Essays on Microeconomic Theory

Adriano Basso

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department ofEconomics
Adviser: Wolfgang Pesendorfer

June 2012

Abstract

This collection of essays investigates issues related to information acquisition in the presence of permanent ambiguity, perception errors or computational constraints. After a brief introduction, Chapter 1 introduces a decision maker who faces a sequence of non-identical experiments with a finite numbers of possible outcomes. The law generating the observations changes from period to period and the decision maker does not know anything about how the laws evolve. She has probabilistic beliefs over sets of possible laws and makes inferences on the true set of laws. I consider updating mechanisms characterized by consequentialism and such that updating can only depend on current beliefs and the current observation. To stay as close as possible to the standard Bayesian framework, I assume that the relative frequencies of observations converges to some limit frequency and that this limit frequency is compatible with only one of the possible sets of laws. I then show that, when the number of possible outcomes is larger than three, the individual may never learn what the true set of laws is.

In Chapters 2 and 3 I consider a decision maker receiving signals. The signals specify a subset of the set S of states of the world in which the true state in included. However, the decision maker does not observe the signals correctly: each signal is perceived as a possibly different subset of S according to a function v mapping true signals into perceived ones. Chapter 2 introduces the concepts of underconfidence and overconfidence for this setting and analyzes the consequences of different behavioral sources of misperception on the set of fixed points of the mapping function v and on the relation between each true signal A and the corresponding perceived signal $v(A)$. In Chapter 3 I add an ex-ante stage to the model. The decision maker is aware of her ability to manage a limited number of different signals, smaller than the total number of signals that she may receive. She therefore chooses an optimal function v mapping received signals into perceived ones, subject to the constraint on the number of the latter. The chapter studies the properties of this optimal mapping.

Acknowledgements

The essays in this dissertation would have never been written without the constant guidance and generous help of my adviser, Professor Wolfgang Pesendorfer. I also want to thank Professor Faruk R. Gul and Professor Stephen E. Morris for their useful comments.

Many people gave me their support during my years of Graduate School, allowing me to overcome the most difficult periods of this intense experience. My family, first of all, and then the friends I had the good fortune to meet in Princeton: Ling Ling Ang, Francesco Bianchi, Elisa Giaccaglia, Francesca Leoni and Alfonso Sorrentino.

Finally, a special thanks to Jayanti Chamling Rai, who has brightened my days during the last year.

To Jayanti, whom I love without ambiguity

Contents

Introduction

The chapters of this dissertations analyze the problem of beliefs updating from three different points of view, exploring issues related to information acquisition in the presence of permanent ambiguity, perception errors or computational constraints. In these contexts, the traditional Bayesian updating algorithm is either inapplicable (as in the case of ambiguity) or it produces unusual results (as in the case of misperceptions).

The model presented in Chapter 1 describe a situation of permanent ambiguity. A decision maker who faces a sequence of non-identical experiments with a finite numbers of possible outcomes. The law generating the observations changes from period to period and the decision maker does not know anything about how the laws evolve. Moreover, she thinks that no information can be obtained on that. She therefore thinks in terms of sets of probability distributions, or laws: from a set of possible laws, one is selected in each period to generate the observation. The decision maker has probabilistic beliefs over sets of laws. The absence of a probabilistic belief over laws (in contrast to one over sets of laws) gives rise to ambiguity. The impossibility of acquiring information on how the laws evolve makes this ambiguity permanent.

As in any model of ambiguity, it is not obvious what a reasonable mechanism for updating beliefs should look like. In this work, I consider updating mechanisms characterized by consequentialism and by a strong restriction of the amount of information that the decision maker can use: updating can only depend on current beliefs and the current observation. The purpose of Chapter 1 is to verify whether, as in a model with no ambiguity, this limited information is sufficient to asymptotically learn the true set of laws. To stay as close as possible to the standard Bayesian framework, I assume that the relative frequencies of observations converges to some limit frequency and that this limit frequency is compatible with only one of the possible sets of laws. The main result is that the individual may never learn what the true set of laws is. Interestingly enough, the impossibility result emerges only when the number of possible outcomes is larger than three. For a smaller number of outcomes, learning is always possible.

In the last two chapter, the decision maker is an expected utility maximizer and updates her beliefs through Bayes rule. Her behavior is, however, non-standard, because perception errors or the impossibility to handle too much information influence the way in which updating takes place.

In Chapters 2 and 3 the decision maker, before choosing an act, receives a signal. The signal specifies a subset of the set S of states of the world in which the true state in included. After observing the signal, the decision maker updates her prior according to Bayes rule and then chooses an act that maximizes her utility. However, she does not observe the signals correctly: each signal is perceived as a possibly different subset of S, according to a function v mapping true signals into perceived ones.

Chapter 2 analyzes the consequences of different behavioral sources of misperception on the characteristics of the mapping function v. In particular, I study the properties of set of fixed points of the mapping function v and the relation between each true signal A and the corresponding perceived signal $v(A)$. The chapter also introduces the concepts of underconfidence and overconfidence for this setting as a relation between received and perceived signals.

Chapter 3 continues the analysis of the model, adding an ex-ante stage. The decision maker is aware of her ability to manage just a limited number of different signals, smaller than the total number of signals that she may receive. Therefore, ex ante, she chooses an optimal function v mapping received signals into perceived ones, subject to the constraint on the number of the latter. Instead of perception errors, the model can be seen as describing a constraint in the amount of information that the decision maker can handle. The difference between received and perceived signals is now endogenous. The chapter studies the properties of the optimal mapping function v.

Chapter 1

Recursive Mechanisms for Updating Beliefs over Sets of Laws

1.1 Introduction

1.1.1 Preliminary Example

A Bayesian individual faces an urn of unknown composition. For the sake of simplicity, assume that the urn contains eight balls of three possible colors, red, blue and green. The individual performs an infinite sequence of experiments. In each period, she draws a ball (with replacement) from the urn and observes its color. She then updates the probabilities assigned to the different events (how likely she thinks it is to observe different sets of sequences of colors).

In this setting, de Finetti's theorem shows that beliefs over events can be reduced to beliefs over the the composition of the urn [1]. We can therefore suppose that the individual is uncertain between two compositions, as shown in Figure 1.1.[1] Her prior

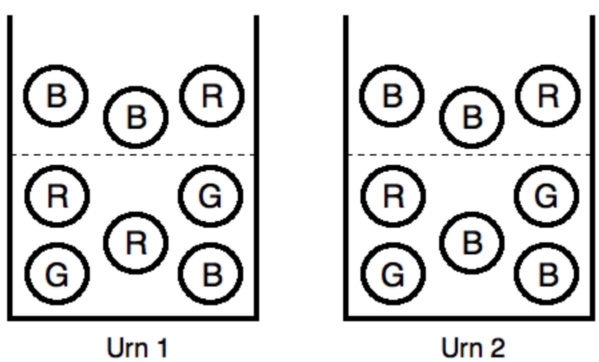

Figure 1.1: Unambiguous Urns

beliefs exhibit probabilistic uncertainty about the true composition. They can be

[1]The meaning of the dotted line will become clear later on.

represented with a probability distribution μ_0 over the two possible compositions. In each period, after observing the color of the ball, the individual updates her current beliefs in a Bayesian way. Once beliefs have been updated, she forgets both the observation and her past beliefs.

This is a very standard model. For future reference, it is useful to emphasize two important features of it. On the one hand, the relative frequencies of the colors observed converge in the long run to the true composition of the urn. On the other hand, the individual's beliefs μ_t converge to a degenerate distribution that assigns probability 1 to the true composition. Notice that this long-run identification is reached even if the individual makes use of a very limited amount of information. In particular, she is not aware of the values of the relative frequencies.

