivatives explicitly, and does not require any proof.

□

## Proofs on the Analysis of the disagreement functions

**Proposition 1**. *The only disagreement functions that satisfy Axioms 1–6 and are such that $D(p, q) = D(q, p)$ for all $p, q$, are proportional to:*

- *the symmetric divergence:*

$$D(p,q) = \sum_j (p_j - q_j)\log\left(\frac{p_j}{q_j}\right) = \sum_j p_j \log\left(\frac{p_j}{q_j}\right) + \sum_j q_j \log\left(\frac{q_j}{p_j}\right);$$

- *the Bhattacharyya distance:*

$$D(p,q) = -\log\left(\sum_j \sqrt{p_j q_j}\right).$$

*Furthermore, the only symmetric disagreement function that is additively separable in the states*[70] *is the symmetric divergence, and the only symmetric disagreement function such that* $D(p,q) < +\infty$ *if and only if* $Supp(p) \cap Supp(q) \neq \emptyset$ *is the Bhattacharyya distance.*

*Proof of Proposition 1.* Consider $p = (1,0)$ and $q = (q_1, 1-q_1)$ with $q_1 \neq 0.5$. Computing $D(p,q)$ for the disagreement functions of equation (1.2) (in Theorem 1) we find that $D(p,q) \neq D(q,p)$ unless $z = 0$. For the functions of equation (1.1), consider $p = (0.5, 0.5)$ and $q = (q_1, 1-q_1)$ and notice that $D(p,q) \neq D(q,p)$ if and only if $a \neq b$. The last statement of the theorem can also be obtained directly.

□

**Proposition 2.** *Let* $p, q, r \in \Delta_n^\circ$ *be beliefs ranked by Likelihood Ratio:*

$$p <_{LR} q <_{LR} r,$$

*then for any disagreement function we have that:*

$$D(p,q) \leq D(p,r).$$

*Proof of Proposition 2.* We prove this result by showing it holds for all the functions $D$

[70]A divergence metric is said to be additively separable in the states if $D(p,q) = \sum_j f_j(p_j, q_j)$, for some functions $f_j : [0,1] \times [0,1] \to \mathbb{R}$.

characterized in Theorem 1.

- let $D(p,q) = \log\left(\sum_i p_i \left(\frac{p_i}{q_i}\right)^{z-0.5}\right)$ with $z > 0.5$. The statement is equivalent to proving that for all $p, q, r$ with $p <_{LR} q <_{LR} r$,:

$$\sum_i p_i \left(\frac{p_i}{q_i}\right)^{z-0.5} \leq \sum_i p_i \left(\frac{p_i}{r_i}\right)^{z-0.5}.$$

To show this, let us prove something stronger, that is, for all $\epsilon \in (0,1)$ we have that $\partial\frac{\exp(D(p,\epsilon r+(1-\epsilon)q))}{\partial\epsilon} \geq 0$. Clearly, if this is true, then $D(p,r) \geq D(p,q)$. It is easy to find that:

$$\frac{\left(\partial\sum_i p_i \left(\frac{p_i}{\epsilon r_i+(1-\epsilon)q_i}\right)^{z-0.5}\right)}{\partial\epsilon} \geq 0$$
$$\Leftrightarrow \sum_i \left(\frac{p_i}{(1-\epsilon)q_i+\epsilon r_i}\right)^{z+0.5} r_i \leq \sum_i \left(\frac{p_i}{(1-\epsilon)q_i+\epsilon r_i}\right)^{z+0.5} q_i,$$

and we have that the latter holds because $\left(\frac{p_i}{(1-\epsilon)q_i+\epsilon r_i}\right)^{z+0.5}$ is decreasing (since $p <_{LR} \epsilon r + (1-\epsilon)q$) and $q <_{LR} r$.

- the case for $|z| < 0.5$ and $z < -0.5$ is analogous;
- the case $D(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right)$ and $D(p,q) = \sum_i q_i \log\left(\frac{q_i}{p_i}\right)$ can be obtained as limit for $z \to 0.5, -0.5$ (see Proposition 3 below) so the same result must hold.

$\square$

**Observation 1.** *For all $z \neq \tilde{z}$, the disagreement functions $D^z$ and $D^{\tilde{z}}$ are not ordinally* *equivalent, i.e. there exist* $p, q, r \in \Delta_n$ *such that:*

$$D^z(p,q) < D^z(p,r) \quad \textit{and} \quad D^{\tilde{z}}(p,q) > D^{\tilde{z}}(p,r).$$

*Proof of Observation 1.* In order to prove this result, notice that by Axiom 6 we have that if $z \neq 0.5, -0.5$:

$$\frac{\partial_i D^z(p,q)}{\partial_j D^z(p,q)} = \frac{\left(\frac{p_i}{q_i}\right)^{z+1} - \left(\frac{p_{i+1}}{q_{i+1}}\right)^{z+1}}{\left(\frac{p_j}{q_j}\right)^{z+1} - \left(\frac{p_{j+1}}{q_{j+1}}\right)^{z+1}},$$

and observe that this is can be interpreted as (minus) the slope of implicit function defined by:

$$D(p,q) = K = D(p, (q_1, \ldots, q_i + \epsilon, q_{i+1} - \epsilon, \ldots, q_j + g_z(\epsilon), q_{j+1} - g_z(\epsilon), \ldots, q_n)),$$

plainly:

$$g_z'(0) = -\frac{\partial_i D^z(p,q)}{\partial_j D^z(p,q)} = -\frac{\left(\frac{p_i}{q_i}\right)^{z+1} - \left(\frac{p_{i+1}}{q_{i+1}}\right)^{z+1}}{\left(\frac{p_j}{q_j}\right)^{z+1} - \left(\frac{p_{j+1}}{q_{j+1}}\right)^{z+1}}. \tag{A.21}$$

Now for any $z \neq \tilde{z}$ we can find $p, q \in \Delta_n^\circ$ with:

$$g_z'(0) \neq g_{\tilde{z}}'(0).$$

This implies that locally around $q$ the ball of radius $D(p,q)$ around $p$ (in the plane spanned by $z_i$ and $z_j$) are different.

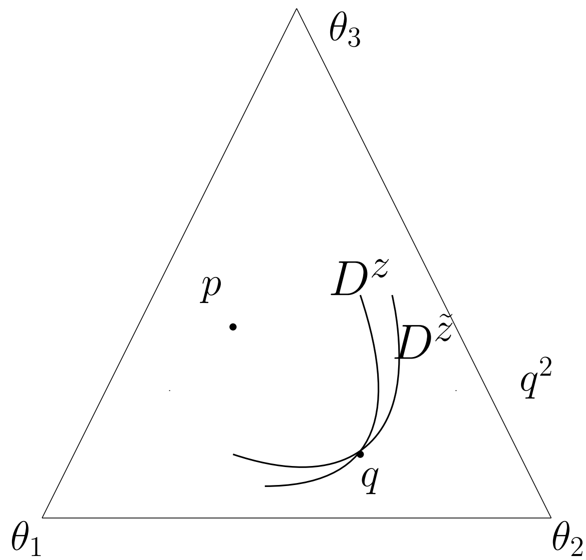



Figure A.1: The two balls of radius $D^z(p,q)$ and $D^{\tilde{z}}(p,q)$. If $z \neq \tilde{z}$ then the tangent in $p$ has different slope (generically), as it can be seen by equation (A.21).

This means we can find $r$ arbitrary close to $q$ such that:

$$D_z(p,q) < D_z(p,r) \quad \text{and} \quad D_{z'}(p,r) < D_{z'}(p,q).$$

If we consider disagreement functions such as those of equation (1.1), the result is analogous. □

**Proposition 3**. *For all $p,q \in \Delta_n^\circ$ and for all $z \neq 0.5, -0.5$:*

1. *$D^z(p,q) = D^{-z}(q,p)$;*
2. *for all $z \in \mathbb{R}$, $z \neq 0.5$:*

$$\frac{1}{z-0.5} D^z(p,q) = \log(\|p(\theta)/q(\theta)\|^p_{z-0.5}),$$

*where the $w$-norm of a function $f : \Theta \to \mathbb{R}^+$, is defined by $\|f(\theta)\|^p_w := (\sum_i |f(\theta_i)|^w p(\theta_i))^{1/w}$;*

3. *$D^z$ can be extended by continuity at $z = 0.5$ and $z = -0.5$:*

$$\lim_{z \to 0.5} \frac{D^z(p,q)}{z - 0.5} = \sum_j p_j \log\left(\frac{p_j}{q_j}\right) =: D^{0.5}(p,q)$$

$$\lim_{z \to -0.5} \frac{D^z(p,q)}{-z - 0.5} = \sum_j q_j \log\left(\frac{q_j}{p_j}\right) =: D^{-0.5}(p,q).$$

*Proof of Proposition 3.* Points 1 and 2 follow simply by writing the definitions of $D^z$, the only point that requires a short proof is 3.

To show this limit recall that for all $f$ positive we have that:

$$\lim_{w \to 0} \|f\|_w^p = \exp\left(\sum_j p_j \log\left(\frac{p_j}{q_j}\right)\right),$$

so that applying this to the function $p(\theta)/q(\theta)$ and taking the logarithm we obtain the first limit result of point 3. To get the second result notice that thanks to point 1 we get:

$$\lim_{z \to -0.5} \frac{D^z(p,q)}{-z - 0.5} = \lim_{z \to -0.5} \frac{D^{-z}(q,p)}{-z - 0.5} \overset{w=-z}{=} \lim_{w \to 0.5} \frac{D^w(q,p)}{w - 0.5},$$

so the result follows from the first limit. □

**Lemma 1.** *If $|z| < 0.5$, then $D^z(p,q) = +\infty$ if and only if $Supp(p) \cap Supp(q) = \emptyset$. If $|z| \geq 0.5$ then:*

- *if $z \geq 0.5$ then $D^z(p,q) = +\infty$ if and only if $Supp(p) \setminus Supp(q) \neq \emptyset$.*
- *if $z \leq -0.5$ then $D^z(p,q) = +\infty$ if and only if $Supp(q) \setminus Supp(p) \neq \emptyset$.*

*Proof of Lemma 1.* Recall that $D^z(p,q)$ is defined, for $|z| < 0.5$

$$D^z(p,q) = -\log\left(\sum_j p_j^{z+0.5} q_j^{0.5-z}\right),$$

hence if $|z| < 0.5$, then $0.5 + z > 0 < 0.5 - z$. Then, we have that the argument of the logarithm is finite, for all $p, q \in \Delta_n$. Therefore we have that $D^z(p,q) = +\infty$ if and only if $\sum_j p_j^{z+0.5} q_j^{0.5-z} = 0$. But since this is a sum of positive terms, it is $0$ only if each term is $0$. On the other hand, for $z > 0.5$, notice that the exponent of $q_j$ is negative, so if $\theta_j \in Supp(p) \setminus Supp(q)$ we have that the sum in the argument of $D^z$ is infinite, because of the $j-$th term alone. For the opposite implication, note that if $Supp(p) \subseteq Supp(q)$ then whenever $q_j = 0$, so is $p_j$, and then $D(p,q) < +\infty$.

□

**Corollary 1.**

1. *if $z \geq 0.5$, $D^z(p,\cdot) : \Delta_n \to \mathbb{R}$ is continuous for all fixed $p \in \Delta_n$. Furthermore $D^z$ depends only on the states $\theta \in Supp(p)$:*

$$D^z(p,q) = \begin{cases} \log\left(\sum_{j\in Supp(p)} p_j \left(\frac{p_j}{q_j}\right)^{z-0.5}\right) & \textit{if } z > 0.5 \\ \sum_{j\in Supp(p)} p_j \log\left(\frac{p_j}{q_j}\right) & \textit{if } z = 0.5 \end{cases}.$$

   *On the other hand, $D^z(\cdot,q) : \Delta_n \to \mathbb{R}$ is continuous if and only if $q \in \Delta_n^\circ$.*

2. *if $z \leq -0.5$, $D^z(\cdot,q) : \Delta_n \to \mathbb{R}$ is continuous for all fixed $q \in \Delta_n$. Furthermore $D^z$*

*depends only on the states $\theta \in Supp(q)$:*

$$D^z(p,q) = \begin{cases} \log\left(\sum_{j\in Supp(q)} p_j \left(\frac{p_j}{q_j}\right)^{z-0.5}\right) & \text{if } z > 0.5 \\ \sum_{j\in Supp(q)} p_j \log\left(\frac{p_j}{q_j}\right) & \text{if } z = 0.5 \end{cases}.$$

*On the other hand, $D^z(p,\cdot) : \Delta_n \to \mathbb{R}$ is continuous if and only if $p \in \Delta_n^\circ$.*

3. *If $|z| < 0.5$, then $D^z : \Delta_n \times \Delta_n \to \mathbb{R}$ is continuous, so it is continuous on both variables separately, and:*

$$D^z(p,q) = -\log\left(\sum_{j\in Supp(p)\cap Supp(q)} p_j \left(\frac{p_j}{q_j}\right)^{z-0.5}\right),$$

*with the convention that $\sum_{j\in\emptyset} = 0$.*

*Proof of Corollary 1.* 1. ($z \geq 0.5$) Consider $z > 0.5$ first. The fact that $D^z(p,\cdot)$ is continuous on $\Delta_n^\circ$ is trivial, because we assumed that $D^z$ be three times continuously differentiable in both variables. To show that it is continuous also on the boundary of $\Delta_n$, consider a sequence of $(q^{(n)})_n$ converging to a point in the boundary, say $q_j^{(n)} \to 0$. If $p_j = 0$ the statement is trivial, because $D^z(p, q^{(n)})$ does not depend on the state $j$. If instead $p_j > 0$, we have that:

$$D^z(p, q^{(n)}) \geq \log\left(p_j^{z+0.5}(q_j^{(n)})^{0.5-z}\right) \geq (0.5-z)\log(q_j^{(n)}) \to +\infty,$$

since $q_j^{(n)} \to 0$ and $(0.5-z) < 0$. Therefore $D^z(p, q^{(n)}) \to +\infty$, and by our conventions (Remark 1) $D(p,r) = +\infty$ anytime $p_j > 0$ and $r_j = 0$. Hence $D^z(p,\cdot)$

is continuous on $\Delta_n$. Given this result we have that:

$$D^z(p,q) = \log\left(\sum_{j\in Supp(p)} p_j \left(\frac{p_j}{q_j}\right)^{z-0.5}\right).$$

Let us prove that $D^z(\cdot, q) : \Delta_n \to \mathbb{R}$ is continuous if and only if $q \in \Delta_n^\circ$. If $q \in \Delta_n^\circ$, then $D^z(\cdot, q)$ is continuous, because it is finite on $\Delta_n$. On the other hand, let us show that if $q \notin \Delta_n^\circ$ then $D^z(\cdot, q)$ is not continuous. Without loss of generality, consider $q$ with $q_1 = 0$, and $q_j > 0$ for all $j = 2, \ldots, n$. Let us prove that if $z > 0.5$, $D^z(p, q)$ is discontinuous in $p$. Consider a sequence $p^{(m)}$ defined by:

$$p_j^{(m)} := \begin{cases} \frac{1}{m} & \text{if } j = 1 \\ \frac{(m-1)}{m(n-1)} & \text{if } > 1 \end{cases}$$

it is easy to see that $D_w(p^{(m)}, q) = +\infty$ for all $m$, because $Supp(p) \setminus Supp(q) \neq \emptyset$, and

$$p^m \to p^* = \left(0, \frac{1}{n-1}, \ldots, \frac{1}{n-1}\right),$$

so $D^z(p^*, q) < +\infty$ because $Supp(q) \setminus Supp(p) = \emptyset$.

The case of $z = 0.5$ is similar.