Consider now a less standard model. There is still a sequence of experiments in which a ball is drawn (with replacement) from an urn and its color is observed. However, the urn does not need to be the same in every period; the experiments are therefore non-identical. We can represent this modified setting with Figure 1.2. In

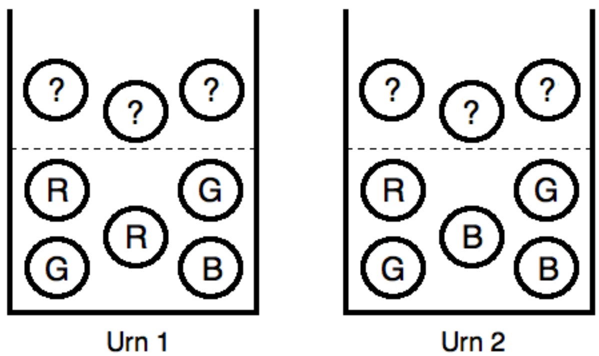

Figure 1.2: Ambiguous Urns

this example, the individual knows that, out of the eight balls in the urn, five balls are the same in every period, while the other three may change from period to period. She has no clues on how the changing balls are chosen in each period; moreover, she thinks that nothing can be known about that. What, at least in principle, can be known is the composition of the five fixed balls. On this regard, the individual is uncertain between two alternative compositions.

We can interpret this setting as a situation of permanent ambiguity. In fact, even if the individual were able to resolve her uncertainty about the composition of the fixed balls, she would still make no improvements in terms of understanding the mechanism that generates the changing balls. This source of ambiguity cannot disappear or be reduced in the long run. We can also think of this as a sort of "objective" ambiguity, intrinsic to the phenomenon itself and that cannot be entirely explained as a form of uncertainty aversion.

Equivalently, we can interpret the individual's beliefs as beliefs over sets of urns. Each set includes all the urns that share the same composition of the five fixed balls. So, for example, Urn 1 in Figure 1.2 correspond to the set of the urns with at least one red, two blue and two green balls. The convex hulls of the sets corresponding to the urns in Figure 1.2 are shown in Figure 1.3.[2] The darker triangle corresponds

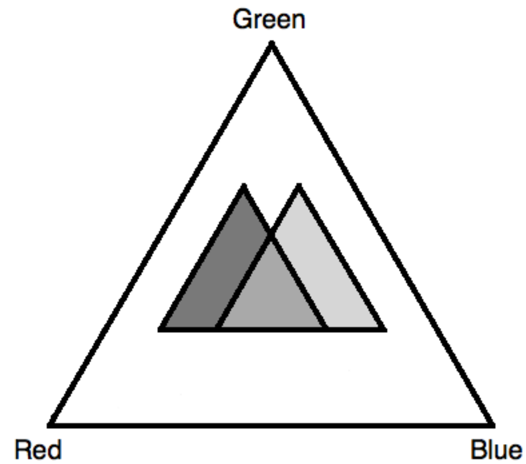

Figure 1.3: Sets of Urns

to Urn 1, the other to Urn 2. Notice that the sets are triangle; more precisely, they are similar to the simplex and differ from each other by a translation. The area of the triangles depends on the ratio between changing and fixed balls. As the example shows, the sets can intersect. We can think of these sets as "sets of laws", where a law is a probability distribution that, in a given period, assigns probabilities to the different colors that can be observed. A set of laws, therefore, can be used as a description of a sequence of non-identical experiments.

In order to remain as close as possible to the standard Bayesian model, I will assume that the individual has probabilistic uncertainty over possible sets of laws; that is, over the two possible compositions of the fixed balls. In each period, the individual draws a ball (with replacement) and observes its color. She then updates her beliefs and, after that, forgets the observation and her past beliefs. Therefore, the information that the individual can use for updating is the same as in the standard model. The class of rules for updating the probabilistic beliefs over sets of laws making use of this limited information will be denoted in this chapter as the class of "recursive updating rules". The term "recursive" underlines the fact that the updating algorithm is the same in every period: it is a function of current beliefs and of the current observation.[3]

To make things even more similar, I make an additional assumption on the sequence of observations. Since nothing is known about the mechanism that selects the changing balls, there is no guarantee of convergence of the relative frequencies of the

[2]In the rest of the chapter, I will consider convex sets. So, in this discrete example, it makes sense to look at the convex hulls.

[3]The standard Bayesian rule is an example of recursive updating rule. However, even in a non-ambiguous setting, one can think of many more recursive rules.

colors observed. By analogy with the Bayesian model, I assume that such a convergence takes place. Moreover, I require that this limit frequency is compatible with one and only one set of urns. That is, I assume that the limit frequency is a point belonging to only one of the convex hulls of the two possible sets, and not to their intersection. With this assumption I exclude the cases of "undecidability"; in fact, it is clear that, if the individual were able to keep track of the relative frequencies, in the long run she would identify the true set of urns.

Given these analogies with the Bayesian model, one may wonder whether similar limit beliefs can be obtained. Since in the non-ambiguous case the beliefs of a Bayesian individual converge to a degenerate distribution assigning probability one to the true urn, can a similar result hold in a model of permanent ambiguity? That is, is it possible for the individual, in the long run, to identify the true composition of the fixed balls using some recursive updating rule? This is the question that I address in the present chapter. The model I use will be, of course, more general than this preliminary example. In particular, I will allow for an arbitrary finite number of outcomes (or colors). Notice, however, that the generalizations would not undermine the convergence result in the non-ambiguous case.

I will prove that, under mild regularity conditions, convergence to the true set of laws is not generally possible. In particular, problems arise when the number of outcomes (colors) is higher than three. The main result of this chapter is therefore an impossibility theorem. For up to three outcomes, on the other hand, we do have a result analogous to the usual Bayesian one.

1.1.2 Existing Literature and Plan of the chapter

The idea that a decision maker could interpret a sequence of experiments as being non-identical was first considered by Epstein and Schneider [3], who define the notion of indistinguishability as opposed to independence. Sequences of non-identical experiments were then explicitly introduced in the economic literature by Epstein and Seo [4]. Notice, however, that similar ideas had been developed in a statistical context; see for example Fierens and Fine [6].

Epstein and Seo provide a behavioral characterization of a utility function that will be used in the present chapter. I will not go over the axioms; here I just want to spend a few words on the interpretation of the function. Let S be the set of outcomes of each experiment; that is, the set of possible colors of a ball. The state space is then given by $\Omega = S^\infty$. Let Σ be the product σ-algebra on Ω. An act f is a Σ-measurable function from Ω to $[0,1]$. The decision maker's preference relation can be represented by a utility function of the form

$$U(f) = \int \left(\min_{\mathcal{L}^\infty} \int f dP \right) d\mu(\mathcal{L}),$$

where

- $\mathcal{L}$ is a closed and convex set of probability distributions over outcomes that can characterize an experiment; it correspond to the (convex hull of) a set of urns as described in the example above; in my terminology, it is a set of laws;

- P is a probability measure on (Ω, Σ); it is a distribution over sequences of outcomes (sequences of colors, in the example above), constructed taking one law in $\mathcal{L}$ for each period and considering the experiments as independent;

- μ is a probability distribution over sets of laws; it describes the decision maker's probabilistic uncertainty over sets (that is, in my example, over compositions of the fixed balls in the urn).

With a change of notation, it is possible to show that this representation is a special case of the multiple-prior utility function axiomatized by Gilboa and Schmeidler [9]. The interested reader can look at [4]. Notice, however, that the source of ambiguity in this model is different from that implied in the traditional interpretation of the Ellsberg paradox. Usually, the decision maker is seen as unable to express her uncertainty on the value of the relevant parameter in terms of a probability distribution. Here, on the other hand, the relevant parameter is the set of laws $\mathcal{L}$, and the decision maker does have probabilistic uncertainty over its value. The ambiguity emerges from the relation between the parameter and the outcomes. The difference becomes crucial when we think about belief updating. In this context, the standard generalized Bayesian rule, as axiomatized in Pires [13], that consists in the Bayesian updating of each prior, does not make sense. The reason is that there is not a natural likelihood function to use. Updating becomes a more subtle problem. The present chapter addresses some of the issues emerging when we try to model the process of updating beliefs over sets of laws.

The problem of belief updating in the case of non-identical experiments was first studied within a statistical framework. Fierens, Rêgo, and Fine [7] summarize the existing literature and provide a complete bibliography on the subject. Their approach involves looking at the sequence of outcomes and choosing a particular set of subsequences to use as inputs for inference. The authors characterize a set of subsequence selection rules that allows to identify the true set of laws with high probability. The rules are quite complicated and depend on the bound imposed to the complexity of the rule for the selection of laws. This approach is unlikely to be applicable to models of decision making. A behavioral characterization of the subsequence selection rules is probably impossible.

A different approach has been proposed by Epstein and Seo [4]. Given the representability of preferences with a utility function of the form seen above, they look for a behavioral characterization of the existence of a likelihood function.[4] Again, the axiomatization can be found in their chapter. Using a notation slightly different from [4], but closer to the one adopted later in the chapter, let Λ be the space of possible sets of laws $\mathcal{L}$. A likelihood is a function $L : \Lambda \to \Delta(\Omega)$. Given L, it is possible to

[4] The definition of likelihood function in [4] is different from the definition that will be used later in the present chapter.

define its one-step-ahead conditional at period n as a function $L_n : S^{n-1} \times \Lambda \to \Delta(S)$. Beliefs are updated according to a rule of the form

$$d\mu_n(\mathcal{L}) = \frac{L_n(s_n|\mathcal{L})}{\overline{L}_n(s_n)} d\mu_{n-1}(\mathcal{L}),$$

where $\overline{L}_n(\cdot) = \int L_n(\cdot|\mathcal{L}) d\mu_{n-1}(\mathcal{L})$. Notice that L_n depends on the entire history of past observations, that is, on the sequence s^{n-1}.

The class of recurses updating rules, as defined in this chapter, are neither a subset nor a superset of the class of likelihood functions in [4].

The rest of the chapter is organized as follows. Section 2 formalizes the model and compares it to the one adopted by Epstein and Seo [4]. Recursive updating rules are defined in Section 3, where they are classified according to their functional forms. In particular, a nonstandard definition of likelihood function is introduced. The main results of the chapter are presented in Section 4. Section 5 concludes. Proofs can be found in Appendix A.

1.2 Non-identical Experiments and Beliefs

1.2.1 The Model

A decision maker faces an infinite sequence of experiments, each yielding an outcome in the finite set $S = \{s_1, ..., s_S\}$ (with a little abuse of notation, I use S to denote both the set of outcomes and its cardinality). The state space is therefore $\Omega = S^\infty$. Let Σ be the product σ-algebra on Ω. An act f is a Σ-measurable function from Ω to $[0, 1]$.

Using the slightly abusive notation introduced by Epstein and Seo [4], I assume that the decision maker's preference relation can be represented by a utility function of the form

$$U(f) = \int \left(\min_{\mathcal{L}^\infty} \int f dP \right) d\mu(\mathcal{L}), \tag{1.1}$$

where $\mathcal{L}$ is a closed and convex set of probability distributions over outcomes that can characterize an experiment (that is, a set of laws), P is a probability measure on (Ω, Σ) and μ is a probability distribution over sets of laws. So the decision maker thinks of all the experiment as independent and governed by the same, but unknown, set of laws; she has probabilistic beliefs over possible sets of laws, represented by μ. The utility of an act is determined minimizing the expected utility over all the distributions $P \in \mathcal{L}^\infty$. Therefore, the decision maker exhibits uncertainty aversion.

This representation has been axiomatized by Epstein and Seo [4]. It is possible to show that such a utility function is a special case of the Gilboa-Schmeidler [9] multiple-prior utility (see [4] for a detailed analysis).[5]

[5]Notice, however, that in my analysis I do not explicitly make use of this utility function. I am only interested in limit beliefs, not in choices. In fact, all I need is a representation of preferences where uncertainty over sets of laws is probabilistic.

In the chapter I focus on the case in which the decision maker is uncertain between a finite number of sets of laws $\{\mathcal{L}_i, i = 1, ..., h\}$. Moreover, I assume $\mathcal{L}_i \subset int(\Delta(S))$ and

$$\mathcal{L}_i = \alpha\{p_i\} + (1-\alpha)\Delta(S),$$

where $p_i \in \Delta(S)$ and $\alpha \in (0, 1]$. It is easy to interpret such a set keeping in mind the urn example:

- p_i is the (unknown) composition of the fixed balls in the urn;
- α is the ratio between the numbers of fixed and changing balls in the urn;
- the term $(1-\alpha)\Delta(S)$ says that the changing balls can assume any possible composition.

The assumption that the sets of laws belong to the interior of the simplex is made just for convenience, to get rid of all the cases of zero-probability events.

The chapter considers issues related to the long run beliefs that can be obtained when updating is based on iterative mechanisms. But what does updating exactly mean in the context of the present model? I am assuming that the decision maker does not know anything about how, among the set of possible laws, the specific law describing the experiment in a given period t is chosen. Moreover, she thinks that nothing can be inferred about this selection mechanism from the observation of the outcomes in earlier periods. Formally, this means that the sets $\mathcal{L}_i$ in the utility functions $U_t(f)$ are the same in every period t.

Following Epstein and Seo [4], I assume that updated conditional preferences at any period t are represented by the same functional form (an assumption that those authors call Basic); moreover, Consequentialism is assumed: the conditional ranking given the sample s^t does not depend on what the act would have yielded under events that have not been realized.

Consequentialism. *If* $f(s^t, .) = f'(s^t, .)$, *then* $f \sim_{t,s^t} f'$.

With these two assumptions, preferences at time t can be represented by a utility function of the form

$$U_t(f) = \int \left(\min_{\mathcal{L}^\infty} \int f(s^t, .)dP \right) d\mu_t(\mathcal{L}).$$

Therefore, updating only affects the probability distribution μ over the sets of laws. For this reason, in the chapter I will often refer to the sets $\mathcal{L}_i$ as "parameters", while the term "beliefs" will be used to denote μ.

Given the decision maker's initial beliefs μ_0, an updating rule is therefore a rule specifying how μ_t is obtained after the agent observes the outcomes $s^t = s_1, ..., s_t$.[6] My analysis will focus on a specific class of updating rules, for which updating does not

[6]Note that I am using the same notation s_i to denote both the ith element of the set S and the observation in the ith period. The interpretation in each instance where s_i is used will be clear form the context.

depend on past observations and beliefs (see Section 1.3). I am interested in particular in the limit beliefs as $t \to \infty$. When is it possible to converge to a probability distribution assigning probability one to the true parameter? Does the answer depend on the cardinality of the set S? These are some of the questions addressed in the following sections.

1.2.2 Differences with Respect to the Epstein-Seo Model

The present model is clearly a special case of the one studied by Epstein and Seo [4]. In fact, I am simply restricting the set of possible parameters, assuming that the decision maker makes a clear distinction between the features that are common to all the experiments and those that are specific to each experiment. In the urn example, the agent knows how many balls are fixed in every period and how many change from period to period. This implies that the different sets of laws among which the agent is uncertain can be obtained translating a single set. Formally, two sets of laws can differ only by the term p_i in the expression above. This is not the case for Epstein and Seo. In their model the decision maker is unsure not only about the colors of the fixed balls, but also about their number.[7]

To understand why this restriction matters, remember that nothing is known about how a specific law is selected in each period. In particular, it cannot be excluded (and the decision maker in my model does not exclude) that in each period one law in chosen randomly among the laws in the true set.