2. this case is analogous to the first one, changing $p$ with $q$ and using the first point of Proposition 3;
3. finally if $|z| < 0.5$ notice that both $z + 0.5$ and $z - 0.5$ are positive, hence there are

no discontinuities at the boundary and trivially:

$$D^z(p,q) = -\log\left(\sum_{j\in Supp(p)\cap Supp(q)} p_j^{z+0.5} q_j^{z-0.5}\right).$$

$\square$

**Proposition 4.** *Let $|z| \geq 0.5$ then for any $\epsilon, \delta > 0$ (arbitrary small) there exist $\bar{p}, \bar{q}, \underline{p}, \underline{q} \in \Delta_n^\circ$ such that:*

$$\|\bar{p} - \bar{q}\|_\infty > \max_{x,y\in\Delta_n} \|x-y\|_\infty - \delta, \qquad \|\underline{p} - \underline{q}\|_\infty < \epsilon,$$

*and $D^z(\underline{p}, \underline{q}) > D^z(\bar{p}, \bar{q})$.*

*If $|z| < 0.5$, instead, $D^z(p,q)$ is uniformly continuous with respect to the metric induced by $\|\cdot\|_\infty$. Namely, for all $\epsilon$ there exist a $\delta$ such that if $\|p-q\|_\infty \leq \delta$ then $D^z(p,q) < \epsilon$.*

*Proof of Proposition 4.* For the case $|z| > 0.5$ it is enough to prove the statement for $n = 2$. Then the result for a general $n$ will follow by considering vectors of the form $(p, 1-p, 0, \ldots, 0)$.

Notice that $\max_{x,y\in\Delta_2} \|x-y\|_\infty = 1$. Fix $\epsilon$ and $\delta$ small. Pick $\bar{p}, \bar{q} \in \Delta_n^\circ$ with $\|\bar{p}-\bar{q}\| > 1-\delta$. Since $\bar{p}, \bar{q} \in \Delta_n^\circ$ we have that $D(\bar{p}, \bar{q}) < +\infty$. Let us show that we can find two sequences $(p^m)_m, (q^m)_\in \Delta_n^\circ$ such that $\|p^m - q^m\| \leq \epsilon$ and $D(p^m, q^m) \to +\infty$ as $m \to +\infty$. Once we prove this, the thesis will follow.

Pick $q^m := (1/m, 1-1/m)$ and $p^m = (1/m+\epsilon, 1-1/m-\epsilon)$. Clearly, $\|p^m - q^m\| \leq \epsilon$. If $z > 0.5$:

$$D^z(p,q) \geq \log\left(p_1^m \left(\frac{p_1^m}{q_m^1}\right)^{z-0.5}\right) = \log\left[\left(\frac{1}{m}+\epsilon\right)\left(\frac{\frac{1}{m}+\epsilon}{\frac{1}{m}}\right)^{z-0.5}\right],$$

and as $m \to +\infty$ the right hand side clearly tends to $\infty$, so we get our result. If $z = 0.5$ we can use the same sequence, that yields:

$$D^{0.5}(p,q) \geq p_1^m \log\left(\frac{p_1^m}{q_1^m}\right) = \left(\frac{1}{m} + \epsilon\right)\log\left(\frac{\frac{1}{m}+\epsilon}{\frac{1}{m}}\right),$$

which also diverges as $m \to +\infty$, for any $\epsilon > 0$. The case $z < -0.5$ is analogous, after inverting $p$ with $q$.

To get that instead $D^z$ is uniformly continuous for $|z| < 0.5$, let us show it first for $n = 2$ and then explain how the same argument holds for $n \geq 2$ too. Consider all the beliefs of the form $p = (x, 1-x)$ and $q = (x+\epsilon, 1-x-\epsilon)$, as $x \in [0, 1-\epsilon]$. It is easy to see that (let $\alpha := z + 0.5$ and notice that $\alpha \in (0,1)$, if $|z| < 0.5$):

$$D^z(p,q) = -\log\left(x^\alpha(x+\epsilon)^{1-\alpha} + (1-x)^\alpha(1-x-\epsilon)^{1-\alpha}\right),$$

and one can check that both $x^\alpha(x+\epsilon)^{1-\alpha}$ and $(1-x)^\alpha(1-x-\epsilon)^{1-\alpha}$ are concave. Therefore $x^\alpha(x+\epsilon)^{1-\alpha} + (1-x)^\alpha(1-x-\epsilon)^{1-\alpha}$ achieves its minimum at $x = 0$ or $x = 1-\epsilon$, and we have that:

$$D^z(p,q) \leq \max\{-\log(\epsilon^\alpha + (1-\epsilon)^{1-\alpha}); -\log((1-\epsilon)^\alpha + \epsilon^{1-\alpha})\}.$$

For $\epsilon \to 0$ we have that both $-\log(\epsilon^\alpha + (1-\epsilon)^{1-\alpha})$ and $-\log((1-\epsilon)^\alpha + \epsilon^{1-\alpha})$ are positive and converge to $0$, thus simply take

$$\delta := \max\{-\log(\epsilon^\alpha + (1-\epsilon)^{1-\alpha}); -\log((1-\epsilon)^\alpha + \epsilon^{1-\alpha})\},$$

and we have the thesis for $n = 2$.

For a general $n$, notice that for any $p, q$ if $p_j = 0$ for some $j$ (say $p_1 = 0$) then $D^z(p,q) \leq D^z(p,(0,q_1+q_2,\dots,q_n))$, so we can reduce ourselves to a case of dimension $n-1$ and the argument follows by induction. Suppose instead that $p \in \Delta_n^\circ$. Pick $q$ such that $\|p-q\| \leq \epsilon$, this implies that $|p_j - q_j| \leq \epsilon$ for all $j$. Now suppose without loss that $p_1 < q_1$ and consider the beliefs:[71]

$$p' = (0, p_1 + p_2, p_3, \dots, p_n), \quad q' = (q_1 - p_1, q_2 + p_1, q_3, \dots, q_n).$$

An argument identical to that done for $n = 2$ shows that:

$$D^z(p,q) \leq D^z(p', q'),$$

but now we have that $p' \notin \Delta_n^\circ$, so we can reduce ourselves to a simplex of dimension $n-1$ and we are done. $\square$

## A.2 Proofs of Section 1.3

**Lemma 2.** *If a family of disagreement functions $D_\Theta$ for any state space $\Theta$ satisfies Axiom 2 then so does the family $(\tilde{D}_n)_{n\geq 2, n\in\mathbb{N}}$, where:*

$$\tilde{D}_n : \Delta(\Theta) \times \Delta(\Theta) \to \mathbb{R}^+ \cup \{+\infty\}, \qquad \tilde{D}_n(p,q) := D_\Theta(p,q),$$

[71]The notation makes sense if $q_2 + p_1 < 1$, notice that if this were not the case we can find another $j$ for which $q_j + p_1 < 1$. I assume such $j$ is 2.

*for any* $\Theta$ *with* $|\Theta| = n$.

*Proof of Lemma 2.* Trivial. $\square$

**Lemma 3.** *If* $D_n$ *satisfies Axiom 3 and Axiom 1 then:*

1. *for all* $p, q \in \Delta_n$ *and for all* $\lambda \in [0, 1]$*:*

$$D_n(p, \lambda p + (1 - \lambda)q) \leq D_n(p, q).$$

2. *For all* $p, q, r \in \Delta_n$ *and* $\lambda \in [0, 1]$*:*

$$D_n(\lambda p + (1 - \lambda)r, \lambda q + (1 - \lambda)r) \leq D_n(p, q).$$

3. *For all* $p, q^1, q^2$ *and* $\lambda \in [0, 1]$*:*

$$D_n(p, \lambda q^1 + (1 - \lambda)q^2) \leq \max\{D_n(p, q^1), D_n(p, q^2)\}.$$

*Also, this last property is equivalent to assuming that the balls:*

$$B(p, \rho] := \{q \in \Delta_n \mid D_n(p, q) \leq \rho\}$$

*are convex for all* $D_n$.

*Proof of Lemma 3.* Notice that the first statement follows directly by taking $p^1 = p^2 = p$ and $q^1 = p$ in Axiom 3. the second statement follows taking $p^1 = p$, $p^2 = r = q^2$, and

$q^1 = q$. Finally the third statement follows by taking $p^1 = p^2 = p$. The fact that balls in the metric induced by $D_n$ are convex follows directly from equation 3. $\square$

**Corollary 2.** *For every $C \subset \Delta_n$ convex:*

$$\sup_{p,q \in C} D_n(p,q) = D_n(\bar{p}, \bar{q}),$$

*for some $\bar{p}, \bar{q} \in \partial C$. This implies that for all $p, q \in \Delta_n$ we have that:*

$$D_n(p,q) \leq D_n(e^1, e^2) \quad (= D_n(e^j, e^i) \quad \forall\, i,j),$$

*where $e^1 = (1,0,0,\ldots,0)$ and $e^2 = (0,1,0,\ldots,0)$.*

*Proof of Corollary 2.* Let $s := \sup_{p,q \in C} D_n(p,q)$ (in general $s \in \mathbb{R}^+ \cup \{+\infty\}$). Consider a sequence of $(p_m, q_m) \in C$ such that $D_n(p_m, q_m) \to s$. Now consider the sequence $(\tilde{p}_m, \tilde{q}_m)$ such that $\tilde{p}_m, \tilde{q}_m \in \partial C$, and[72] $[p_m, q_m] \subseteq [\tilde{p}_m, \tilde{q}_m]$. By Lemma 3 we have that $D_n(\tilde{p}_m, \tilde{q}_m) \geq D_n(p_m, q_m)$. So we have that $\lim_m D_n(\tilde{p}_m, \tilde{q}_m) \geq s$, which implies that $\lim_m D_n(\tilde{p}_m, \tilde{q}_m) = s$, because $s$ is the sup on $C$.

The set $\partial C$ is compact (it is closed and bounded, and we are in a normed space), so $(\tilde{p}_m, \tilde{q}_m)$ admits a converging subsequence, with limit in $\partial C$. Call such limit $(p^*, q^*)$, it must be that $D_n(p^*, q^*) = s$ (by continuity of $D$), hence the thesis. $\square$

**Proposition 5.** *Let $D$ be a smooth measure of disagreement satisfying Axioms 1–4. If $D$*

[72] For any two $a, b \in \Delta_n$, the set $[a,b]$ denotes the segment joining them:

$$[a,b] := \{q \in \Delta_n \mid q = \lambda a + (1-\lambda) b, \quad \exists \lambda \in [0,1]\}.$$

*satisfies Property 1 and is not locally constant, there exists a strictly increasing function* $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ *such that the disagreement function* $\tilde{D} := \phi(D)$ *satisfies Axiom 5.*

*Proof of Proposition 5.* Recall by Corollary 2 we have that $D_n(e^1, e^2) \geq D_n(p, q)$ for all $p, q \in \Delta_n$, where $e^1 = (1, 0, \ldots, 0)$ and $e^2 = (0, 1, 0, \ldots, 0)$. Also, notice that for all $p, q \in \Delta_n$ we have that:

$$D_n(p, q) = D_{n+1}((p_1, \ldots, p_n, 0), (q_1, \ldots, q_n, 0)),$$

by Lemma 4.

Let $n = 2$ and notice that there are two mutually exclusive and collectively exhaustive cases:

1. $D_2(e_1, e_2) > D_2(e_1, (x, 1-x))$ for all $x \in (0, 1)$;
2. $D_2(e_1, e_2) = D_2(e_1, (x, 1-x))$ for some $x \in (0, 1)$;

Consider the first case. Let us prove that for all $p, q \in \Delta_n$ there must exist an $x \in (0, 1)$ such that:

$$D_n(p, q) = D_2(e^1, (x, 1-x)),$$

and for brevity define the function $\phi(x) = D_2(e^1, (x, 1-x))$. This follows directly from the fact that $\phi(1) = 0$ and $\phi(x) \to \infty$ as $x \to 0$, therefore for all $p, q \in \Delta_n$ there exists $x$ such that $D_n(p, q) = D_2(e^1, (x, 1-x))$. Using then that $D_2$ cannot be locally constant implies that such $x$ is unique.

Let us now show that for all $D_n$, we can find $\psi$ such that $\psi(D_n(p * q, r * s)) =$

$\psi(D_n(p,r)) + \psi(D_n(q,s))$. Fix a measure $D_n$, and suppose that for some $x, y, z$:

$$D_4(e^1 * e^1, (x, 1-x) * (y, 1-y)) = D_2(e^1, (z, 1-z)).$$

We define $\psi$ so that if this is the case, then:

$$\psi(D_2(e^1, (x, 1-x))) + \psi(D_2(e^1, (y, 1-y))) = \psi(D_2(e^1, (z, 1-z)))$$

To show that this is a well-defined function, we need to show that if for some $x'$ and $y'$ we have that:

$$D_4(e^1 * e^1, (x, 1-x) * (y, 1-y)) = D_4(e^1 * e^1, (x', 1-x') * (y', 1-y')), \qquad \text{(A.22)}$$

then $\psi(\phi(x)) + \psi(\phi(y)) = \psi(\phi(x')) + \psi(\phi(y'))$. To prove this last point, suppose, without loss, that $\phi(x') > \phi(x)$. Then there exists $w$ such that:

$$D_4(e^1 * e^1, (w, 1-w) * (x, 1-x)) = D_2(e^1, (x', 1-x')),$$

and then by Property 1 we have that:

$$D_8(e^1 * e^1 * e^1, (w, 1-w) * (x, 1-x) * (y, 1-y))$$
$$= D_4(e^1 * e^1, (x', 1-x') * (y, 1-y)),$$

and by (A.22) we get that:

$$
\begin{aligned}
D_8(e^1 * e^1 * e^1, (w, 1-w) * (x, 1-x) * (y, 1-y))& \\
&= D_8(e^1 * e^1 * e^1, (w, 1-w) * (x', 1-x') * (y', 1-y')),
\end{aligned}
$$

so combining these last two equations (and eliminating the $x'$ term) we get that:

$$D_4(e^1 * e^1, (w, 1-w) * (y', 1-y')) = D_2(e^1, (y, 1-y)).$$

Now since $\phi(x') > \phi(x)$, it must be that $\phi(y') < \phi(y)$, so there exists $w'$ such that:

$$D_4(e^1 * e^1, (w', 1-w') * (y', 1-y')) = D_2(e^1, (y, 1-y)),$$

and repeating the same procedure just carried out for $w$ we obtain that:

$$D_2(e^1, (w, 1-w)) = D_2(e^1, (w', 1-w')),$$

which implies that $\psi(\phi(y')) - \psi(\phi(y)) = -(\psi(\phi(x')) - \psi(\phi(x)))$ and then yields that $\psi(\phi(x)) + \psi(\phi(y)) = \psi(\phi(x')) + \psi(\phi(y'))$.

Let us now prove that for all $p, q, r, s$ we have that:

$$\psi(D_{nm}(p * q, r * s)) = \psi(D_n(p, r)) + \psi(D_m(q, s)).$$

Consider $x_{p,r} \in (0,1)$ and $x_{q,s}$ to be defined as follows:

$$D_n(p,r) = D_2(e^1,(x_{p,r},1-x_{p,r})) \quad \text{and} \quad D_m(q,s) = D_2(e^1,(x_{q,s},1-x_{q,s})). \tag{A.23}$$

As proved earlier such $x_{p,r}$ and $x_{q,s}$ are unique. Let us now show that:

$$D_{nm}(p*q,r*s) = D_{nm}(e^1*e^1,(x_{p,r},1-x_{p,r})*(x_{q,s},1-x_{q,s})). \tag{A.24}$$

This follows directly from Property 1, as its first condition can be used for first with $D_n(p,r) \leq D_2(e^1,(x_{p,r},1-x_{p,r}))$, and then with $D_n(p,r) \geq D_2(e^1,(x_{p,r},1-x_{p,r}))$. This shows that:

$$D_{nm}(p*q,r*s) = D_{nm}(e^1*q,(x_{p,r},1-x_{p,r})*s),$$

and applying the same argument to $q, s$ we find equation (A.24).