Consider now an example with $S = 2$ and suppose that the decision maker is uncertain between two sets of probabilities as represented in Figure 1.4 below. This

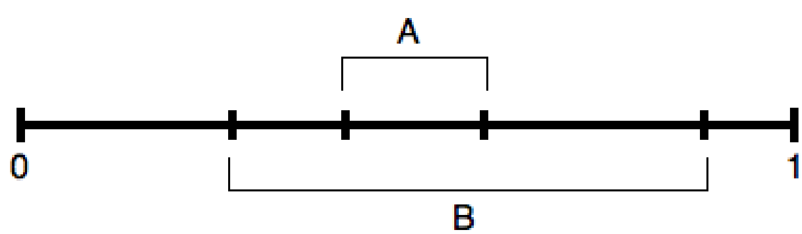

Figure 1.4: Excluded Case

is a case which is excluded in my model (the two sets do not differ by a translation), but which is perfectly possible in the model by Epstein and Seo. Assume that the true parameter (set of laws) corresponds to the smaller set A. In this case, there is no reasonable way to identify the parameter. In fact, any sequence of outcomes obtained with a random selection rule over A can be equally obtained with a random selection rule over the the set $B \setminus A$.[8] Restricting the parameters to translations of

[7]Epstein and Seo [4] provide an axiomatic characterization of such beliefs. I do not have axioms characterizing the special case I adopt.

[8]One way to address this problem is to restrict the set of possible selection rules, excluding, in particular, random rules. This is the approach followed in the statistical literature (see [7]).

a single set eliminates the possibility for the decision maker to be uncertain between a set of laws and one of its subsets. In this case, whatever the true parameter, there are always sequences of outcomes for which a correct identification is, at least theoretically, possible. Consider Figure 1.5 below.

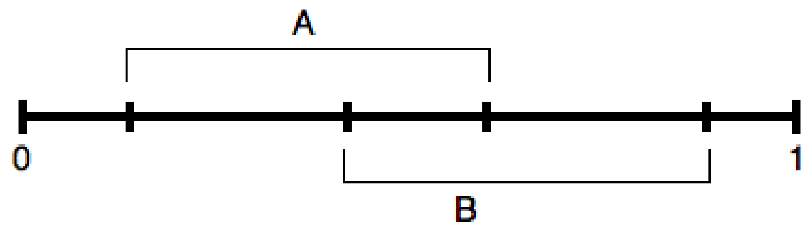

Figure 1.5: Acceptable Case

If, for example, the sequence of outcomes generates frequencies that converge to to a point in $A \setminus B$ ($B \setminus A$), than any individual with enough information can learn that the true set of laws is A (B). Such limit frequencies cannot be generated by a sequence of laws selected from the set B (A). Since in this chapter I am interested in determining the ability of a decision maker with limited information to learn the true set of laws, I need to restrict myself to cases in which, at least in principle, the true set can be inferred from the observations.

1.3 Recursive Updating Rule

In the classic Bayesian model, the decision maker makes use of a limited amount of information. The only input of the Bayesian algorithm is the current observation. In my model, I want to preserve this constraint, therefore considering learning processes characterized by an incremental and recursive nature. Unlike Epstein and Seo [4], I am not interested in a behavioral characterization of such a class of updating rules. My aim will be to see whether this limited information is still sufficient for the convergence of beliefs to the true value of the parameter, that is, to the true set of laws.

As stated in the previous section, I am assuming that conditional preferences can be represented by utility functions of the same functional form. To that assumption I then add the requirement that the decision maker cannot remember anything about the past. We then get what I call "recursive updating rules". We can think of these rules as the class of mechanisms for updating the value of μ that make use of a very limited amount of information. In particular, at each period t, only two pieces of information can be used:

- the current observation s_t, that is, the outcome of the current experiment;
- μ_{t-1}, the agent's beliefs at the beginning of period t.

However, this is technically complicated and it is difficult to imagine a behavioral characterization of such a restriction.

Notice that, when applied to the non-ambiguous setting, this class includes Bayes rule, among many others.

It is useful to classify recursive updating rules into subgroups. I will consider only deterministic mechanisms, in which beliefs are updated applying the same rule in each period. Therefore, there exists a function $f : (S, \Delta(\{\mathcal{L}_i\})) \to \Delta(\{\mathcal{L}_i\})$ such that, given the observation s_t in period t,

$$\mu_t | s_t = f(s_t, \mu_{t-1}).$$

In the literature on beliefs updating it is common to focus on mechanisms making use of likelihood functions. In the present chapter I use a nonstandard definition: I call likelihood function any function $g : S \to \mathbb{R}^h_{++}$ such that

$$(\mu_t | s_t)_i = \frac{g(s_t)_i \mu_{t-1,i}}{\sum_{j=1,...,h} g(s_t)_j \mu_{t-1,j}}, \quad i = 1, ..., h.$$

The difference with respect to the common notion of likelihood function is that I do not require $\sum_{s \in S} g(s)_i = 1$ for all $i = 1, ..., h$. So, strictly speaking, g cannot be interpreted as the probability of the current observation given a specific value of the parameter. When I talk about a likelihood function, I am therefore thinking of any updating mechanism showing a multiplicative separation of s and μ. Again, Bayes rule is a special case of this class of rules.

It is useful to consider further subsets of mechanisms of the likelihood function type. Let $\mathcal{L}_i = \alpha\{p_i\} + (1 - \alpha)\Delta(S)$. We can think of two relevant special cases:

1. Likelihood functions associate to each parameter probabilities belonging to the correspondent set of laws: $g(s)_i \in \alpha p_i(s) + (1 - \alpha)[0, 1]$. We can think of these as the only "meaningful" likelihoods, since, given any outcome and any parameter, they assume one of the possible values of the conditional probability of the outcome given the parameter.

2. Likelihood functions select conditional probabilities that have the same "position" within each set of laws: $g(s)_i = \alpha p_i(s) + (1 - \alpha)q_s(s)$, where $q_s \in \Delta(S)$ is the same for all is. So, in the urn example, we can interpret different likelihood functions as different hypotheses on the composition of the variable part of the urn.

Clearly, the second subset of mechanisms is a special case of the first.

In the rest of the chapter, I will consider cases in which the decision maker is uncertain between two possible sets of rules only, $\mathcal{L}_1$ and $\mathcal{L}_2$; that is, $h = 2$. The assumption simplifies the proofs, but does not reduce the power of the theorems. Any result can be easily generalized to an arbitrary finite h. A more serious restriction is the introduction of some mild regularity conditions.

Well-behavedness. *Given $h = 2$, let's simplify the notation redefining f as a function mapping $(S, (0, 1))$ to $[0, 1]$, where $\mu \in (0, 1)$ is now the probability associated to $\mathcal{L}_1$. I require that, for any $s \in S$,* *i)* $\lim_{\mu \to 0} f(s, \mu) = 0$ *and* $\lim_{\mu \to 1} f(s, \mu) = 1$*;*

ii) $f'(s, .)$ *has limits (possibly infinite) for* $\mu \to 1$ *and* $\mu \to 0$*.*

The first condition has a simple interpretation. As beliefs approach certainty, a single observation cannot change them dramatically. This is a restriction on the weight that can be assigned to observations relative to prior beliefs or, more precisely, a restriction on the speed with which this weight can increase as beliefs approach certainty. The second condition can be interpreted as an additional restriction on the weights given to observations for different prior beliefs. I assume that, as beliefs approach certainty, these weights don't oscillate indefinitely. The relevance of observations for updating beliefs converges to a singe value, not necessarily finite. Notice that I do not require $f(s, .)$ to be continuous on the entire interval $(0, 1)$. Observations may in general be weighted very differently for arbitrarily close priors. What I need is that this should not be the case in the limit; that is, I need $f(s, .)$ to be definitely uniformly continuous.

A function f satisfying these conditions will be called "well-behaved". Assuming that f is well-behaved is a significant restriction. The assumption clearly reduces the power of Theorem 1.4.1 below. In order to prove that the amount of information that the decision maker can use in my model is not sufficient for a correct identification of the true set of laws, one would like to get rid of these regularity conditions, however mild they may appear. Whether the same result can be proved for generic functions f is still an open question.

1.4 Updating Rules and Long Run Beliefs

Before studying the decision maker's limit beliefs when updating is recursive, I want to introduce an additional assumption. The assumption will make the similarity with the standard Bayesian model even stronger. Since experiments are non-identical and not even independent, there is no guarantee that the frequencies of the outcomes in S will converge in the limit. In the following theorems, however, I will always assume that convergence takes place, in analogy with the model without ambiguity. Informally, we can interpret the experiments as independent "in the limit". Notice that, since my main result will be an impossibility theorem, such a restriction does not affect the strength of the result.

I want to study the possibility for the decision maker of learning which of the two sets of laws is the true one. However, as we have seen in the Introduction, there are cases that are objectively undecidable. Consider $\mathcal{L}_1$ and $\mathcal{L}_2$ such that $\mathcal{L}_1 \cap \mathcal{L}_2 \neq \emptyset$. Suppose that the relative frequencies of the outcomes converge to a point in the intersection. Remember that the decision maker does not know anything about how a law is selected in each period from the true set. Moreover, she thinks that any selection mechanism is possible. In particular, a random mechanism cannot be excluded. For such a decision maker, limit frequencies in $\mathcal{L}_1 \cap \mathcal{L}_2$ are compatible with both sets of laws. This is true no matter how much information she has on the history of past observations.

On the other hand, if the decision maker were able to observe the frequencies of the various outcomes, it would make sense for her beliefs to converge to $\mu(\mathcal{L}_1) = 1$ ($\mu(\mathcal{L}_2) = 1$) if the limit frequencies belong to the set $\mathcal{L}_1 \setminus \mathcal{L}_2$ ($\mathcal{L}_2 \setminus \mathcal{L}_1$). In fact, a sequence of outcomes whose limit frequencies fall outside $\mathcal{L}_i$ cannot be generated by a sequence of laws belonging to $\mathcal{L}_i$. Clearly, since the decision maker in my model uses a recursive updating rule, she does not have information on the empirical frequencies. This, however, is not a problem in the standard Bayesian case: the true parameter will be correctly identified in the long run. The question I want to address is whether such a result is still possible in the present model, where permanent ambiguity is introduced. More precisely, I am going to require something a little weaker: I ask for a correct identification of the true set only when limit frequencies fall in $int(\mathcal{L}_i \setminus \mathcal{L}_j)$. As the following theorem shows, even this is generally impossible for well-behaved updating rules.

Theorem 1.4.1. *Suppose $S \geq 4$. Let $r_t(s)$ be the relative frequency of the outcome s after the first t periods. Let $f : (S, \Delta(\{\mathcal{L}_i\})) \to \Delta(\{\mathcal{L}_i\})$ be any well-behaved recursive updating rule. Then, there exist $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mu_0 \in \Delta(\{\mathcal{L}_i\})$ such that, for some $p \in int(\mathcal{L}_1 \setminus \mathcal{L}_2)$, $r_t \to p$ but $\mu_t \nrightarrow \delta_{\mathcal{L}_1}$.*

Theorem 1.4.1 says that, if there are more than three possible outcomes, it may be impossible for the decision maker to learn the true set of laws, even if the sequence of observations is compatible with one of the two sets only. Therefore, the identification of the true parameter, possible in the standard Bayesian model, does not extend to the case of permanent ambiguity. This is true, at least, if we restrict ourselves to well-behaved updating rules. The proof of the theorem makes clear that the problems in identifying the true set arise only when the sets intersect (and not even in all of these cases).

What is generally not possible when $S \geq 4$ becomes possible if $S \leq 3$. In this case there always exists an updating rule with a likelihood function g such that the parameter $\mathcal{L}_1$ ($\mathcal{L}_2$) is correctly identified if the limit frequency falls in $int(\mathcal{L}_1 \setminus \mathcal{L}_2)$ ($int(\mathcal{L}_2 \setminus \mathcal{L}_1)$).

Theorem 1.4.2. *Let $S \leq 3$ and let the decision maker be uncertain between the parameters $\mathcal{L}_1 = \alpha p_1 + (1-\alpha)\Delta(S)$ and $\mathcal{L}_2 = \alpha p_2 + (1-\alpha)\Delta(S)$. There exists a likelihood function $g : S \to \Delta(S)$ such that, if the limit frequency falls in $int(\mathcal{L}_1 \setminus \mathcal{L}_2)$ ($int(\mathcal{L}_2 \setminus \mathcal{L}_1)$), then the decision maker's beliefs converge to $\delta_{\mathcal{L}_1}$ ($\delta_{\mathcal{L}_2}$).*

It would probably be possible to prove that the likelihood function can be chosen such that $g(s)_i \in \alpha p_i(s) + (1-\alpha)[0,1]$ (the first of the special cases considered in Section 1.3). For the special case in which $int(\mathcal{L}_1) \cap int(\mathcal{L}_2) \neq \emptyset$, this result is easy to prove, as shown in Appendix A (Remark 1). I have not tried to prove whether this is true in general.

On the other hand, for $S = 3$, it turns out that it is not possible to have a likelihood function satisfying the second special case in Section 1.3. In fact, as the following lemma shows, we can find examples where there is no $q_s \in \Delta(S)$ such that $g(s)_i = \alpha p_i(s) + (1-\alpha) q_s(s)$ for $i = 1, 2$.

Lemma 1.4.1. *For $S = 3$, there exist $\mathcal{L}_1 = \alpha p_1 + (1-\alpha)\Delta(S)$ and $\mathcal{L}_2 = \alpha p_2 + (1-\alpha)\Delta(S)$ such that no likelihood functions satisfy both Theorem 4.2 and the condition $g(s)_i = \alpha p_i(s) + (1-\alpha)q_s(s)$, for $i = 1, 2$ and $q_s \in \Delta(S)$.*

Whether this becomes possible for $S = 2$ is still an open question.

1.5 Conclusions

In this chapter I have analyzed some issues arising from the problem of updating beliefs when the observations come from a sequence of non-identical experiments. I have considered the case in which the individual has severe limitations on the amount of information she can carry from one period to the next. In particular, she forgets all past observations and past beliefs; all she can remember are her beliefs at the end of the previous period. I focused on this special case to preserve an analogy with the classical Bayesian updating model. Moreover, while forgetfulness of all past observations can be considered an extreme case, forgetfulness of past beliefs is a very sensible assumption. In particular, the possibility of choosing an updating rule as a function of initial beliefs μ_0 assigns to those beliefs a special status that is hardly justifiable. In fact, what should be the intrinsic difference between initial beliefs and those in following periods? Not to mention the fact that in real situations the notion of "initial period" is ambiguous at best.

To compensate for this limitations, I have restricted myself to a framework that, intuitively, would be the most favorable in order to learn the set of laws that characterize the sequence of experiments. Therefore I have considered the case of two only possible sets of laws and, within each set, I have assumed that there is a clear distinction, known by the individual, between what is common to all the laws and what changes from law to law. I have also excluded the possibility of sequences of observations compatible with both sets of laws. Finally, I have assumed that relative frequencies converge to a limit distribution.

The main result (Theorem 1.4.1) is an impossibility statement: given the above limits on available information, there are no "well-behaved" updating rules assuring that beliefs converge in the limit to a distribution that assigns probability one to the true set of laws. "Well-behavedness" is an assumption on the form of the updating function that imposes some regularity conditions in the limit, as the probability assigned to one of the sets of laws approaches one.