But on pair of beliefs of the form $(e^1*e^1,(x_{p,r},1-x_{p,r})*(x_{q,s},1-x_{q,s}))$ we proved earlier that:

$$\begin{aligned}&\psi(D_4(e^1*e^1,(x_{p,r},1-x_{p,r})*(x_{q,s},1-x_{q,s})))\\ &\qquad\qquad = \psi(D_2(e^1,(x_{p,r},1-x_{p,r}))) + \psi(D_2(e^1,(x_{p,r},1-x_{p,r}))).\end{aligned}$$

Now using Property 1 we have that:

$$D_4(e^1*e^1,(x_{p,r},1-x_{p,r})*(x_{q,s},1-x_{q,s})) = D_{nm}(p*q,r*s),$$

and using equation (A.23) we obtain:

$$D_{nm}(p * q, r * s) = D_n(p, r) + D_m(q, s).$$

This concludes the first case, in which:

$$D_2(e_1, e_2) > D_2(e_1, (x, 1-x)), \quad \forall x \in (0, 1).$$

If instead we have that $D_2(e_1, e_2) = D_2(e_1, (x, 1-x))$ for some $x \in (0, 1)$, then we can consider a similar construction in which we define $\phi(z) := D_2((z, 1-z), (x, 1-x))$, for all $z \in (0, 1)$. A similar function $\psi$ can then be defined so that $\psi(D(\cdot, \cdot))$ is additive on all vectors $p, q, r, s \in \Delta_2$. And finally one can show that it must be additive on $\Delta_n$, analogously to what was done in the first case (as the only difference between these two cases is the references point – earlier it was $e^1$ and now it is $x$). $\square$

**Lemma 4.** *If $D_n$ satisfies Axioms 1–5, then $\forall p, q \in \Delta_n$:*

$$D_n(p, q) = D_{n+1}((p_1, \ldots, p_n, 0), (q_1, \ldots, q_n, 0)).$$

*Proof of Lemma 4.* By the coarsening axiom, Axiom 4, we have that for all $m \geq n$:

$$D_{m+1}(\bar{p}^{m+1}, \bar{q}^{m+1}) \geq D_m(\bar{p}^m, \bar{q}^m), \tag{A.25}$$

so the sequence $D_m(\bar{p}^m, \bar{q}^m)$ is non-decreasing in $m$.

By Axiom 5, we can construct the beliefs $p * (1, 0)$:

$$p * (1, 0) := (p_1, \ldots, p_n, 0, \ldots, 0) = \bar{p}^{2n}$$

and $q * (1, 0) = \bar{q}^{2n}$. By the independence axiom, Axiom 5, we have that:

$$D_{2n}(\bar{p}^{2n}, \bar{q}^{2n}) = D_{2n}(p * (1, 0), q * (1, 0)) = D_n(p, q) + D_2((1, 0), (1, 0)),$$

but $D_2((1, 0), (1, 0)) = 0$ by Axiom 1, so that we get $D_{2n}(\bar{p}^{2n}, \bar{q}^{2n}) = D_n(p, q)$, and by induction one can prove that:

$$D_{mn}(\bar{p}^{mn}, \bar{q}^{mn}) = D_n(p, q), \tag{A.26}$$

for all $m \in \mathbb{N}$. Hence we have that the sequence of numbers $D_m(\bar{p}^m, \bar{q}^m)$ is weakly increasing in $m$ (by equation (A.25)) and periodic (by equation (A.26)), but then it must be that it is a constant sequence, and hence we get the thesis. $\square$

**Lemma 5.** *Let $p, q \in \Delta_n$ and suppose that $Supp(p) \cap Supp(q) = \emptyset$, meaning $p_j q_j = 0$, for all $j$. Then if $D$ satisfies Axioms 1–5, $D(p, q) = +\infty$.*

*Proof of Lemma 5.* Notice that by Axiom 4 we have that for all $p, q \in \Delta_n$ with disjoint support:

$$D(p, q) \geq D((1, 0), (0, 1)),$$

simply by taking the partition of $\Theta$ consisting in $A_1 := Supp(p)$, $A_2 = \Theta \setminus Supp(p)$. Hence if we prove that $D((1, 0), (0, 1)) = +\infty$, the general statement follows.

Assume that $D((1,0),(0,1)) < \infty$, and consider the joint distributions:

$$(1,0) * (1,0) = (1,0,0,0) \quad \text{and} \quad (0,1) * (0,1) = (0,0,0,1).$$

By Lemma 4 we have that $D((1,0,0,0),(0,0,0,1)) = D((1,0),(0,1))$, and by Axiom 5, $D((1,0) * (1,0),(0,1) * (0,1)) = 2D((1,0),(0,1))$.

Hence:

$$D((1,0),(0,1)) = 2D((1,0),(0,1)),$$

then assuming $D((1,0),(0,1)) < \infty$ we obtain that $D((1,0),(0,1)) = 0$ - which contradicts the fact that $D(p,q) > 0$ for all $p \neq q$, so $D((1,0),(0,1)) = \infty$. □

**Proposition 6.** *Let $p, q \in \Delta_n^\circ$ be two beliefs, and suppose that $\frac{p_1}{q_1} = \frac{p_2}{q_2}$. Then:*

$$D_n(p,q) = D_{n-1}((p_1 + p_2, p_3, \ldots, p_n),(q_1 + q_2, q_3, \ldots, q_n)).$$

*Proof of Proposition 6.* Let's prove this by showing that both inequalities hold. Namely:

1. $D_n(p,q) \geq D_{n-1}((p_1 + p_2, p_3, \ldots, p_n),(q_1 + q_2, q_3, \ldots, q_n))$;
2. $D_n(p,q) \leq D_{n-1}((p_1 + p_2, p_3, \ldots, p_n),(q_1 + q_2, q_3, \ldots, q_n))$;

(1) follows directly from the coarsening axiom, Axiom 4:

$$D(p,q) \geq D((p_1 + p_2, p_3, \ldots, p_n),(q_1 + q_2, q_3, \ldots, q_n)).$$

In order to prove (2) notice that by the expandability Lemma (Lemma 4) we have that:

$$
\begin{aligned}
&D((p_1+p_2,p_3,\dots,p_n),(q_1+q_2,q_3,\dots,q_n)) \\
&\qquad = D((0,p_1+p_2,p_3,\dots,p_n),(0,q_1+q_2,q_3,\dots,q_n))
\end{aligned}
$$

and:

$$
\begin{aligned}
&D((p_1+p_2,p_3,\dots,p_n),(q_1+q_2,q_3,\dots,q_n)) \\
&\qquad = D((p_1+p_2,0,p_3,\dots,p_n),(q_1+q_2,0,q_3,\dots,q_n)).
\end{aligned}
$$

Notice also that:

$$p = \lambda(p_1+p_2,0,p_3,\dots,p_n) + (1-\lambda)(0,p_1+p_2,p_3,\dots,p_n),$$

with $\lambda = \frac{p_1}{p_1+p_2}$, and similarly:

$$q = \lambda'(q_1+q_2,0,q_3,\dots,q_n) + (1-\lambda')(0,q_1+q_2,q_3,\dots,q_n),$$

with $\lambda' = \frac{q_1}{q_1+q_2}$. But if $\frac{p_1}{p^2} = \frac{q^1}{q^2}$ then $\lambda = \lambda'$ and then we can use Axiom 3 to obtain that:

$$
\begin{aligned}
D(p,q) \leq \max\{&D((0,p_1+p_2,p_3,\dots,p_n),(0,q_1+q_2,q_3,\dots,q_n)), \\
&D((p_1+p_2,0,p_3,\dots,p_n),(q_1+q_2,0,q_3,\dots,q_n))\} = \\
&D((p_1+p_2,p_3,\dots,p_n),(q_1+q_2,q_3,\dots,q_n)), \quad \text{(A.27)}
\end{aligned}
$$

where the last equality follows from Lemma 4. This proves (2) and gives the thesis. □

**Corollary 3.** *Fix a belief* $p \in \Delta_n^\circ$ *and consider the segment joining* $q^1 = (q_1 + q_2, 0, q_3, \dots, q_n)$ *and* $q^2 = (0, q_1 + q_2, q_3, \dots, q_n)$*:*

$$[q^1, q^2] = \{r \in \Delta_n \mid r = \lambda q^1 + (1-\lambda)q^2, \lambda \in [0,1]\}.$$

*The minimum of the distance between* $p$ *and* $[q^1, q^2]$ *is reached at the* $q^* \in [q^1, q^2]$ *satisfying:*

$$\frac{q_1^*}{q_2^*} = \frac{p_1}{p_2},$$

*And for all* $r, r' \in [q^1, q^2]$ *we have that if* $r \in [r', q^*]$ *then*

$$D_n(p, r) \leq D_n(p, r').$$

*Proof of Corollary 3.* Thanks to Proposition 6 we have that if $\frac{p_1}{q_1^*} = \frac{p_2}{q_2^*}$ then:

$$D(p, q^*) = D((p_1 + p_2, \dots, p_n), (q_1 + q_2, \dots, q_n)),$$

and since $\forall r \in [q^1, q^2]$ we have that:

$$D(p, r) \geq D((p_1 + p_2, \dots, p_n), (q_1 + q_2, \dots, q_n)),$$

by Axiom 4, then we obtain that for all $r \in [q^1, q^2]$:

$$D(p, r) \geq D(p, q^*), \quad \forall r \in [q^1, q^2].$$

The latter part of the result is then easily obtained by the quasi-convexity of the function $f(q) := D(p, q)$. Namely, pick $r$ and $r'$ with $r \in [r', q^*]$, then by Axiom 3 we have that:

$$D(p, r) \leq \max\{D(p, r'), D(p, q^*)\},$$

but since $D(p, q^*) \leq D(p, r')$ we have that $D(p, r) \leq D(p, r')$. $\square$

## A.3 Proofs of Section 1.4

**Proposition 7.** *For any $p, q \in \Delta_n^\circ$ with $p \neq q$ and for all $D$ measures of disagreement there exists an experiment $\pi = (S, f(s|\theta))$ and a signal $s' \in S$ such that:*

$$D(p(s'), q(s')) < D(p, q).$$

*For any $D$ measure of disagreement, there exist $p, q \in \Delta_n$, an experiment $\pi = (S, f(s|\theta))$, and a signal $s \in S$ such that:*

$$D(p(s), q(s)) > D(p, q).$$

*Proof of Proposition 7.* The first statement is trivial, as we can pick $s'$ that fully reveals state $\theta_1$, $f(s'|\theta_1) = 1$ and $f(s'|\theta_j) = 0$ for all $j = 2, \ldots, n$. Since $p, q$ are fully mixed we have that $p(s') = q(s') = (1, 0, \ldots, 0)$, and hence:

$$D(p(s'), q(s')) = 0 < D(p, q),$$

because $p \neq q$.

To prove the second statement, notice that for any $p' \neq q'$ we can find a signal $s'$ such that $D(p'(s'), q(s')) < D(p', q')$. But then notice that we can define a signal $s$ by:

$$f(s|\theta) = \frac{k}{f(s'|\theta)}, \quad \forall\theta,$$

where $k$ is a constant that makes sure $f(s|\theta) < 1$. Now, defining $p := p'(s')$, $q := q'(s')$ we have that $p(s) = p'$ and $q(s) = q'$ and therefore:

$$D(p, q) = D(p'(s'), q(s')) < D(p', q') = D(p(s), q(s)),$$

which yields the thesis.

$\square$

**Theorem 2.** *Let $p$ be a degenerate distribution on the true state of the world, and let $q \in \Delta_n$. Then for all measures of disagreement, and for all $\tilde{\pi} \preceq \pi$ we have that:*

$$\mathbb{E}_p^{\pi}[D(p(s), q(s))] \leq \mathbb{E}_p^{\tilde{\pi}}[D(p(\tilde{s}), q(\tilde{s}))].$$

*Proof of Theorem 2.* Let us consider the characterization of the disagreement functions given by equation (1.3). If $z \geq 0.5$ the results follows from the proof of Theorem 3 below, so we do not repeat it here (notice how the statement of this Theorem is a particular case of point 3 in Theorem 3).

Now, let $z \leq -0.5$ and let $p = (1, 0, \dots, 0)$ without loss. Let $S$ be the signals of the experiment $\pi$ such that $f(s|\theta_1) > 0$ and $\tilde{S}$ those of $\tilde{\pi}$ with the same properties. If for all $s \in S$ we have that $f(s|\theta_j) = 0$ for all $j \geq 2$, then we have that:

$$\mathbb{E}_p^\pi[D(p(s), q(s))] = 0,$$

so the statement of the theorem is true. If instead there exists a signal $s$ such that $f(s|\theta_j) > 0$ for some $\theta_j \geq 2$, then there exists a $\tilde{s} \in \tilde{S}$ such that $g(\tilde{s}|\theta_j) > 0$ because:

$$g(\tilde{s}|\theta_j) = \sum_s \lambda_{s,\tilde{s}} f(s|\theta_j), \tag{A.28}$$

and $\sum_{\tilde{s}} \lambda_{s,\tilde{s}} = 1$. But then observe that upon observing such signal $\tilde{s}$ (which has positive probability according to $p$) we have that $q_j(\tilde{s}) > 0 = p_j(\tilde{s})$, and then $D^z(p(\tilde{s}), q(\tilde{s})) = \infty$ and therefore we obtain:

$$\mathbb{E}_p^{\tilde{\pi}}[D(p(\tilde{s}), q(\tilde{s}))] = \infty,$$

so the statement of the Theorem is trivially true.

The only case that we are left to analyze is $|z| < 0.5$, and this is the only non-trivial part of the proof. Notice that, for all $s \in S$, $p_1(s) = 1$ and then $D^z(p(s), q(s)) = -\log\left(q_1(s)^{-z+0.5}\right)$, and similarly for $\tilde{s}$, therefore we have to prove that:

$$-(0.5 - z)\sum_s f(s|\theta_1)\log(q_1(s)) \leq -(0.5 - z)\sum_{\tilde{s}} g(\tilde{s}|\theta_1)\log(q_1(\tilde{s})),$$

and since $-(0.5-z)<0$ for $|z|<0.5$ this is equivalent to

$$\sum_{\tilde{s}} g(\tilde{s}|\theta_1)\log(q_1(\tilde{s})) \le \sum_s f(s|\theta_1)\log(q_1(s))$$

Rewrite the RHS using the fact that, for all $s$, $\sum_{\tilde{s}} \lambda_{s,\tilde{s}} = 1$, that gives:

$$\sum_s f(s|\theta_1)\log(q_1(s)) = \sum_s \sum_{\tilde{s}} \lambda_{s,\tilde{s}} f(s|\theta_1)\log(q_1(s)) = \sum_{\tilde{s}} \sum_s \lambda_{s,\tilde{s}} f(s|\theta_1)\log(q_1(s)),$$

and now multiply and divide by $g(\tilde{s}|\theta_1)$ finding:

$$\begin{aligned}\sum_s f(s|\theta_1)\log(q_1(s)) &= \sum_{\tilde{s}} \sum_s \lambda_{s,\tilde{s}} f(s|\theta_1)\log(q_1(s)) \\ &= \sum_{\tilde{s}} g(\tilde{s}|\theta_1) \sum_s \frac{\lambda_{s,\tilde{s}} f(s|\theta_1)}{g(\tilde{s}|\theta_1)} \log(q_1(s)), \quad \text{(A.29)}\end{aligned}$$

now notice that the terms $\frac{\lambda_{s,\tilde{s}} f(s|\theta_1)}{g(\tilde{s}|\theta_1)}$ are probability distributions as $\sum_s \frac{\lambda_{s,\tilde{s}} f(s|\theta_1)}{g(\tilde{s}|\theta_1)} = 1$, and then using the fact that $-\log(\cdot)$ is convex we get that:

$$\begin{aligned}&\sum_s \frac{\lambda_{s,\tilde{s}} f(s|\theta_1)}{g(\tilde{s}|\theta_1)}(-\log(q_1(s)^{-1})) \\ &\qquad \ge -\log\left(\sum_s \frac{\lambda_{s,\tilde{s}} f(s|\theta_1)}{g(\tilde{s}|\theta_1)} \frac{f(s;q)}{q_1 f(s|\theta_1)}\right) = -\log\left(\frac{\sum_s \lambda_{s,\tilde{s}} f(s;q)}{q_1 g(\tilde{s}|\theta_1)}\right),\end{aligned}$$

and since $\sum_s \lambda_{s,\tilde{s}} f(s;q) = f(\tilde{s};q)$, we have that:

$$-\log\left(\frac{\sum_s \lambda_{s,\tilde{s}} f(s;q)}{q_1 g(\tilde{s}|\theta_1)}\right) = \log(q_1(\tilde{s})),$$

and plugging this into (A.29) we get:

$$\sum_s f(s|\theta_1)\log(q_1(s)) \geq \sum_{\tilde{s}} g(\tilde{s}|\theta_1)\log(q_1(\tilde{s})),$$

which is the thesis.

$\square$

**Lemma 6.** *$D$ satisfies Axioms 1–7 if and only if $D(p,q) = aD^z(p,q)$ for some $z \geq 0.5$, and $a > 0$.*

*Proof of Lemma 6.* Notice that for all $z \geq 0.5$, $D^z$ satisfies Axioms 7, so it is enough to show that for $z < 0.5$ there exists $p, q$ with $q_j = 0 < p_j$ such that $D(p,q) < +\infty$. To this extent, just consider $q = (1,0)$ and $p = (\alpha, 1-\alpha)$ and the result follows.

$\square$

**Theorem 3.** *Let $D$ be a measure of disagreement. Then the following statements are equivalent:*

1. *$D$ satisfies Axiom 7;*
2. *$D(p,q) = aD^z(p,q)$ for some $z \geq 0.5$ (and $a > 0$);*
3. *for all experiments $\pi \succeq \tilde{\pi}$ and priors $p, q \in \Delta_n$:*

$$\mathbb{E}_p^{\pi}[D(p(s), q(s))] \leq \mathbb{E}_p^{\tilde{\pi}}[D(p(\tilde{s}), q(\tilde{s}))].$$

*Proof of Theorem 3.* The equivalence of 1 and 2 was established in Lemma 6 above. To prove that 2$\Leftrightarrow$ 3 we will prove separatly that 2 implies 3 and vice versa:

(2 $\Rightarrow$ 3) First off, notice that we can assume without loss of generality that $p, q \in \Delta_n^\circ$ (if $q \notin \Delta_n^\circ$ then we have that on both sides disagreement will be infinite, so the statement is trivial; if $p \notin \Delta_n^\circ$ then disagreement would not change if we set $\Theta$ to be the support of $p$). For the rest of the proof, then, assume $p, q \in \Delta_n^\circ$.

We will use one of the characterizations of sufficiency proved by Blackwell and Girshick (1954):

**Result 9**. *We have that* $\tilde{\pi} \preceq \pi$ *if and only if for all* $\phi : \Delta_n \to \mathbb{R}$ *concave and all priors* $p \in \Delta_n$*:*

$$\sum_{s \in S} \mathbb{P}_p(s)\phi(p(s)) \leq \sum_{\tilde{s} \in \tilde{S}} \mathbb{P}_p(\tilde{s})\phi(p(\tilde{s})).$$

Therefore we are done if we prove that the functions $D_z(p(s), q(s))$ can be written as a concave function of $p(s)$ – for all $z \geq 0.5$. To do this, we apply the formula proved by Alonso and Camara (2016) that allows to write the posterior $q(s)$ as a function of $p(s)$ (and of both priors, $p$ and $q$).

$$q_j(s) = \frac{\frac{q_j}{p_j} p_j(s)}{\sum_i \frac{q_i}{p_i} p_i(s)} = \frac{r_j p_j(s)}{p(s) \cdot r}, \tag{A.30}$$

where we defined the likelihood ratio $r_j := \frac{q_j}{p_j}$ and denoted by $\cdot$ the scalar product in $\mathbb{R}^n$. Using this formula we get:

$$D_z(p(s),q(s)) = \log\left(\sum_j p_j(s)\left(\frac{p_j(s)}{q_j(s)}\right)^{z-0.5}\right)$$
$$= \log\left(\sum_j p_j(s)\left(\frac{p(s)\cdot r}{r_j}\right)^{z-0.5}\right)$$
$$= (z-0.5)\log\left(p(s)\cdot r\right) + \log(p(s)\cdot r^z), \quad \text{(A.31)}$$

where we defined $r_j^z := \frac{1}{r_j^{z-0.5}}$.

If $z > 0.5$, then, we have that $D_z(p(s),q(s))$ is a concave function of $p(s)$ because $r^z$ is strictly positive (so $\log(p(s)\cdot r^z)$ is concave), and $(z-0.5)\log\left(p(s)\cdot r\right)$ is also concave because $r \gg 0$ and $z-0.5>0$.

On the other hand, if $z=0.5$, we get:

$$D_{0.5}(p(s),q(s)) = \sum_j p_j(s)\log\left(\frac{p_j(s)}{q_j(s)}\right)$$
$$= \sum_j p_j(s)(\log(p(s)\cdot r) - \log(r_j)) = -p(s)\cdot\log(r) + \log(p(s)\cdot r), \quad \text{(A.32)}$$

where we defined the vector $\log(r) := (\log(r_1),\dots,\log(r_n))$.

Then we have that $D_{0.5}(p(s),q(s))$ is the sum of a linear and a concave function of $p(s)$, and then it is a concave function of $p(s)$.

($3 \Rightarrow 2$) To prove this implication, we divide the proof in two parts:

1. firstly, we prove that for all $z<0.5$, we can find $p,q$ and an experiment $\pi$ such that:

$$\mathbb{E}_p^\pi(D^z(p(s),q(s))) > D(p,q);$$

2. secondly, we show there exist $p, q$ and $\pi$ such that:

$$\mathbb{E}_p^\pi(aD^{-0.5}(p(s), q(s)) + bD^{0.5}(p(s), q(s))) > aD^{-0.5}(p, q) + bD^{0.5}(p, q))$$

These two steps show that the only measures of disagreement that satisfy 3 are the measures described in 2.

Fix a $z < 0.5$. Let $p \in \Delta_2^\circ$ and $q^n = (1/n, 1 - 1/n)$. Notice that for all $n$, $D^z(p, q) < D^z(p, (0, 1)) < +\infty$, let us show that we can find a sequence of experiments $\pi(\epsilon)$ such that as $\epsilon \to 0$, $\mathbb{E}_p^{\pi(\epsilon)}\left(D(p(s), q^n(s))\right) \to +\infty$ (provided we pick $n$ big enough). This will imply that 3 is violated (with $\tilde{\pi}$ being the null experiment). Pick $\pi(\epsilon)$ to be an experiment with two signals, and conditional probabilities defined as follows:

$$f(s_0|\theta_1) = 1 - \epsilon, \quad f(s_1|\theta_1) = \epsilon,$$

$$f(s_1|\theta_2) = \epsilon \quad f(s_0|\theta_2) = 1 - \epsilon.$$

Consider the posteriors after observing signal $s_1$:

$$p_2(s_0) = \frac{\epsilon p_2}{\epsilon p_2 + (1 - \epsilon)p_1}; \quad q_2(s_0) = \frac{\left(1 - \frac{1}{n}\right)\epsilon}{\left(1 - \frac{1}{n}\right)\epsilon + \frac{1}{n}(1 - \epsilon)}.$$

Notice that if we take $\epsilon = n^{-0.5}$, as $n \to +\infty$ we have that $p_2(s_0) \to 0$ and $q_2(s_0) \to 1$. Therefore as we let $n \to \infty$ we have that:

$$\mathbb{E}_p^{\pi(n^{-0.5})}\left(D(p(s), q(s))\right) \geq \mathbb{P}_p(s_0)D^z((1 - \delta, \delta), (\delta, 1 - \delta)), \tag{A.33}$$

for all $\delta \in (0,1)$. Since for all $n$ we have $\mathbb{P}_p(s_0) > 0$ and $D^z(p, q^n) < D^z(p, (0,1)) < \infty$ we obtain that there exists an $n$ such that:

$$\mathbb{E}_p^{\pi(n^{-0.5})}\left(D^z(p(s), q(s))\right) > D^z(p, q^n),$$

because we know that for all $z$, $D^z((1-\delta,\delta),(\delta,1-\delta)) \to \infty$ as $\delta \to 0$ and hence we can use equation (A.33) to conclude.

Now consider the case of $D(p,q) = aD^{-0.5}(p,q) + bD^{0.5}(p,q)$. Let $p = (0.5-\epsilon, \epsilon, 0.5)$ and $q = (0.5-\epsilon, \delta, 0.5+\epsilon-\delta)$. One can easily check that:

$$\begin{aligned} D(p,q) = a\left(\delta\log\left(\frac{\delta}{\epsilon}\right) + (0.5+\epsilon-\delta)\log\left(\frac{0.5+\epsilon-\delta}{0.5}\right)\right) \\ + b\left(\epsilon\log\left(\frac{\epsilon}{\delta}\right) + 0.5\log\left(\frac{0.5}{0.5+\epsilon-\delta}\right)\right). \end{aligned}$$

Let us prove that considering $\pi = (\{s_1, s_2\}, f(s|\theta))$ with $f(s_1|\theta_1) = f(s_1|\theta_2) = 1$ and $f(s_2|\theta_3) = 1$ we have that expected posterior disagreement exceeds the disagreement in the priors, for some values of $\epsilon, \delta$. The experiment $\pi$ essentially reveals whether the state of the world is $\theta_3$ or it is not. In other words, $p(s_2) = q(s_2) = (0,0,1)$ (for all $\epsilon, \delta$), whereas:

$$p(s_1) = \left(\frac{0.5-\epsilon}{0.5}, \frac{\epsilon}{0.5}, 0\right) \quad \text{and} \quad q(s_1) = \left(\frac{0.5-\epsilon}{0.5-\epsilon+\delta}, \frac{\delta}{0.5-\epsilon+\delta}, 0\right).$$

Also, $\mathbb{P}_p(s_1) = 0.5$ for all $\epsilon, \delta$, therefore expected posterior disagreement is:

$$0.5D(p(s_1), q(s_1)) = 0.5aD^{-0.5}(p(s_1), q(s_1)) + 0.5bD^{0.5}(p(s_1), q(s_1)).$$

Let us show that letting $\delta \to 0$ and $\epsilon \to 0.5$ we have that: $D^{0.5}(p, q) - 0.5D^{0.5}(p(s_1)q(s_1))$ remains bounded, whereas $D^{0.5}(p, q) - 0.5D^{0.5}(p(s_1)q(s_1)) \to -\infty$, which proves the thesis.

First notice that:

$$\begin{aligned} &D^{0.5}(p, q) - 0.5D^{0.5}(p(s_1)q(s_1)) \\ &\qquad = \epsilon \log\left(\frac{\epsilon}{\delta}\right) + (0.5 + \epsilon - \delta)\log\left(\frac{0.5 + \epsilon - \delta}{0.5}\right) \\ &- 0.5\left(\frac{0.5 - \epsilon}{0.5}\log\left(\frac{0.5}{0.5 - \epsilon + \delta}\right) + \frac{\epsilon}{0.5}\left(\log\left(\frac{\epsilon}{\delta}\right) + \log\left(\frac{0.5 - \epsilon + \delta}{\delta}\right)\right)\right) \\ &\qquad = -\delta\log\left(\frac{0.5 + \epsilon - \delta}{0.5}\right) - \epsilon\log\left(\frac{0.5 - \epsilon + \delta}{\delta}\right). \end{aligned} \tag{A.34}$$

Now, picking $\delta = \sqrt{0.5 - \epsilon}$ and letting $\epsilon \to 0.5$ we get that $D^{0.5}(p, q) - 0.5D^{0.5}(p(s_1)q(s_1))$ is bounded.

On the other hand, considering $D^{-0.5}(p,q) - 0.5D^{-0.5}(p(s_1)q(s_1))$ we find:

$$\begin{aligned}
&D^{-0.5}(p,q) - 0.5D^{-0.5}(p(s_1)q(s_1)) \\
&\qquad = \delta \log\left(\frac{\delta}{\epsilon}\right) + (0.5+\epsilon-\delta)\log\left(\frac{0.5+\epsilon-\delta}{0.5}\right) \\
&\qquad\quad - 0.5\Bigg(\frac{0.5-\epsilon}{0.5-\epsilon+\delta}\log\left(\frac{0.5-\epsilon+\delta}{\delta}\right) \\
&\qquad\quad + \frac{\delta}{0.5-\epsilon+\delta}\left(\log\left(\frac{\delta}{\epsilon}\right) + \log\left(\frac{\delta}{0.5-\epsilon+\delta}\right)\right)\Bigg). \qquad \text{(A.35)}
\end{aligned}$$

Notice that as $\epsilon \to 0.5$ and $\delta = \sqrt{0.5-\epsilon}$ we have that $D^{-0.5}(p,q) = \delta\log\left(\frac{\delta}{\epsilon}\right) + (0.5+\epsilon-\delta)\log\left(\frac{0.5+\epsilon-\delta}{0.5}\right)$ is bounded. On the other hand, plugging in $\delta = \sqrt{0.5-\epsilon}$ we find that $-0.5D(p(s_1), q(s_1)) \to -\infty$, which yields the thesis.

□

**Theorem 4.** *Let $p^1, \ldots, p^J \in \Delta_n^\circ$, and let $(u_j)_j$ be CRRA utility functions with parameters $z_j \in (0,1)$. The equilibrium of the economy exists and is unique, and it can be characterized as follows:*

- *$(\Pi_i)_i$ is the unique solution to the problem:*

$$\min_{q\in\Delta(I)} \sum_j \frac{1}{1-z_j} D^{z_j+0.5}(q, p^j(\cdot|I)), \qquad \text{(A.36)}$$

  *i.e. the prices of the Arrow Debreu securities are the beliefs that minimize weighted disagreement with the agents;*

- *the expected volume of trade is given by:*

$$V_j((x_i^j)_i) = \exp(D^{z_j+0.5}(p^j(\cdot|I),\Pi) + D^{-z_j+0.5}(\Pi, p^j(\cdot|I))) - 1 \quad \forall j \in J; \quad \text{(A.37)}$$

- *the gains from trade are given by:*

$$G_j((x_i^j)_i) = \frac{1}{\frac{1}{z_j}-1}\left(1-\exp\left(-\frac{D^{z_j-0.5}(\Pi, p^j(\cdot|I))}{z_j}\right)\right), \quad \forall j \in J. \qquad \text{(A.38)}$$

*Proof of Theorem 4.* To prove the first point, let us solve the maximization problem of agent $j$:

$$\max_{(x_i^j)_i} \sum_i p_i^j \frac{\left(x_i^j\right)^{1-\frac{1}{z_j}}}{1-\frac{1}{z_j}} \quad \text{subject to} \quad \sum_i \Pi_i x_i \leq 0,$$

this gives first order conditions (for the amount of good $x_i^j$):

$$p_i^j (x_i^j)^{-1/z_j} - \lambda \Pi_i = 0,$$

where $\lambda$ is the Lagrainge multiplier. Isolating $x_i^j$ we get:

$$x_i^j = \left(\frac{p_i^j}{\lambda \Pi_i}\right)^{z_j}, \qquad \text{(A.39)}$$

and in equilibrium the budget constraint binds, $\sum_i \Pi_i x_i^j = 1$, so that:

$$\lambda^{z_j} = \sum_i \Pi_i \left(\frac{p_i^j}{\Pi_i}\right)^{z_j},$$

so plugging this into (A.39) we get:

$$x_i^j = \frac{\left(\frac{p_i^j}{\Pi_i}\right)^{z_j}}{\sum_k \Pi_k \left(\frac{\Pi_k}{p_k^j}\right)^{-z_j}}, \tag{A.40}$$

so summing across agents $j$ we get:

$$1 = \sum_j x_{i,j} = \sum_j \frac{\left(\frac{p_i^j}{\Pi_i}\right)^{z_j}}{\sum_k \Pi_k \left(\frac{\Pi_k}{p_k^j}\right)^{-z_j}},$$

which in particular implies that the RHS does not depend on $i$ and we obtain (picking $i$ and $i+1$):

$$\sum_j \frac{\left(\frac{p_i^j}{\Pi_i}\right)^{z_j} - \left(\frac{p_{i+1}^j}{\Pi_{i+1}}\right)^{z_j}}{\sum_k \Pi_k \left(\frac{\Pi_k}{p_k^j}\right)^{-z_j}} = 0,$$

and notice that this can be rewritten as:

$$\sum_j \partial_i^1 \left( \frac{1}{1-z_j} D^{z_j+0.5}(\Pi, p^j(\cdot|I)) \right),$$

where $\partial_i^1 D$ stands for the derivative of $D$ on the first argument ($\Pi$, in this case) along the direction $z_{i,i+1}$. Thanks to the quasi-convexity of $D$ (Axiom 3) we have the first result of our theorem.