One weakness of the paper is apparent: Theorem 1.4.1 works under a specific assumption on the class of updating rules we are allowed to consider, namely well-behavedness. It is maybe a weak assumption, but it is still annoying. It would be interesting to verify whether the theorem holds when this assumption is dropped. Is it true that there are no recursive updating rules, no matter how they behave when beliefs approach certainty, that allow the individual to asymptotically identify the true set of laws?

Finally, it may be interesting to find general conditions under which identification is possible. Theorem 1.4.2 states that, when the number of different outcomes is at most three, beliefs converge to a distribution assigning probability one to the true set of laws. Alternatively, we may ask what minimum amount of information needs to be retained in order to have the same limit result with an arbitrary finite number of outcomes. Does the individual need to remember all past observations? Is there some sufficient statistics?

The model of ”objective ambiguity” introduced by Epstein and Seo [4] is surely worth additional investigation. The present work, and the conclusive suggestions, are just one of the possible ways to understand its deepest implications.

Chapter 2

Updating and Misperception of Signals

2.1 Introduction

Beliefs updating has generated an immense literature. Several rules (Bayes, Generalized Bayes, Dempster-Shafer, etc.) have been studied and applied to preferences with different utility representations. For recent developments see, for example, Pires [13], Wang [15], Hanany and Klibanoff [11], Eichberger, Grant and Kelsey [2] and Shmaya and Yariv [14]. In these works, the decision maker observes an event or, more generally, a signal. Given this information, she updates her prior beliefs. Looking at the conditional preferences and and having observed the signal, an external observer can determine the rule the decision maker uses to update her beliefs (or, more precisely, he can model the change in preferences as if the decision maker was updating her beliefs according to a certain rule). In all these models, the decision maker and the external observer see the same signals. But what happens if the decision maker does not observe the true signal?

Consider the following example. An urn contains balls with colors in the set S. A ball is drawn and a signal is generated. The signal gives some information on the color of the ball: it says that the color belongs to a certain subset of S. If the color is s, the possible signals can be identified with subsets of S whose elements include s. Assume that, given the true color s, any signal that includes s may be generated. So, for example, the signal $\{s_1, s_2\}$ can be observed under both states s_1 and s_2. An external observer is aware of the true signal, but the decision maker is not: her observational capabilities can be described with a function that assigns to each signal some other set of colors. This means that, although the decision maker may not observe the signal correctly, she still interprets signals as subsets of S. Assuming a correspondence between signals and sets of states is clearly a serious restriction to the applicability of the model. However, I think that even in this limited setting interesting questions can be addressed.

Over this structure one can accommodate any kind of updating rules. Given initial beliefs and the signal as observed by the decision maker, one can apply any of the updating rules that have been studied in the economic literature. In the present work, however, I am only interested in the characteristics of the function mapping each true signal to what the decision maker perceives. I will introduce assumptions of the sort "given a certain relation between two true signals, the corespondent signals perceived by the decision maker must satisfy some other relation." These assumptions formalize the classes of mistakes that the decision maker can or cannot make. I will then look at what those assumptions imply in terms of the set of signals that are correctly perceived and at how we can restrict the class of perceived signals that the decision maker may associate to each true one.

Section 2.2 introduces two additional assumptions that will be always maintained throughout the chapter. One is a standard non-degeneracy assumption for conditional preferences. The other, that I will call Coherence, says that any true signal is mapped into a subset of S that would be correctly recognized if observed. The idea is that there exists a function $v : \Sigma \to \Sigma$ such that, after receiving the signal A, the decision maker behaves as if she had observed the signal $v(A)$. But this interpretation makes sense only if, when the decision maker receives the signal $v(A)$, she correctly perceives it. Therefore, for any signal A, I require that $v(A)$ is a fixed point of the function v.

Some relations between real and perceived signals can be interpreted as an indication of the level of the decision maker's confidence on what she observes. So we can interpret a situation in which $A \subset v(A)$ as a case of underconfidence: the decision maker does not feel she can exclude all the states of the world that are impossible given the signal. On the other hand, she can be seen as overconfident if $v(A) \subset A$: after receiving the signal A, she is so confident about some of the states of the world that she dismisses the others even though the signal does not justify this exclusion. It is tempting to interpret underconfidence as a version of conservatism. However, in the traditional definition of conservatism, as given by Phillips and Edwards [12], the decision maker's updated beliefs lie somewhere between prior beliefs and what would be implied by Bayesian updating. This is not necessarily the case in the present context, in which nothing is assumed about beliefs over conditionally non-null states.

In Sections 2.3 and 2.4 I will study two specific sets of behavioral assumptions from which underconfidence and overconfidence can emerge. In the first model, the decision maker never excludes any state that is possible under the signal, but may not be able to distinguish between some states. So the signal she perceives is the one containing all the states that are indistinguishable from some state in the true signal. The second model adds a second source of misperception. Some sets of states are more vivid than others. When the true signal contains a set of states that is more vivid compared to some other state in the signal, this last state is excluded by the decision maker. These assumptions have some interesting consequences with regards to the set of fixed points of the function v and to how $v(A)$ is computed for each true signal A.

Section 2.5 presents some considerations on how the decision maker interprets signals that differ in the amount of information they provide. In particular, I will analyze the relation between $v(A)$ and $v(B)$ on one side and $v(A \cup B)$ or $v(A \cap B)$ on the other.

Proofs can be found in Appendix B.

2.2 Framework

Let S be a finite set of states of the world; Σ, the set of all subsets of S, is the set of events. Let X be the set of outcomes and define acts as functions $f : S \to X$ mapping states of the world into outcomes. L will denote a set of acts and $\succsim$ a preference relation over L.

The decision maker receives a signal providing information on the state of the world. I am going to distinguish between what an external observer (let's call it the "experimenter") can see, and what can be perceived by the decision maker. The experimenter observes the signal correctly. Moreover, he knows what the probability of any signal is, conditional on each state; that is, he knows the "true" likelihood function λ. This is not necessarily the case for the decision maker.

Let's denote with H the set of possible signals. Throughout the chapter I will assume that, for any $\eta_1, \eta_2 \in H$, there exists $s \in S$ such that either $\lambda(\eta_1|s) > 0$ and $\lambda(\eta_2|s) = 0$, or $\lambda(\eta_1|s) = 0$ and $\lambda(\eta_2|s) > 0$. That is, different signals correspond to different subsets of Σ.

For example, imagine a situation in which a ball is drawn from an urn. The possible colors of the ball constitute the set of states S. A signal is then generated. The signal gives some information on the color of the ball: it says that the color belongs to a certain subset of S. Given the state s, the set of possible signals can be identified with a subset of S whose elements include s. Assume that, for any signal and for any s included in it, the likelihood of observing the signal given s is positive. So, for example, the signal $\{s_1, s_2\}$ can be observed under both states s_1 and s_2. Such an example satisfies the assumptions of my model.

Notice that the decision maker may have a wrong belief on how signals are generated given states; or she may be unable to distinguish between different signals. However, I will always assume that the decision maker associates signals with subsets of S, so that different signals (according to what she can perceive) correspond to different subsets. This will imply that conditional beliefs can be identified with a set of conditionally non-null states. That is, there cannot be two different conditional beliefs that share the same set of non-null states.

The framework can accommodate decision makers with various ways of misinterpreting the signals. In the chapter I will consider some behavioral assumptions that may seem quite reasonable, and I will ask what class of updating mechanisms describes a decision maker behaving in this way.

Let's formalize the above discussion, redefining the properties in terms of preferences. Given the correspondence between signals and elements of Σ, in what follows I will use the same notation for both signals and events. The meaning in each instance will be clear from the context. Therefore, the symbol $\succsim^A$ denotes the preference relation conditional on the signal corresponding to the set A. Notice that this is the "true" signal, that is the one observed by the experimenter. I will also focus on the special case in which $H = \Sigma$, that is, for any subset of S there exists a corresponding signal.[1] As I said, I want to assume that even for the decision maker there is an analogous relation between signals and events. Formally, this corresponds to the following assumption.