To obtain the second result, simply notice that using equation (A.40) we find:

$$V_j((x_i^j)_i) = \sum_i p_i^j x_i^j - 1 = \frac{\sum_i p_i^j \left(\frac{\Pi_i}{p_i^j}\right)^{-z_j}}{\sum_k \Pi_k \left(\frac{\Pi_k}{p_k^j}\right)^{-z_j}} - 1$$

$$= \exp(D^{z_j+0.5}(p^j(\cdot|I), \Pi) + D^{-z_j+0.5}(\Pi, p^j(\cdot|I))) - 1.$$

Finally, to get the formula for the gains from trade, notice that $u_j(1) = \frac{1}{1-\frac{1}{z_j}}$, and:

$$u_j(x_i^j) = \frac{1}{1-\frac{1}{z_j}} \left( \frac{\left(\frac{\Pi_i}{p_i^j}\right)^{-z_j}}{\sum_k \Pi_k \left(\frac{\Pi_k}{p_k^j}\right)^{-z_j}} \right)^{1-\frac{1}{z_j}}$$

$$\Rightarrow \sum_i p_i^j u(x_i^j) = \frac{1}{1-\frac{1}{z_j}} \left( \sum_i p_i^j \left(\frac{\Pi_i}{p_i^j}\right)^{1-z_j} \right) \cdot \left( \sum_k \Pi_k \left(\frac{\Pi_k}{p_k^j}\right)^{-z_j} \right)^{\frac{1}{z_j}-1},$$

so the gains from trade are:

$$G_j((x_i^j)_i) = \frac{1}{\frac{1}{z_j}-1} \left( 1 - \exp\left( -\frac{D^{z_j-0.5}(\Pi, p^j(\cdot|I))}{z_j} \right) \right), \quad \forall j \in J.$$

□

**Theorem 5.** *Let $(u_j)_{j\in J}$ be a family of strictly concave, twice continuously differentiable utility functions defined in a compact neighborhood of $1$. Suppose (without loss of generality) that $u_j'(1) = 1$, and assume that $-1/u_j''(1) =: z_j \in (0,1)$. Let $(p^{j,(m)})_{m\in\mathbb{N}}^{j\in J}$ be any family of merging and uniformly mixed beliefs.*

*For all $m$ the equilibrium exists and is unique, denote it by $((x_i^{j,(m)})_{i=1,\ldots,n}^{j\in J}, (\Pi_i^m)_i)$. For all $m$, let $\tilde{\Pi}^m$ be the solution of the problem $\min_{q\in\Delta_n^\circ} \sum_j \frac{1}{1-z_j} D^{z_j+0.5}(q, p^{j,(m)})$. We have* *that:*

- *for all* $j$, $\lim_{m\to\infty} \frac{D^{z_j+0.5}(p^{j,(m)},\Pi^m)}{D^{z_j+0.5}(p^{j,(m)},\tilde{\Pi}^m)} = 1$, *so the disagreement between any belief* $p^j$ *and the approximate equilibrium* $\tilde{\Pi}^m$ *is asymptotically equivalent to the disagreement with the market belief* $\Pi^m$.

- *The volume of trades is asymptotic to the volume of trades in the economy with CRRA utility functions:*

$$\lim_{m\to\infty} \frac{\sum_i p_i^{j,(m)}(x_i^{j,(m)} - 1)}{D^{z_j+0.5}(p^{j,(m)}, \tilde{\Pi}^m) + D^{-z_j+0.5}(\tilde{\Pi}^m, p^{j,(m)})} = 1. \tag{A.41}$$

- *The gains from trade are asymptotic to the volume of trades in the economy with CRRA utility functions:*

$$\lim_{m\to\infty} \frac{\sum_i p_i^{j,(m)} u_i(x_i^{j,(m)})}{\frac{1}{1-z_j} D^{z_j-0.5}(\tilde{\Pi}^m, p^{j,(m)})} = 1. \tag{A.42}$$

*Proof of Theorem 5.* The existence and uniqueness of the equilibrium is granted by the strictly concavity of the utility functions, and by the compactness of the domain. Notice also that for $m$ large enough such solution must be in the interior of the domain, because as $p^{j,(m)}$ merge, the equilibrium $x_i^{j,(m)} \to 1$, which is in the interior of the domain of $u_j$ (by hypothesis). In the rest of the proof we assume that first order conditions hold with equality, even though in general this will hold only eventually.

Let $\tilde{u}_j$ to be the CRRA approximation of $u_j$ at $1$, that is:

$$\tilde{u}_j(x) := \frac{x^{1-\frac{1}{z_j}}}{1-\frac{1}{z_j}},$$

where $z_j = -\frac{1}{u''(1)}$.

For a given $J$-tuple of beliefs $\mathbf{p} := ((p_i^1)_i, \ldots, (p_i^J)_i)$ denote by $((x_i^j(\mathbf{p}))_{i,j}, (\Pi_i(\mathbf{p}))_i, (\lambda_j(\mathbf{p}))_j)$ to be the solution of the model with beliefs $\mathbf{p}$ and utility functions $u_j$, where $x_i^j$ are the allocations, $\Pi_i$ are the prices, and $\lambda_j$ are the Lagrainge multipliers. Analogously, let $((\tilde{x}_i^j(\mathbf{p}))_{i,j}, (\tilde{\Pi}_i(\mathbf{p}))_i, (\tilde{\lambda}_j(\mathbf{p}))_j)$ be the solution to the problem with utility functions $\tilde{u}_j$.

It is a well-known result of smooth economies that the functions $\mathbf{p} \mapsto x_i^j(\mathbf{p})$ are local diffeomorphism (for all $j, i$) and so are $\mathbf{p} \mapsto \Pi_i(\mathbf{p})$ (for all $i$) and $\mathbf{p} \mapsto \lambda_j(\mathbf{p})$ (for all $j$). Furthermore, denoting by $\mathbf{p}^* = (p^*, \ldots, p^*)$ the vector of agreement beliefs we have that:

$$\partial_{\mathrm{v}} x_i^j(\mathbf{p}^*) = \partial_{\mathrm{v}} \tilde{x}_i^j(\mathbf{p}^*), \tag{A.43}$$

where we denote by $\partial_v$ the directional derivative in the direction $\mathbf{v} = (v^1, \ldots, v^J)$, where for all $j \in J$: $\sum_i v_i^j = 0$, so that $p^j + \epsilon v^j \in \Delta_n^\circ$ for $\epsilon$ small. The equalities in (A.43) follow from the fact that $((x_i^j(\mathbf{p}))_{i,j}, (\Pi_i(\mathbf{p}))_i, (\lambda_j(\mathbf{p}))_j)$ and $((\tilde{x}_i^j(\mathbf{p}))_{i,j}, (\tilde{\Pi}_i(\mathbf{p}))_i, (\tilde{\lambda}_j(\mathbf{p}))_j)$ are the solutions to a system of equations given by first order conditions (FOC), budget constraints (BC), and market clearing conditions (MC). Equations BC and MC are the same in the general economy (utility functions $u_j$) and in the CRRA approximation (utility functions $\tilde{u}_j$). On the other hand, FOC depend on the utility functions for general vectors of beliefs $\mathbf{p}$. Nonetheless, the FOC:

$$p_i^{j,(m)} u_j'(x_i^{j,(m)}) - \lambda_j^m \Pi^m = 0, \quad \text{and} \quad p_i^{j,(m)} \tilde{u}_j'(\tilde{x}_i^{j,(m)}) - \tilde{\lambda}_j^m \tilde{\Pi}^m = 0,$$

and similarly:

$$\partial_{\mathbf{v}}\Pi_i(\mathbf{p}^*) = \partial_{\mathbf{v}}\tilde{\Pi}_i(\mathbf{p}^*) \quad \text{and} \quad \partial_{\mathbf{v}}\lambda_j(\mathbf{p}^*) = \partial_{\mathbf{v}}\tilde{\lambda}_j(\mathbf{p}^*). \tag{A.44}$$

Now, for any sequence of beliefs $\mathbf{p}$ convergent to $\mathbf{p}^*$, let us prove that (for all $i, j$):

$$\lim_{\mathbf{p}\to\mathbf{p}^*} \frac{x_i^j(\mathbf{p}) - 1}{\tilde{x}_i^j(\mathbf{p}) - 1} = 1, \tag{A.45}$$

where the limit is taken in the usual norm on the product space $\Delta_n^\circ \times \cdots \times \Delta_n^\circ$. This result implies the statement of the theorem for any sequence of merging beliefs.

Notice that because of the first order conditions $x_i^j = (u')^{-1}\left(\frac{\lambda_j(\mathbf{p})\Pi_i(\mathbf{p})}{p_{i,j}}\right)$, and $\tilde{x}_i^j(\mathbf{p}) = \left(\frac{\tilde{\lambda}_j(\mathbf{p})\tilde{\Pi}_i(\mathbf{p})}{p_{i,j}}\right)^{-z_j}$, where in the latter we used the fact that $\tilde{u}_j$ is CRRA. Denote by $\phi(\mathbf{p}) := \frac{\lambda_j(\mathbf{p})\Pi_i(\mathbf{p})}{p_{i,j}}$, and similarly define $\tilde{\phi}(\mathbf{p})$. Thanks to the local results at $\mathbf{p}^*$ (equation (A.44)) we have that:

$$\partial_{\mathbf{v}}\phi(\mathbf{p}^*) = \partial_{\mathbf{v}}\tilde{\phi}(\mathbf{p}^*), \quad \forall \mathbf{v}.$$

The limit in equation (A.45) is of the form $0/0$ and because of the smoothness assumptions we can apply the multidimensional de l'Hopital rule. In other words, if we prove that for all $\mathbf{v}$:

$$\lim_{\mathbf{p}\to\mathbf{p}^*} \frac{\partial_{\mathbf{v}}(x_i^j(\mathbf{p}) - 1)}{\partial_{\mathbf{v}}(\tilde{x}_i^j(\mathbf{p}) - 1)} = 1, \tag{A.46}$$

then the limit in equation (A.45) follows. Rewriting the limit (A.46) we get:

$$\lim_{\mathbf{p}\to\mathbf{p}^*} \frac{-\frac{1}{u''((u')^{-1}(\phi(\mathbf{p})))}\partial_{\mathbf{v}}\phi(\mathbf{p})}{-z_j\tilde{\phi}(\mathbf{p})\partial_{\mathbf{v}}\tilde{\phi}(\mathbf{p})}, \tag{A.47}$$

and $\phi(\mathbf{p}) \to 1$, $\tilde{\phi}(\mathbf{p}) \to 1$, $\frac{1}{u''((u')^{-1}(\phi(\mathbf{p})))} = -z_j$, and $\partial_{\mathbf{v}}\phi(\mathbf{p}), \partial_{\mathbf{v}}\tilde{\phi}(\mathbf{p}) \to \partial_{\mathbf{v}}\phi(\mathbf{p}^*) \neq 0$.[73] This proves equation (A.45).

Summing on states and weighting by $p_i^j$ we get:

$$\lim_{\mathbf{p}\to\mathbf{p}^*} \frac{\sum_i x_i^j(\mathbf{p})p_i^j}{\sum_i \tilde{x}_i^j(\mathbf{p})p_i^j} = 1,$$

and the denominator is asymptotic to $D^{z_j+0.5}(p^j, \tilde{\Pi}(\mathbf{p})) + D^{-z_j+0.5}(\tilde{\Pi}(\mathbf{p}), p^j)$ as $\mathbf{p} \to \mathbf{p}^*$. Therefore equation (A.42) follows and (A.41), is obtained analogously (thanks to the continuity of $u_j$ and $\tilde{u}_j$).

To obtain equation disagreement between the prices $\Pi$ and the approximated prices $\tilde{\Pi}$ notice that:

$$\frac{p_i^j}{\Pi_i^j(\mathbf{p})} = \lambda_j(\mathbf{p})u'(x_i^j(\mathbf{p})) \quad \text{and} \quad \frac{p_i^j}{\tilde{\Pi}_i^j(\mathbf{p})} = \tilde{\lambda}_j(\mathbf{p})\tilde{u}'(\tilde{x}_i^j(\mathbf{p})),$$

so equation (A.45) implies that for all $i, j$:

$$\lim_{\mathbf{p}\to\mathbf{p}^*} \frac{\frac{p_i^j}{\Pi_i^j(\mathbf{p})}}{\frac{p_i^j}{\tilde{\Pi}_i^j(\mathbf{p})}} = 1,$$

and then the first statement of the theorem follows from the continuity of the logarithm and the sum. $\square$

[73] The fact that these differentials are not null follows from the fact that $\phi$ is a local diffeomorphism at $\mathbf{p}^*$.

## A.4 Renyi's Divergence Axiomatization

The first main difference between Renyi's approach and ours is that Renyi's postulates involve *generalized distributions*, that is finite dimensional vectors $(x_i)_i$ satisfying:

$$x_i \geq 0, \quad \text{and} \quad \sum_i x_i \leq 1.$$

We will denote the set of generalized distributions in dimension $n$ as $G_n$. Even though this appears as a small difference between Renyi's and our axiomatization, it will have important implications, because Renyi's axioms will apply to a much larger set than our (all vectors in $G_n$, not only on the vectors in $\Delta_n$). Let us introduce Renyi's postulates (with the numbering of Rényi (1961)):

(P6) $I(x|y)$ is unchanged if the elements of $x$ and $y$ are rearranged;

(P7) for all $n \in N$ and for all $x, y \in G_n$, $I(x,y) \geq 0$ if $x_i \geq y_i$ for all $i$; and $I(x|y) \leq 0$ if $x_i \leq y_i$ for all $i$;

(P8) $I(1, 1/2) = 1$;

(P9) for all $n, m \in \mathbb{N}$ and for any $x^1, y^1 \in G_n$ and $x^2, y^2 \in G_m$: $I(x^1 * x^2|y^1 * y^2) = I(x^1|y^1) + I(x^2|y^2)$;

(P10) for all $x \in G_n$ and $y \in G_m$, denote by $x \cup y := (x_1, \ldots, x_n, y_1, \ldots, y_m) \in G_{n+m}$. There exists strictly increasing function $g : \mathbb{R} \to \mathbb{R}$ such that if $x^1, x^2, y^1, y^2, x^1 \cup x^2$

and $y^1 \cup y^2$ are generalized distributions:[74]

$$I(x^1 \cup x^2 | y^1 \cup y_2) = g^{-1}\left(\frac{\left(\sum_i x_i^1\right) g(I(x^1|x^2)) + \left(\sum_j x_j^2\right) g(I(y^1|y^2))}{\sum_i x_i^1 + \sum_j x_j^2}\right)$$

The first postulate, $(P6)$, is equivalent to our Axiom 2 (this axiom is assumed by all the axiomatization mentioned in this literature review). $(P7)$ has few implications on our analysis since we consider probability distributions. It only implies our Axiom 1, because for proper probability distributions $x, y$, $x_i \geq y_i$ for all $i$ if and only if $x = y$, and then $(P7)$ implies that $I(p|p) = 0$. $P(8)$ is a normalization. $(P9)$ is very similar to our Axiom 5, but it is much stronger because it applies for all generalized distributions. In particular, considering $x^1, y^1, x^2, y^2 \in G_1 = [0, 1]$ it implies that:

$$I(x^1 x^2 | y^1 y^2) = I(x^1|y^1) + I(x^2|y^2).$$

This equation (and a smoothness condition on $I$) imply that for $x, y \in [0, 1]$, $I(x|y) = \beta \log(x/y)$, for some $\beta \in \mathbb{R}$. Therefore $(P9)$ alone implies that $I$ is the logarithm on generalized distributions in dimension 1 (i.e. real numbers $x, y \in [0, 1]$). Our Axiom 5, instead, apply only to actual distributions so does not have the implications of $(P9)$. Finally, $(P10)$ imposes a functional form on the divergence measures $I(x|y)$, and since

[74]Observe that for any $x \in G_n$ and $y \in G_m$, $x \cup y$ is a generalized distribution if and only if:

$$\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j \leq 1.$$

for $z, z' \in [0,1]$, $I(z|z') = \beta \log(z/z')$, $(P10)$ implies that for all $x, y \in G_n$:

$$I(x|y) = g^{-1}\left(\sum_j \frac{x_j}{\sum_i x_i} g(\alpha \log(x_j/y_j))\right),$$

by induction. The final part of Renyi's proof shows that the only increasing functions $g$ that make sure $I$ satisfies $(P9)$ are parametrized by $g(x) := \gamma e^{\alpha x}$. Finally, the normalization $(P8)$ imply that:

$$I(x|y) = \frac{1}{\alpha - 1} \log\left(\frac{\sum_j x_j \left(\frac{x_j}{y_j}\right)^{\alpha-1}}{\sum_j x_j}\right),$$

which boils down to the expression $R_\alpha(p,q)$ when applied to probability distributions:

$$R_\alpha(p,q) = \frac{1}{\alpha - 1} \log\left(\sum_j p_j \left(\frac{p_j}{q_j}\right)^{\alpha-1}\right).$$

Summarizing, the main differences between Renyi's axiomatization and Chapter 1 are the following: Rényi (1961) models generalized distributions, and this (with additivity) buys that the divergence of two events is given by the logarithm of the ratio of the probabilities.[75] In our model, we also find that disagreement measures involve the dispersion of the likelihood ratio, but we derive such functional form from other axioms (that are related to properties of disagreement). Secondly. Renyi's $(P10)$ constraints the di-

[75]Aczel and Daroczy (1975) provided an axiomatization of Renyi's divergences that does not use any generalized distributions, but Aczel and Daroczy (1975) constraints the divergence to be a weighted mean of $\log(p_i/q_i)$.

vergences to have the functional form of a generalized mean,[76] whereas we impose only local conditions the disagreement measures (Axiom 6). Thirdly, Renyi's postulate $(P10)$ is not symmetric in $x$ and $y$, meaning: if $(x,y) \mapsto I(x|y)$ satisfies it, $(x,y) \mapsto I(y|x)$ does not. In our case, instead if $(x,y) \mapsto D(x,y)$ satisfies our set of axioms, then so does $(x,y) \mapsto D(y,x)$.[77] This implies that the function $D(p,q) = \sum_j q_j \log\left(\frac{q_j}{p_j}\right)$ satisfies our axioms and is *not* a Renyi divergence. Finally, our characterization is based on basic principles that intuitively describe disagreement, whereas Renyi's motivation is generalizing the entropy and relative entropy.

[76]The generalized mean of $n$ real numbers $x_1, \ldots, x_n \in \mathbb{R}$ is defined as:

$$M_\phi^p(x_1, \ldots, x_n) = \phi^{-1}\left(\frac{\sum_j p_j \phi(x_j)}{n}\right),$$

for some $\phi(\cdot)$ increasing.

[77]This result is obvious for Axioms 1–5, while Axiom 6 is apparently asymmetric in $p$ and $q$. Nonetheless it is clear from our characterization that Axiom 6 holds for both $p$ and $q$.

Chapter B

# *Appendix of Chapter 2*

**Lemma 7** (Redundant signals). *Suppose that for two signals $s_1, s_2$ we have that:*

$$f(s_1|\theta)f(s_2|\theta') = f(s_1|\theta')f(s_2|\theta) \tag{B.1}$$

*for all $\theta, \theta'$. Then considering $\gamma : S \to S'$ such that $\gamma(s_1) = \gamma(s_2)$ and $\gamma(s) \neq \gamma(s')$ for any $s \neq s'$ ($s, s' \neq s_1, s_2$) we have that:*

$$c_n(\gamma \circ \pi) = c(\pi).$$

*Proof of Lemma 7.* Thanks to Postulate 4 we have that $c_n(\gamma \circ \pi) \leq c_n(\pi)$. Let us prove that if equation (B.1) holds then it is also true that $c_n(\gamma \circ \pi) \geq c_n(\pi)$. Equation (B.1) is equivalent to having that for some $k > 0$:

$$kf(s_1|\theta) = f(s_2|\theta) \quad \forall \theta \in \Theta,$$

and with this notation we have that:

$$\gamma \circ \pi = \left(\begin{pmatrix} (1+k)f(s_1|\theta_1) & \dots & (1+k)f(s_1|\theta_n) \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix}\right)$$

$$= \begin{pmatrix} \frac{1+k}{k}f(s_2|\theta_1) & \dots & \frac{1+k}{k}f(s_2|\theta_n) \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix},$$

and thanks to Postulate 2 we have that:

$$c_n(\gamma \circ \pi) = c_n \begin{pmatrix} (1+k)f(s_1|\theta_1) & \dots & (1+k)f(s_1|\theta_n) \\ 0 & \dots & 0 \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix} = c_n \begin{pmatrix} 0 & \dots & 0 \\ \frac{1+k}{k}f(s_2|\theta_1) & \dots & \frac{1+k}{k}f(s_2|\theta_n) \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix}, \qquad (B.2)$$

and notice that we can write $\pi$ as follows:

$$\pi = \frac{1}{k+1} \begin{pmatrix} (1+k)f(s_1|\theta_1) & \dots & (1+k)f(s_1|\theta_n) \\ 0 & \dots & 0 \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix} + \frac{k}{k+1} \begin{pmatrix} 0 & \dots & 0 \\ \frac{1+k}{k}f(s_2|\theta_1) & \dots & \frac{1+k}{k}f(s_2|\theta_n) \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix}.$$

Therefore we can apply Postulate 3 and find that:

$$c_n(\pi) \leq \max \Bigg\{ c_n \left(\begin{pmatrix} (1+k)f(s_1|\theta_1) & \dots & (1+k)f(s_1|\theta_n) \\ 0 & \dots & 0 \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix}\right),$$

$$c_n \left(\begin{pmatrix} 0 & \dots & 0 \\ \frac{1+k}{k}f(s_2|\theta_1) & \dots & \frac{1+k}{k}f(s_2|\theta_n) \\ f(s_3|\theta_1) & \dots & f(s_3|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix}\right) \Bigg\},$$

and using equation (B.2) we obtain that:

$$c_n(\pi) \leq c_n(\gamma \circ \pi),$$

which yields the thesis. $\square$

**Proposition 8.** *Let $\pi$ be sufficient for $\pi'$. If $c_n$ satisfies Postulates 1–4 then $c_n(\pi) \geq c_n(\pi')$.*

*Proof of Proposition 8.* If $\pi = (S, (f(s|\theta))_{s\in S,\theta\in\Theta})$ is sufficient for $\pi' = (S', (g(s'|\theta))_{s'\in S',\theta\in\Theta})$ we have that for some $\lambda_{s,s'} \geq 0$ with $\sum_{s'\in S'} \lambda_{s,s'} = 1$:

$$g(s'|\theta) = \sum_{s} \lambda_{s,s'} f(s|\theta). \tag{B.3}$$

Define $\bar{s}$ to be the largest $s$ such that $f(s|\theta) > 0$ (for some $\theta$) and $\bar{s}'$ the largest $s'$ such that $g(s'|\theta) > 0$.

Notice that thanks to Lemma 8 we can "split" the rows of $\pi$ without changing its cost:

$$c_n(\pi) = c_n\left(\begin{pmatrix} \lambda_{s_1,s'_1} f(s_1|\theta_1) & \dots & \lambda_{s_1,s'_1} f(s_1|\theta_n) \\ \lambda_{s_1,s'_2} f(s_1|\theta_1) & \dots & \lambda_{s_1,s'_2} f(s_1|\theta_n) \\ \dots & \dots & \dots \\ \lambda_{s_1,\bar{s}} f(s_1|\theta_1) & \dots & \lambda_{s_1,\bar{s}} f(s_1|\theta_n) \\ f(s_2|\theta_1) & \dots & f(s_2|\theta_n) \\ \dots & \dots & \dots \\ f(s_m|\theta_1) & \dots & f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix}\right),$$

where we used the fact that $\sum_{s'=1}^{\bar{s}'} \lambda_{s_1,s'} = 1$. We can do the same for all rows of $\pi$ obtaining:

$$c_n(\pi) = c_n\left(\begin{pmatrix} \lambda_{s_1,s'_1} f(s_1|\theta_1) & \dots & \lambda_{s_1,s'_1} f(s_1|\theta_n) \\ \lambda_{s_1,s'_2} f(s_1|\theta_1) & \dots & \lambda_{s_1,s'_2} f(s_1|\theta_n) \\ \dots & \dots & \dots \\ \lambda_{s_1,\bar{s}} f(s_1|\theta_1) & \dots & \lambda_{s_1,\bar{s}} f(s_1|\theta_n) \\ \lambda_{s_2,s'_1} f(s_2|\theta_1) & \dots & \lambda_{s_2,s'_1} f(s_2|\theta_n) \\ \dots & \dots & \dots \\ \lambda_{s_2,\bar{s}} f(s_2|\theta_1) & \dots & \lambda_{s_2,\bar{s}} f(s_2|\theta_n) \\ \dots & \dots & \dots \\ \lambda_{s_m,s'_1} f(s_m|\theta_1) & \dots & \lambda_{s_m,s'_1} f(s_m|\theta_n) \\ \dots & \dots & \dots \\ \lambda_{s_m,\bar{s}} f(s_m|\theta_1) & \dots & \lambda_{s_m,\bar{s}} f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix}\right),$$

Now we can apply Postulate 4 summing on rows with the same $s'$, and we obtain a weakly less costly experiment. Reordering the rows (i.e. using Postulate 2) we get:

$$c_n\left(\begin{pmatrix} \lambda_{s_1,s_1'}f(s_1|\theta_1) & \dots & \lambda_{s_1,s_1'}f(s_1|\theta_n) \\ \lambda_{s_1,s_2'}f(s_1|\theta_1) & \dots & \lambda_{s_1,s_2'}f(s_1|\theta_n) \\ \dots & \dots & \dots \\ \lambda_{s_1,\bar{s}}f(s_1|\theta_1) & \dots & \lambda_{s_1,\bar{s}}f(s_1|\theta_n) \\ \lambda_{s_2,s_1'}f(s_2|\theta_1) & \dots & \lambda_{s_2,s_1'}f(s_2|\theta_n) \\ \dots & \dots & \dots \\ \lambda_{s_2,\bar{s}}f(s_2|\theta_1) & \dots & \lambda_{s_2,\bar{s}}f(s_2|\theta_n) \\ \dots & \dots & \dots \\ \lambda_{s_m,s_1'}f(s_m|\theta_1) & \dots & \lambda_{s_m,s_1'}f(s_m|\theta_n) \\ \dots & \dots & \dots \\ \lambda_{s_m,\bar{s}}f(s_m|\theta_1) & \dots & \lambda_{s_m,\bar{s}}f(s_m|\theta_n) \\ \dots & \dots & \dots \end{pmatrix}\right)$$

$$\geq c_n\left(\begin{pmatrix} \sum_s \lambda_{s,s_1'}f(s|\theta_1) & \dots & \sum_s \lambda_{s,s_1'}f(s|\theta_n) \\ \sum_s \lambda_{s,s_2'}f(s|\theta_1) & \dots & \sum_s \lambda_{s,s_2'}f(s|\theta_n) \\ \dots & \dots & \dots \\ \sum_s \lambda_{s,\bar{s}'}f(s|\theta_1) & \dots & \sum_s \lambda_{s,\bar{s}}f(s|\theta_n) \end{pmatrix}\right),$$

and notice that the matrix on the right hand side corresponds to $\pi'$, by equation (B.3), therefore we found that $c_n(\pi) \geq c_n(\pi')$. □

**Proposition 9.** *Let $\phi : \mathscr{E}_n \to \mathbb{R} \cup \{+\infty\}$ be any function. If $\phi$ satisfies Postulate 7,*[78] *and $\pi$ is the fully informative experiment, then either $\phi(\pi) = 0$ or $\phi(\pi) = \infty$.*

*Proof of Proposition 9.* Observe that if $\pi$ is fully informative then the experiment $\tilde{\pi} := \pi \otimes \pi$ is also fully informative, therefore:

$$c_n(\pi) = c_n(\tilde{\pi}) = c_n(\pi \otimes \pi) = c_n(\pi) + c_n(\pi),$$

and we have that either $c_n(\pi) = 0$ or $c_n(\pi) = \infty$. □

**Corollary 4.** *Let $H : \Delta(\Theta) \to \mathbb{R}$ be any bounded function such that $H(\mu) > H(\delta_j)$ for all $\mu \in \Delta^\circ(\Theta)$ and $j = 1, \dots, |\Theta|$. For a fixed $\mu \in \Delta^\circ(\Theta)$, define the function $\phi_\mu : \mathscr{E}_n \to \mathbb{R}$*

[78] I.e. if:

$$\phi(\pi_1 \otimes \pi_2) = \phi(\pi_1) + \phi(\pi_2), \quad \forall \pi_1, \pi_2 \in \mathscr{E}_n.$$

*by:*

$$\phi_\mu(\pi) := |H(\mu) - \sum_{s\in S} \mathbb{P}_\mu(s) H(\mu(s))|.^{79}$$

$\phi_\mu(\pi)$ *does not satisfy Postulate 7.*

*Proof.* Observe that if $\pi$ is fully informative, then for all possible $s$ (i.e. for all $s$ with $\mathbb{P}_\mu(s) > 0$), $\mu(s) = \delta_j$ for some $j$. Therefore we obtain that $H(\mu) > \sum_{s\in S} \mathbb{P}_\mu(s) H(\mu(s))$, which implies that $\phi_\mu(\pi) > 0$ for $\pi$ fully informative. Then the thesis of Proposition 9 implies that $\phi_\mu(\pi) = \infty$, which contradicts the fact that $H$ is bounded. $\square$

**Theorem 6**. *For any* $\Theta$ *with* $|\Theta| > 3$ *and* $S$, $c_{\Theta,S}$ *satisfies Postulates 1–7 if and only if it is proportional to:*

$$\sum_s \sum_{i,j} (f(s|\theta_i) - f(s|\theta_j)) \log\left(\frac{f(s|\theta_i)}{f(s|\theta_j)}\right) = \sum_{i,j} D(f(\cdot|\theta_i), f(\cdot|\theta_j)),$$

*where* $D$ *is the symmetric Kullback-Leibler divergence, defined by* $D(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) + \sum_i q_i \log\left(\frac{q_i}{p_i}\right)$.

*Proof of Theorem 6.* Thanks to Lemma 8 we have that:

$$c_{\Theta,S} = \sum_{\theta,\theta'} c_2(\pi(\theta,\theta')),$$

Theorem 7 then implies that either

$$c_{\Theta,S}(\pi) = \sum_{\theta,\theta'} c_2^{KL}(\pi(\theta,\theta')),$$

[79] We write $\mathbb{P}_\mu(s)$ as a short-cut for $\sum_\theta \mu(\theta) f(s|\theta)$, i.e. the ex-ante probability that an agent with belief $\mu$ assigns to the signal $s$ being realized.

or:

$$c_{\Theta,S}(\pi) = \sum_{\theta,\theta'} c_2^B(\pi(\theta,\theta')),$$

but Lemma 9 implies that the second case does not satisfy Postulate 5 so we are left with only one possibility:

$$c_{\Theta,S}(\pi) = \sum_{\theta,\theta'} c_2^{KL}(\pi(\theta,\theta')),$$

which is the thesis. □

**Corollary 5.** *$c_{\Theta,S}(\pi) = \infty$ if and only if $\pi$ precludes a state, i.e. if $\exists s \in S, \theta, \theta' \in \Theta$ such that $f(s|\theta) > 0 = f(s|\theta')$.*

*Proof of Corollary 5.* We have that $c_{\Theta,S}(\pi) = 0$ if and only if $D(f(\cdot|\theta), f(\cdot|\theta')) = \infty$ for at least a pair of $\theta, \theta' \in \Theta$. But $D(f(\cdot|\theta), f(\cdot|\theta')) = \infty$ if and only if there exists (at least) a signal $s$ such that $f(s|\theta) > 0 = f(s|\theta')$, or $f(s|\theta) = 0 < f(s|\theta')$.