$\mathcal{A}1$ (Signals as events) For any $A, B \in H$, if the set of $\succsim^A$-non-null events is equal to the set of $\succsim^B$-non-null-events, then, for any $f, g \in L$, $f \succsim^A g \iff f \succsim^B g$.

When the decision maker observes the signal E, she interprets it as saying that some events are impossible. So she divides Σ into two groups of conditionally null and non-null events. It is this reformulated information that enters the updating mechanism, so that the new preferences $\succsim^E$ are completely characterized by the set of $\succsim^E$-non-null events. The axiom excludes the existence of different conditional preferences $\succsim^A$ and $\succsim^B$ sharing the same set of conditionally non-null events. Thanks to Axiom 1, we can represent the updating rule with a function $v : \Sigma \to \Sigma$ mapping each signal into the corresponding set of conditionally non-null states of the world.

The role of Axiom 1 is to limit the class of updating rule under consideration. I cannot exclude that my approach could generalize to a more general set of rules. The axiom, however, is reasonable enough not to obliterate the interest of the results in this limited setting.

In the chapter I will always consider updating rules that satisfy an additional axiom, that I call "Coherence".

$\mathcal{A}2$ (Coherence) Given any signal E, let E' be the set of $\succsim^E$-non-null states, that is, for any $F \subseteq E'$, $\exists f, g, h \in L$ such that $f|^h_{F^c} \succ^E g|^h_{F^c}$, where $f|^h_{F^c}$ denotes an act that coincides with f on F and with h on F^c. Then, E' is also the set of $\succsim^{E'}$-non-null states.

In other words, Axiom 2 states that the function v maps any signal into a fixed point: $\forall E \in \Sigma$, $v(v(E)) = v(E)$. Notice that, up to this point, I have not made any general assumption on the relation between E and $v(E)$. It may well be the case that, for some signal E, $v(E) \subset E$, or $E \subset v(E)$, or that E and $v(E)$ intersect in a more generic way. It may even be the case that E and $v(E)$ are disjoint.

The reason for calling this property "coherence" is obvious. Consider a signal A and let B be the set of $\succsim^A$-non-null states. We can say that, after receiving the signal A, the decision maker behaves as if she has observed the signal B. If this is the interpretation, it must be the case that the decision maker can correctly perceive the signal B.

Finally, I add the standard requirement that conditional preference are non-degenerate.

$\mathcal{A}3$ (Non-degeneracy) For any $E \in \Sigma$, $\exists f, g \in L : f \succ^E g$.

Axioms 1-3 lead to the following simple result.

[1] Results similar to those obtained in the chapter may hold in the more general case in which $H \subseteq \Sigma$. However, axioms must be modified to adapt to the new setting.

Proposition 2.2.1. *Given Axiom 1, conditional preferences satisfy Axioms 2 and 3 if and only if there exists a unique collection $\mathcal{V}$ of sets such that:*

(i) *For any $V \in \mathcal{V}$ and for any $E \in \Sigma$, E is $\succsim^V$-non-null if and only if $E \cap V \neq \emptyset$;*

(ii) *For any $E \in \Sigma$, $\exists!\ V^E \in \mathcal{V}$ such that $\forall f, g \in L$, $f \succsim^E g \iff f \succsim^{V^E} g$.*

Property (i) defines $\mathcal{V}$ as the set of fixed points of the function v. Setting $V^E = v(E)$, Proposition 2.2.1 is just a different way to express coherence.

2.3 Indistinguishable Signals

Imagine a decision maker who cannot distinguish between two or more signals. What may be a plausible reason? An interesting possibility is given by the following axiom.

$\mathcal{A}4$ (Strong Indistinguishability) Given a signal $A \in \Sigma$ and a state $s \in S$,

$$\big(\exists f, g \in L \text{ with } f(s') = g(s'),\ \forall s' \neq s \text{ s.th. } f \succ^A g\big)$$

$$\iff \left(\exists s'' \in A \text{ s.th. } \forall f, g \in L,\ f \succsim^{\{s\}} g \iff f \succsim^{\{s''\}} g\right).$$

The axiom says that, given a signal A, a state s is conditionally non-null if and only if there exists a state $s'' \in A$ such that the decision maker cannot distinguish between the signals $\{s\}$ and $\{s''\}$. Notice that I call two signals indistinguishable when updated preferences are the same conditional on any of them.[2] Consider the following example. An urn contains red, yellow, and black balls: $S = \{r, y, b\}$. The decision maker can correctly observe the color yellow, that is $v(\{y\}) = \{y\}$, but she cannot distinguish red from black: $v(\{r\}) = v(\{b\}) = \{r, b\}$. Suppose a ball is drawn and the decision maker receives a signal $\{r, y\}$, corresponding to the event "the ball is either yellow or red." By Axiom 4, she must consider all three colors as possible. Not only cannot the colors corresponding to the "true" signal be excluded, but the same must hold for any color indistinguishable from one of them. So $v(\{r, y\}) = \{r, y, b\}$.

Notice that Axiom 4 implies that all the states contained in a signal are non-null. Moreover, if there exists a signal E such that $v(E) \neq E$, the decision maker satisfies what I call "global underconfidence", as formalized in the following definition.

Definition 2.3.1 (Underconfidence)**.** Given the event $A \in \Sigma$, the decision maker is weakly underconfident if, for any $B \subseteq A$, $\exists f, g, h \in L$ such that $f|^h_{B^c} \succ^A g|^h_{B^c}$. If, in addition, there exists a set C with $A \cap C = \emptyset$ satisfying the same property, the decision maker is said to be underconfident. She is globally underconfident if she is weakly underconfident for any signal in Σ and underconfident for some signals.

[2] In what follows, I will say that two states are indistinguishable when the corresponding singletons are.

The question now is what kind of updating rules correspond to such an assumption. That is, I want to find the properties of the set of fixed points of v and to determine how, given a signal, conditional non-null states can be computed. The answer is given by the following proposition.

Proposition 2.3.1. *Given Axiom 1, conditional preferences satisfy Axiom 4 if and only if there exist a unique set $\mathcal{V} \subseteq \Sigma$ such that (i) and (ii) in Proposition 2.2.1 hold and*

(iii) *for any signal $E \in \Sigma$, $E \subseteq V^E$ and $\nexists\, V \in \mathcal{V} \;:\; E \subseteq V \subset V^E$;*

(iv) *$\mathcal{V} \cup \emptyset$ is an algebra.*

Notice that Axiom 4 implies both Coherence and Non-degeneracy, as is clear from the comparison of Propositions 2.2.1 and 2.3.1.

2.4 Dominated States

t may be interesting to look for updating rules which do not imply that the decision maker is always underconfident. In fact, it is not unreasonable to allow for signals after which she shows overconfidence, a behavior defined as follows.

Definition 2.4.1 (Overconfidence)**.** Given the event $A \in \Sigma$, we say that the decision maker is overconfident if

(i) there exists no B with $A \cap B = \emptyset$ such that $\exists f, g, h \in L$ with $f|^h_{B^c} \succ^A g|^h_{B^c}$;

(ii) there exists some $C \subset A$ such that, for any $f, g, h \in L$, $f|^h_{C^c} \sim^A g|^h_{C^c}$.

Simply put, overconfidence consists in the set of conditionally non-null states being a subset of the signal. What may be a reason for overconfidence? We can imagine that some event is so more vividly perceived by the decision maker compared to some other state, that, every time the signal does not exclude that event, the decision maker disregard the other state. In the urn example, suppose that the decision maker is so much impressed by the color red and so little but the color yellow that, when red is possible, she does not pay attention to the possibility of the drawn ball being yellow. So, for example, we have $v(\{r, y\}) = \{r\}$. This intuitive argument has an important implication: if, given the signal A, vividness considerations exclude state $s \in A$, then the same s must be excluded given any signal B such that $A \subset B$. In fact, if s is excluded because it's "dominated" by some event $E \subset A$ (that is, because E is "more vivid" than s), it must be excluded also given B, since $E \subset B$. Moreover, it is consistent with the intuition of relative vividness to assume that all states that are excluded from a signal are dominated by some set of conditionally non-null states. The restriction here is the requirement that the "dominating" states be conditionally non-null. This behavior is formally described in the following axiom.