□

**Theorem 7.** *The only cost functions $c_2$ that satisfy postulates 1–5 and 7 are:*[80]

1. $c_2^{KL}(\pi) = \sum_s [f(s|\theta_1) - f(s|\theta_2)][\log(f(s|\theta_1)) - \log(f(s|\theta_2))]$;
2. $c_2^B(\pi) = -\log\left(\sum_s \sqrt{f(s|\theta_1)f(s|\theta_2)}\right)$.

*Proof of Theorem 7.* This result mirrors Proposition 1 in Chapter 1, as the axioms in this paper can be mapped into axioms of Chapter 1. The proof of Theorem 7 consists in showing that the hypothesis of Proposition 1 in Chapter 1 are met, and hence we can use its

[80]The subscript 2 refers to the fact that $|\Theta| = 2$; the superscript $KL$ and $B$ instead stand for Kullback-Leibler and Bhattacharyya respectively, as explained after the theorem.

thesis. As illustrated in the main body of the paper, the analogy is drawn by identifying the beliefs $p, q$ (of Chapter 1) with the conditional distributions $f(\cdot|\theta), f(\cdot|\theta')$ of this paper.

Postulate 1 implies Axiom 1 and the fact that $D(p, q) < \infty$ for any two fully mixed beliefs $p, q$ (that was another hypothesis of Chapter 1, not stated as a postulate). Postulate 2 implies Axiom 2 and the fact that $D$ is symmetric, $D(p, q) = D(q, p)$, which is an hypothesis of Proposition 1. Postulate 3 with $n = 2$ is exactly equivalent to Axiom 2. Postulate 4 is equivalent to Axiom 4, and it shows clearly that the states ($\theta$) in Chapter 1 are mapped to signals ($s$) in this paper. Postulate 5 is equivalent to separability (Property 2 in Chapter 1), which strengthens Axiom 3. Finally, Postulate 7 is equivalent to Axiom 5.

Given that each axiom in Chapter 1 can be obtained by assuming the axioms in this paper, and that furthermore Axiom 1 implies symmetry, we obtain Theorem 7. □

**Lemma 8**. *If a cost function $c_n$ satisfies Axioms 1,2 and 7 then $f_n$ (as defined in Axiom 6) must be the sum:*

$$f_n(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i, \quad \forall (x_i)_i \in [0, \infty]^n.$$

*Proof of Lemma 8.* This proof proceeds by induction. Consider first the case of $|\Theta| = 3$, we have that:

$$c_3(\pi) = c_2(\pi(\theta_1, \theta_2)) + f_2(c_2(\pi(\theta_3, \theta_2)), c_2(\pi(\theta_1, \theta_3))). \tag{B.4}$$

Using Axiom 2 we also have that:

$$c_3(\pi) = c_2(\pi(\theta_2, \theta_3)) + f_2(c_2(\pi(\theta_1, \theta_2)), c_2(\pi(\theta_1, \theta_3))),$$

so equating the two expressions we get that:

$$\begin{aligned} &c_2(\pi(\theta_1, \theta_2)) + f_2(c_2(\pi(\theta_3, \theta_2)), c_2(\pi(\theta_1, \theta_3))) \\ &\qquad = c_2(\pi(\theta_2, \theta_3)) + f_2(c_2(\pi(\theta_1, \theta_2)), c_2(\pi(\theta_1, \theta_3))). \end{aligned} \tag{B.5}$$

Now, for a fixed $c_2$, we can consider change the distribution $f(\cdot|\theta_3)$ such that: (i) $c_2(\pi(\theta_1, \theta_3))$ remains constant (there is an hyperplane of $\Delta(S)$ for which this happens, thanks to the smoothness of $c_2$); and (ii) $\partial c_2(\pi(\theta_1, \theta_3)) \neq 0$.[81] Considering this change and taking derivatives on both sides of (B.5) we get:[82]

$$\partial_x f_2(c_2(\pi(\theta_3, \theta_2)), c_2(\pi(\theta_1, \theta_3)))\partial c_2(\pi(\theta_1, \theta_3)) = \partial c_2(\pi(\theta_1, \theta_3)).$$

Since we assumed that $\partial c_2(\pi(\theta_1, \theta_3)) \neq 0$, this buys us that:

$$\partial_x f_2(c_2(\pi(\theta_3, \theta_2)), c_2(\pi(\theta_1, \theta_3))) = 1.$$

[81]We loosely refer to the change in the cost $c_2(\pi(\theta_1, \theta_3))$ under this transformation as $\partial c_2(\pi(\theta_1, \theta_3))$.

[82]$\partial_x f_2$ represents the derivative of $f_2$ with respect to the first coordinate.

Because of Axiom 2, we get the same result for $\partial_y f_2 = 1$, and therefore we get that:

$$f_2(x, y) = x + y + k,$$

where $k \in \mathbb{R}$ is a constant obtained by integrating the derivatives. Now notice that applying equation (B.4) to the non-informative experiment we get (using Axiom 1):

$$0 = c_3(\pi) = 0 + f(0, 0), \quad \Rightarrow \quad f(0, 0) = 0,$$

which concludes the proof.

$\square$

**Lemma 9.** *For any $n = |\Theta| > 2$, the function*

$$c_{\Theta,S}(\pi) = \sum_{\theta,\theta'} c_2^B(\pi(\theta, \theta')).$$

*does not satisfy Axiom 5.*

*Proof of Lemma 9.* Consider an experiment with $\Theta = \{\theta_1, \ldots, \theta_3\}$ and $f(s|\theta) > 0$ for $s = s_1, \ldots, s_6$ (while $f(s|\theta) = 0$ for $s = s_7, \ldots$). A consequence of Axiom 5 is that the relative change in $c$ due to an increase in $f(s_1|\theta_1)$ and a decrease in $f(s_2|\theta_1)$, divided by a similar change in $f(s_3|\theta_1)$ and $f(s_4|\theta_1)$ should be independent of signals $s_5, s_6$.[83]

[83]For a more detailed description of this property, see Chapter 1.

Nonetheless, notice that we can rewrite:

$$c_{\Theta,S}(\pi) = -\log\left(\prod_{\theta,\theta'}\left(\sum_s \sqrt{f(s|\theta)f(s|\theta')}\right)\right),$$

and it is clear that taking the derivatives described above we obtain that they depend on all the signals $s$ with $f(s|\theta) > 0$. As a matter of fact, the product of the sum would include all possible interaction terms between signals.

□

**Proposition 10.** *Let $\pi$ be the non-informative experiment. There exists a sequence of experiments $(\pi^n)_n$ such that $\pi^n \to \pi$ and $c(\pi^n) \to \infty$. On the other hand, if for some sequence $(\pi^n)_n$ we have that $c(\pi^n) \to 0$, then $\pi^n \to \pi$ non-informative.*

*Proof of Proposition 10.* Consider the uninformative experiment experiment $\pi = \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ \dots & \dots \end{pmatrix}$ and let $\pi^n$ be defined by

$$\pi^n = \begin{pmatrix} 1-\frac{1}{n} & 1 \\ \frac{1}{n} & 0 \\ \dots & \dots \end{pmatrix}.$$

It is clear that $\pi^n \to \pi$ and $c(\pi^n) = \infty$ for all $n$, so in particular $c(\pi^n) \to \infty$.

To show that if $c(\pi^n) \to 0$ then $\pi^n$ converges to the non-informative experiment, observe that:

$$c(\pi^n) \geq (f^n(s|\theta_i) - f^n(s|\theta_j)) \log\left(\frac{f^n(s|\theta_i)}{f^n(s|\theta_j)}\right) \geq 0$$

for all fixed $s, \theta_i, \theta_j$, and in order for the latter to converge to $0$ it must be that:

$$f^n(s|\theta_i) - f^n(s|\theta_j) \to 0,$$

which implies that $|f^n(s|\theta_i) - f^n(s|\theta_j)| \to 0$ (for all $s, \theta_i, \theta_j$). This allows us to conclude using equation (2.4), which defines convergence in norm. $\square$

**Lemma 10**. *For any $\mu \in \Delta(\Theta)$ with $\mu(\theta) > 0$ for all $\theta \in \Theta$, the cost of an experiment $\pi$ that does not preclude any state can be written as:*

$$c(\pi) = \sum_s \mathbb{P}_\mu(s) \sum_\theta \frac{\mu(\theta|s)n}{\mu(\theta)} \left(\log(\mu(\theta|s) - \log[a(\mu(s))]\right)) =: \sum_s \mathbb{P}_\mu(s)\phi_I(\mu(s);\mu),$$

*where we defined $\phi_I(\mu(s);\mu) := \sum_\theta \frac{\mu(\theta|s)n}{\mu(\theta)} \left(\log(\mu(\theta|s) - \log[a(\mu(s))]\right))$.*

*Proof of Lemma 10.* For brevity, denote by $\mu_j(s) := \mu(\theta_j|s)$ and $\mu_j := \mu(\theta_j)$. Since $f(s|\theta_j) = \frac{\mu_j(s)\mathbb{P}_\mu(s)}{\mu_j}$ for any $\mu \in \Delta(\Theta)$, we have that:

$$\begin{aligned} c(\pi) &= 2\sum_s \sum_{i,j} f(s|\theta_i)(\log(f(s|\theta_i)) - \log(f(s|\theta_j))) \\ &= 2\sum_s \sum_{i,j} \frac{\mu_i(s)\mathbb{P}_\mu(s)}{\mu_i}(\log(\mu_i(s)) - \log(\mu_j(s)) - \log(\mu_i) + \log(\mu_j)) \\ &= \sum_s \sum_{i,j} \frac{\mu_i(s)\mathbb{P}_\mu(s)}{\mu_i}(\log(\mu_i(s)) - \log(\mu_j(s))) + \sum_s \sum_{i,j}(\log(\mu_j) - \log(\mu_i)) \\ &= \sum_s \mathbb{P}_\mu(s) \sum_{i,j} \frac{\mu_i(s)}{\mu_i}[\log(\mu_i(s)) - \log(\mu_j(s))], \end{aligned} \quad \text{(B.6)}$$

and isolating the sum on $i$ and $j$ we obtain:

$$\begin{aligned}\phi_I(\mu(s);\mu) = \sum_{i,j} \frac{\mu_i(s)}{\mu_i}[\log(\mu_i(s)) - \log(\mu_j(s))]\\ = \sum_i \frac{\mu_i(s)}{\mu_i}\sum_j[\log(\mu_i(s)) - \log(\mu_j(s))]\\ = \sum_i \frac{\mu_i(s)n}{\mu_i}[\log(\mu_i(s)) - \log(a(\mu(s))))],\end{aligned} \tag{B.7}$$

where $a(\mu(s))$ is defined as the geometric mean of $\mu(s)$:

$$a(\mu(s)) = \sqrt[n]{\mu_1(s)\cdots\mu_n(s)},$$

so plugging this into (B.6) we obtain:

$$c(\pi) = \sum_s \mathbb{P}_\mu(s)\sum_i \frac{\mu_i(s)n}{\mu_i}[\log(\mu_i(s)) - \log(a(\mu(s)))].$$

$\square$

**Lemma 11**. *For all $\mu \in \Delta(\Theta)$, we have that:*

- $\phi_H(\mu;\mu) = \phi_I(\mu;\mu) = 0$*;*
- $\phi_H(\cdot;\mu)$ *and* $\phi_I(\cdot;\mu)$ *are convex function;*
- $\lim_{x\to\partial\Delta(\Theta)} \phi_H(x;\mu) < \infty$ *whereas* $\lim_{x\to\partial\Delta(\Theta)} \phi_I(x;\mu) = \infty$.[84]

[84]The set $\partial\Delta(\Theta)$ denotes the boundary of $\Delta(\Theta)$, formally:

$$\partial\Delta(\Theta) = \{\mu \in \Delta(\Theta) \mid \mu(\theta), \exists\theta\}$$

*Proof of Lemma 11.* The first equality follows by direct computation. The second condition follows from the fact that both cost functions are increasing in Blackwell informativeness, and then they must be convex function of the posterior (see Result 9 in Chapter 1). Finally the last condition can be proved using the continuity of $H$ to show that $\Phi_H(x; \mu)$ be bounded in $x$; and by direct computation for $\phi_I$. $\square$

Chapter C

# *Appendix of Chapter 3*

## C.1 Non-Manipulability

In the marriage model with finitely many agents, we had the following result (Theorem 4.6 from Roth and Sotomayor (1992)):

> When any stable mechanism is applied to a marriage market in which preferences are strict and there is more than one stable matching, then at least one agent can profitably misrepresent his or her preferences, assuming the others tell the truth.

On the other hand, with countably many agents the failure of the Rural Hospital Theorem implies the failure of that results - as we show by mean of an example.

Consider the preferences of Example 3, and let the stable mechanism be given by the women proposing Deferred Acceptance algorithm. Under such algorithm if every agent reports truthfully, the outcome is the matching $\mu_W$ defined by:

$$\mu_W(w_i) = m_{i+1} \quad \text{and} \quad \mu_W(m_1) = m_1.$$

Notice that no woman has strict incentives to misreport, as she is already matched to the best agent according to her preferences. Also, each man has no incentive to misreport:

- if man $m_1$ manipulates his preferences, he remains single, as no woman finds him acceptable - except for $w_1$, who is matched to $m_2$;
- if man $m_j$ ($j \geq 2$) manipulates his preferences then:
  - if he truncates his preferences to his best choice, $w_j$, then he remains single;
  - if he adds further women to the list of preferences or shuffles them, he remains matched to $w_{j-1}$ - because she is the only one proposing to him.

## C.2 Proofs

Most of the proofs in this Appendix are adapted from Roth and Sotomayor (1992), where the authors provide the same results for the finite marriage markets. Some of those proofs rely uniquely on the existence and definition of blocking pair, and then can be applied directly to our setting too.

**Lemma 12.** *If the preferences $P(x)$ of an agent $x$ do not satisfy Assumption 1 then there exist a market $(M, W, \boldsymbol{P})$ in which there are no stable matching.*

*Proof of Lemma 12.* Suppose that the preference of an agent does not satisfy Assumption 1, without loss let such agent be a woman $w \in W$. This means there exist a subset $\tilde{M} \subseteq M$ of *acceptable men* on which $>_w$ admits no maximum. Mathematically:

$$\forall m \in \tilde{M},\ \exists m' \in \tilde{M} \quad \text{s.t.} \quad m' >_w m, \tag{C.1}$$

and $m >_w w$, for all $m \in \tilde{M}$.

Now construct the set of preferences for men in the following way:

(Condition 1) if $m \in M \setminus \tilde{M}$, let $m >_m w$, - i.e. all men in $M \setminus \tilde{M}$ find $w$ unacceptable;

(Condition 2) if $\tilde{m} \in \tilde{M}$ let $w \geq_{\tilde{m}} w'$ for all $w' \in W$ - i.e. all men in $\tilde{M}$ have $w$ as their best choice;

Let us show that if preferences are so defined, there exists no stable matching. If a matching $\mu$ leaves $w$ single, it is clearly not stable as taking any $\tilde{m} \in \tilde{M}$ we have that $(\tilde{m}, w)$ constitutes as blocking pair. If instead $w$ is matched to a man $m$, then if $m \in M \setminus \tilde{M}$ the matching is not individually rational, because of Condition 1; if $m \in \tilde{M}$ then by mean of property (C.1) we must have that there exists a $m' \in \tilde{M}$ such that $m' >_w m$. But then notice that $(m', w)$ constitute a blocking pair, because $m'$ ranks $w$ as his first choice, because of Condition 2. $\square$

**Theorem 8**. *The $\mu_M$ so defined is a stable matching.*

*Proof of Theorem 8.* First off, let us prove that such function is in fact a matching: we need to prove that either a man is single, or he is matched to a woman (for the women it is trivial). If $m$ is not single, then $\mu_M(m) = x \neq m$. But then, according to the algorithm, it means that $(\mu^{(j)}(m))_j$ is eventually $x$, and we know that $\mu^{(j)}(m) \in W \cup \{m\}$ for all $j$, so $x \in W$. Identically, if a woman is not single under $\mu_M$, it means that $\mu^{(j)}(w) \in M$ for infinitely many $j$.

Also, it cannot be the case that $\mu_M(m) = \mu_M(m') \in W$, for some $m \neq m'$. If this was the case, then $\mu^{(j)}(m)$ and $\mu^{(j)}(m')$ eventually agree, so in particular $\exists K \in \mathbb{N}$ such that for all $k' \geq K$, $\mu^{(k')}(m) = \mu^{(k')}(m')$. But this can never happen in the algorithm defined, as in step $K$ the woman $w = \mu^{(K)}(m)$ would have rejected either $m$ or $m'$, who then cannot re propose to her. Therefore, we get an absurd.

Let us prove that such a matching is also stable. In order to prove it, observe first that:

$$\mu^{(j)}(w) \geq_w \mu^{(l)}(w), \quad \forall j \geq l, \tag{C.2}$$

which means that as the algorithm proceeds each woman can only do better (according to her preferences). This follows directly from the fact that the women withhold the best proposal they got so far, so either $\mu^{(j+1)}(w) = \mu^{(j)}(w)$ (if they reject the new proposal they got, if any), or $\mu^{(j+1)}(w) >_w \mu^{(j)}(w)$ (if they decide to reject the previous pick and withhold the new proposal).

Now, let us prove that the matching obtained is stable:

- no agent is matched with an unacceptable agent: this is clear on the side of the men, since they only propose to acceptable women; and similarly easy for women, as they accept only proposal by acceptable men;
- no pair blocks the matching: suppose there exists a pair $(m, w)$ such that $w >_m \mu_M(m)$. Then according to the algorithm, we would have that $m$ proposed to $w$ at some step $j$, and at some step $j' \geq j$, $w$ rejected $m$ in favor, say, of $m'$. But then, using equation (C.2) we have that:

  $$\mu_M(w) >_w m,$$

  so the pair $(m, w)$ could not block the matching.

□

**Theorem 9** (Optimality of $\mu_M$ (under $>_M$)). *For any $\mu$ stable, we have that $\mu_M \geq_M \mu$.*

*Proof of Theorem 9.* In order to prove such theorem, it is convenient to introduce an expression we will use frequently.

**Definition 20** (Achievable Matching). A woman $w$ is achievable by a man $m$ in a marriage market $(M, W, \mathbf{P})$ if there exists $\mu$ stable such that $\mu(m) = w$.

Now observe that the statement of the theorem is equivalent to saying that: *along the man proposing Deferred Acceptance algorithm no man is rejected by an achievable woman.* As a matter of fact, if this is the case, it means that for any $m$, $\mu_M(m) \geq \mu(m)$, for any $\mu$ stable, which is the thesis.

We will prove the theorem by induction on the steps of the algorithm. Suppose that in the first $j$ steps of the algorithm no man has been rejected by an achievable woman, we argue by contradiction supposing that in the step $j+1$ some man is rejected by an achievable woman. Suppose that this achievable woman $w$ rejects $m$ in favor of $m'$: this implies that:

$$m' >_w m. \tag{C.3}$$

Now, since $m'$ was not rejected by any achievable woman (by the induction hypothesis), we must have that:

$$w >_{m'} \mu(m'), \quad \forall \mu \text{ stable}. \tag{C.4}$$

Now consider $\tilde{\mu}$ to be the stable matching that matches $m$ to $w$ (such matching must exist, since by assumption $w$ is achievable by $m$). Now, notice that by equations (C.3) and

(C.4) we have that $(m', w)$ is a blocking pair for $\tilde{\mu}$ (as (C.4) holds *for any* stable matching), thus we get the contradiction that implies the thesis. $\square$

**Lemma 13.** *Let $\mu$ and $\mu'$ be stable matching of the marriage market $(M, W, \boldsymbol{P})$, we have that:*

$$\mu >_M \mu' \Leftrightarrow \mu' >_W \mu.$$

*Proof of Lemma 13.* Observe that the proof of $\Rightarrow$ and $\Leftarrow$ are symmetric, therefore it is enough to prove that for any $\mu$ and $\mu'$ stable:

$$\mu >_M \mu' \Rightarrow \mu' >_W \mu.$$

Let us argue by contradiction: suppose that $\mu >_M \mu'$, but for a woman $w$:

$$\mu(w) >_w \mu'(w). \tag{C.5}$$

This implies that under $\mu$, $w$ is matched to a man (otherwise she would be indifferent between the two outcomes, as $\mu(w) = w = \mu'(w)$). Let $m = \mu(w)$. Observe that since $\mu >_M \mu'$ we also have, for man $m$:

$$\mu(m) \geq_m \mu'(m),$$

and actually even more, $\mu(m) >_m \mu'(m)$, as if this was not the case, then $\mu'(m) = \mu(m) = w$. But being $\mu$ and $\mu'$ matching we would then have that $\mu'(w) = m = \mu(w)$, thus contradicting equation (C.5) - which is strict, so it rules out indifference. Therefore we can conclude that:

$$\mu(m) >_m \mu'(m). \tag{C.6}$$

But then notice that using equations (C.5) and (C.6) (and the fact that $\mu(w) = m$) we have that $(m, w)$ constitutes a blocking pair for $\mu'$, hence $\mu'$ is not stable, which contradicts the hypothesis of $\mu$ and $\mu'$ being both stable. □

**Corollary 6.** *For all $\mu$ stable we have that:*

$$\mu_M \geq_M \mu \geq_M \mu_W,$$

*so in particular $\mu_W$ is man pessimal.*

*Proof of Corollary 6.* The first inequality follows from Theorem 9 for men, while the second inequality follows with the same Theorem for women, jointly with Lemma 13. □

**Theorem 10.** *Let $\mu$ and $\nu$ be two stable matching on $(M, W, \boldsymbol{P})$, the functions $\vee$ and $\wedge$ defined in* (3.3) *and* (3.4) *yield stable matching.*

*Proof of Theorem 10.* Since the proofs are identical, we will only prove that $\lambda := \mu \vee \nu$ is a stable matching. The fact that $\lambda$ is the least upper bound is trivial - given it's definition - hence we have to prove:

1. $\lambda$ is a matching, i.e. if $\lambda(m) = w$, then $\lambda(w) = m$;[85]
2. $\lambda$ is stable;

Let us prove both statements by contradiction:

[85]The further property that if $\lambda(m) \neq m$, then $\lambda(m) \in W$ is trivial.

1. suppose that $\lambda(m) = w$ but $\lambda(w) \neq m$. Suppose furthermore (wlog) that $\lambda(m) = \mu(m)$, which implies that:

$$w = \mu(m) >_m \nu(m).$$

   If $\lambda(w) \neq m$, it means $\lambda(w) = \nu(w)$, and then:

$$m = \mu(w) >_w \nu(w).$$

   But then notice that $\nu$ is instable because $(m, w)$ are a blocking pair.

2. now to prove that the matching $\lambda$ is stable, notice that:

   - individual rationality is trivial - given the individual rationality of $\mu$ and $\nu$;
   - for blocking pairs, suppose there is one. I.e. suppose that:

$$w >_m \lambda(m) \tag{C.7}$$

$$m >_w \lambda(w) \tag{C.8}$$

   suppose wlog that $\lambda(m) = \mu(m)$ and $\lambda(w) = \nu(w)$.[86] This in turn buys us that:

$$\lambda(m) = \mu(m) >_m \nu(m) \tag{C.9}$$

$$\mu(w) >_w \nu(w) = \lambda(w) \tag{C.10}$$

[86]Clearly it cannot be that $\lambda(m) = \mu(m)$ and $\lambda(w) = \mu(w)$ otherwise $\mu$ would be unstable.

Observe then that by (C.9) and (C.7):

$$w >_m \lambda(m) >_m \nu(m),$$

we are done if we prove that $m >_w \nu(w)$, but notice that:

$$m >_w \lambda(w) = \nu(w),$$

by (C.10) - and hence $\nu$ has a blocking pair too, absurd.

$\square$

**Lemma 14.** *Let $\mu$ and $\mu'$ be stable matching. If $\mu \leq_M \mu'$ then:*

$$M(\mu) \subseteq M(\mu').$$

*In particular, for any $\mu$ we have that:*

$$M(\mu_W) \subseteq M(\mu) \subseteq M(\mu_M);$$

$$W(\mu_M) \subseteq W(\mu) \subseteq W(\mu_W).$$

*Proof of Lemma 14.* Notice that if $m \in M(\mu)$ and $\mu \leq_M \mu'$ then $m's$ match under $\mu'$ must be at least as good as that under $\mu$, i.e. $\mu(m) \leq_m \mu'(m)$. Since $\mu(m) \in W$, by individual rationality of $\mu$ we must have that $\mu'(m) \in W$, too, hence $m \in M(\mu')$. We then obtain that $M(\mu) \subseteq M(\mu')$, and the second part of the statement follows directly. $\square$

**Proposition 11**. *The Rural Hospital Theorem holds if and only if:*[87]

$$M = \mu_W \circ \mu_M(M).$$

*Proof of Proposition 11.* Suppose the Rural Hospital theorem does not hold. Then either $M(\mu_W) \subsetneq M(\mu_M)$, or $W(\mu_M) \subsetneq W(\mu_W)$. In the first case pick $m \in M(\mu_M) \setminus M(\mu_W)$: clearly $m \in M$, but $m \notin \mu_W \circ \mu_M(M)$, as $\mu_M(m) \in W$, and $m \notin \mu_W(W)$ (as $\mu_W(m) = m$ by hypothesis). In the second scenario, there exists a woman $w \in W(\mu_W) \setminus W(\mu_M)$. Pick $m := \mu_W(w)$: let us prove that $m \in M \setminus \mu_W \circ \mu_M(M)$. We have that: $\mu_M$ maps $\mu_M(M)$ into $W \setminus \{w\}$, but then $m \notin \mu_W(W \setminus \{w\})$, thus $m \in M \setminus \mu_W \circ \mu_M(M)$.

Now, for the opposite implication, suppose the Rural Hospital theorem holds, then for every $m \in M(\mu_M)$, we have that $m \in M(\mu_W)$ (and similarly for women). Let us prove that then $M = \mu_W \circ \mu_M(M)$. Trivially, $M \supset \mu_W \circ \mu_M(M)$, as:

- if $m$ is single under $\mu_M$, then so he is under $\mu_W$, hence $\mu_W(\mu_M(m)) = m$;
- if $m$ is matched to $w$ under $\mu_M$, then $w$ is matched to a man under $\mu_W$, hence $\mu_W(\mu_M(m)) \in M$.

Now, suppose by contradiction that $\exists m \in M \setminus \mu_W \circ \mu_M(M)$. Clearly, $m$ cannot be single under $\mu_M$, as if he were, then he would be single also under $\mu_W$ (because $\mu_M(m) \geq_m \mu_W(m)$ - and $\mu_W$ is individually rational), then $\mu_W \circ \mu_M(m) = m \Rightarrow m \in \mu_W \circ \mu_M(M)$.

If $m$ is not single under $\mu_M$, then he is neither under $\mu_W$ (by the Rural Hospital The-

[87]We write $f \circ g$ to mean the composition map: $x \mapsto f(g(x))$, and as always for any set $A$, $f(A) = \{y \mid y = f(x), \exists x \in A\}$.

orem), and then $\exists w \in W$ such that

$$\mu_W(w) = m. \tag{C.11}$$

This woman is matched under $\mu_W$, and therefore she is also matched by $\mu_M$, and therefore $\exists \tilde{m} \in M$ such that

$$\mu_M(\tilde{m}) = w. \tag{C.12}$$

Observe then that by equations (C.11) and (C.12):

$$\mu_W \circ \mu_M(\tilde{m}) = m \Rightarrow m \in \mu_W \circ \mu_M(M),$$

which contradicts the hypothesis and then proves the theorem. □

**Corollary** 7. *The Rural Hospital holds if one of the following conditions holds:*

- $|M(\mu)| < +\infty$ *for some* $\mu$ *stable;*
- *there exists a partition of* $M = \bigcup_j M_j$ *and* $W = \bigcup_j W_j$*, such that for all* $j$ $|M_j|, |W_j| < +\infty$*, and for all* $m \in M_j$*,* $m$ *finds acceptable only the women in* $W_j$*, mathematically:*

$$m >_m w, \quad \forall m \in M_j, w \in W \setminus W_j.$$

*Proof of Corollary 7.* The proof of the first point is trivial, given Lemma 14. The second point is then obvious, because the infinite problem boils down to a sequence of independent finite problems, and for finite problems the Rural Hospital theorem holds. □

**Theorem 11**. *The set of man maximal matching is a non-empty sublattice of the set of sta-*

*ble matching (under the meet and join operations defined in* (3.3) *and* (3.4)*, and the man-preferred or woman-preferred order) .*

*Proof of Theorem 11.* Notice that the non-emptiness of either maximal set is trivial, since by Lemma 14 we have that:

$$M(\mu_M) \supseteq M(\mu), \quad \forall\, \mu \text{ stable},$$

and hence the set of man maximal matching always contains $\mu_M$. To prove that the set of man maximal matching is also a sublattice, notice that if $\mu$ and $\mu'$ are man maximal, then in particular $M(\mu) = M(\mu')$. But if $m \in M(\mu) \cap M(\mu')$ then $m \in M(\mu \vee \mu')$ and $m \in M(\mu \wedge \mu')$. Hence:

$$M(\mu \vee \mu') = M(\mu \wedge \mu') = M(\mu') = M(\mu),$$

and then $\mu \vee \mu'$ and $\mu \wedge \mu'$ are maximal too. □

**Corollary 8.** *The set of maximal matchings is a (possibly empty) sublattice of the set of stable matching.*

*Proof of Corollary 8.* Observe that the set of maximal matchings is the intersection of the set of man maximal agents, and the set of woman maximal agents. Then being the intersection of two sublattices, it is a (possibly empty) sublattice.[88] □

---

[88]Certain textbooks define a sublattice as a *nonempty* subset closed under meet and join operations. If one refers to that definition, then a correct statement would be to say that the set of maximal matching is *either empty or a sublattice* of the set of stable matching.

**Theorem 12.** *If all men and women have preferences for the future (or the past) then the dynamic matching market is isomorphic to a sequence of "disjoint" finite matching markets in which all agents find only women in their year acceptable.*[89]

*If all men have preferences for the past, and women have preferences for the future, then the market is in general not isomorphic to a sequence of finite markets.*

*Proof of Theorem 12.* We will prove the first statement by assuming that both men and women have preferences for the future.

Consider a man $m \in M^t$, by assumption the set of women acceptable by $m$ is a subset $\bigcup_{s \geq t} W^s$. Since each woman in $\bigcup_{s>t} W^s$ has preferences for the future, she will not find $m$ acceptable. So we can eliminate from $m$'s preferences all the women entering the market in the future - without changing the set of stable matching. Doing the same with all the other men and women we obtain an isomorphic problem in which each man in $M^t$ (resp. woman in $W^t$) finds only women in $W^t$ (resp. men in $M^t$) acceptable. Clearly in this modified problem, each period $t$ is independent from any other period $t'$, so the conclusion follows.

For the second part, consider the Example 3 above, where the failure of the Rural Hospital Theorem was shown. Then divide the set of men and women in sets of two agents and defining $M^i := \{m_{2i}, m_{2i+1}\}$, and $W^i := \{m_{2i}, m_{2i+1}\}$. Then in the market thus defined women have preferences for the future, and men have preferences for the past - and plainly the market is not isomorphic to a sequence of finite markets, otherwise the Rural Hospital theorem would hold, thanks to Corollary 7. $\square$

[89]In this context isomorphic refers to the set of stable matchings. Formally, two markets $(M, W, \mathbf{P}_1)$ and $(M, W, \mathbf{P}_2)$ with different preferences are isomorphic if the set of stable matchings is the same.