$\mathcal{A}5$ (Relative vividness) Given a signal A, let B be the set of conditionally non-null states (as defined in Axiom 2). Then for any $s \in A \setminus B$

(i) for any $f, g \in L$, $f \succsim^B g$ if and only if $f \succsim^{B \cup \{s\}} g$;

(ii) for any signal A' such that $A \subset A'$, s is $\succsim^{A'}$-null.

Notice that Axiom 5 does not say that any $\succsim^A$-null state s must also be $\succsim^{B \cup \{s\}}$-null. This property has to hold just for $s \in A$. We can interpret (ii) as a sort of confidence monotonicity: the decision maker is relatively more confident for less informative signals. This must be interpreted in a very weak way, though. It is not necessarily the case that overconfidence for A implies overconfidence for any A' such that $A \subset A'$. However, if the decision maker is so "confident" that she can exclude some objectively possible state conditionally on a signal, she will still show this "confidence" conditionally on any less informative signal.

We can imagine a decision maker characterized by both "relative vividness" and "indistinguishability". Given a signal A, if she does not exclude some state $s \notin A$, the reason is that she cannot distinguish s from some $s' \in A$. On the other hand, if she exclude some $s'' \in A$, it means that the set of conditionally non-null state is more "vivid" than s''. To describe such a decision maker we need to add the following axioms.

$\mathcal{A}6$ (Indistinguishability) Given a signal A and given $s \notin A$ such that $\exists f, g, h \in L$ such that $f|^h_{s^c} \succ^A g|^h_{s^c}$, there exists $s' \in A$ such that, $\forall f, g \in L$, $f \succ^{\{s\}} g$ if and only if $f \succ^{\{s'\}} g$.

Axiom 6 says that if, given a signal A, some state s not included in A is conditionally non-null, then there there must exists a state in A that the decision maker cannot distinguish from s. Notice that this axiom is weaker than Axiom 4. Here we allow for the possibility that states indistinguishable from a state in A may still be conditionally null. A difficulty arises from the interaction of Axioms 5 and 6. It might be the case that some state $s' \notin A$ is non-null because indistinguishable from some $s \in A$, but s is null because of vividness considerations. To exclude this possibility, I introduce an additional axiom.

$\mathcal{A}7$ (Equi-vividness) Given a signal A, let the states $s \in S$ and $s' \in A$ be such that $\forall f, g \in L$, $f \succ^{\{s\}} g$ if and only if $f \succ^{\{s'\}} g$. Then $\exists f, g, h \in L$ such that $f|^h_{s^c} \succ^A g|^h_{s^c}$ if and only if the same is true for s'.

Axiom 7 makes two requirements. On one hand, it says that if a state $s \in A$ is conditionally non-null, then the same must be true for all the states that the decision maker cannot distinguish from s.[3] On the other hand, if s is null (because some other event is more vivid), all the indistinguishable states must be null, too. Informally, I want that indistinguishable states share the same vividness properties.

What kind of updating rules are implied by these axioms? The following proposition provides a result.

[3]This implies that, for states in A that are conditionally non-null, the converse of Axiom 6 holds.

Proposition 2.4.1. *Given Axioms 1 and 5, conditional preferences satisfy Axiom 2, 3, 6 and 7 if and only if there exists a unique set $\mathcal{V} \subseteq \Sigma$ such that (i) and (ii) in Proposition 2.2.1 hold and*

(iii) *$\mathcal{V} \cup \emptyset$ is a semialgebra; moreover, $\forall A, B \in \mathcal{V},\ A \setminus B \in \mathcal{V} \cup \emptyset$.*

(iv) *for any $A \in \Sigma$, V^A satisfies the conditions:*

1. *$V^A \cap A \neq \emptyset$;*
2. *$V^A \cap A \not\subset V \cap A,\ \forall\ V \in \mathcal{V}$;*
3. *$V^A \subseteq V,\ \forall\ V \in \mathcal{V}$ s.th. $V \cap A = V^A \cap A$.*

Given $\mathcal{V}$ and a signal A, property (iv) provides the restrictions that must be satisfied by the set of $\succsim^A$-non-null states. Notice that the proposition implies that, given a signal A, if $\mathcal{V}$ contains supersets of A, then V^A is the smallest of these supersets, where existence of a smallest superset is guaranteed by closure under intersection. A weakness of Proposition 2.4.1 is the absence of a sufficient condition for Axiom 5. Finding it is not a trivial task. In Proposition 2.3.1, there is a clear rule associating each signal to the set of conditionally non-null states. This set can be determined independently for each signal. Here, on the other hand, a single set $\mathcal{V}$ may allow for different preference relations: for many signals, there are multiple sets of conditionally non-null sets that are compatible with $\mathcal{V}$. Of course, not all the sets satisfying (iii) and (iv) above can be taken: the choice of V^A for some signal A constrains the correspondent choice for other signals. However, this is not a restriction that can be included in the proposition: it would be nothing more that a re-statement of Axiom 5.

The following corollary establishes a relation between the classes of indistinguishable signals and the respective sets of conditionally non-null states.

Corollary 2.4.1. *Take $V \in \mathcal{V}$ and let $A_1, ..., A_k$ be all the signals such that*

- *$V^{A_i} = V$;*
- *there is no signal $B \subset A_i$ such that $V^B = V$.*

Given Axiom 1, if conditional preferences satisfy Axiom 2, 3, 5-7, then $V = \bigcup_{i=1}^{k} A_i$.

2.5 Response to Information Increases

We have seen that the general model can accommodate for updating rules with very different properties. The same is true with respect to the problem of information acquisition. For example, the model is compatible with a setting in which "objectively" less informative signals provide more information to the decision maker. Consider the following example. There is an urn with red, yellow and green balls; so $S = \{r, y, g\}$. The function v is as follows:

$$v(\{y\}) = \{y\} \quad v(\{r, y\}) = \{r, y\} \quad v(\{y, g\}) = \{y, g\}$$

$$v(\{r\}) = v(\{g\}) = v(\{r, g\}) = v(\{r, y, g\}) = \{r, y, g\}.$$

We can imagine that the decision maker is normally sleepy, and expects an alarm clock to ring when a signal arrives. However, the alarm clock works only if the signal does not exclude the color y. When asked to make her choice, a decision maker who did not hear the alarm believes that no signal arrived. The result is that, for example, the more informative signal $\{r\}$ does not provide any information to the decision maker, while the less informative $\{r, y\}$ gets identified.

Let's now introduce a new axiom.

$\mathcal{A}8$ (Information monotonicity) Given $A, B \in \Sigma$, if for any $f, g, h \in L$, $f|^h_{B^c} \sim^A g|^h_{B^c}$, then for any $A' \subset A$, $f|^h_{B^c} \sim^{A'} g|^h_{B^c}$.

Axiom 8 states that if an event is null conditionally on a given signal, it must still be null conditionally to more informative signals.

In this section I will consider only updating rules that exhibit global underconfidence. Therefore, the following propositions do not provide a complete axiomatic characterization of classes of updating rules. They say that, given a globally underconfident decision maker, if we add some requirement on the way information is acquired, the updating rule exhibits some additional features. The following proposition shows the implications of Information monotonicity.

Proposition 2.5.1. *Given the representation as in Proposition 2.2.1 and global underconfidence, conditional preferences satisfy Axiom 8 if and only* $\forall A \in \Sigma$, *there is no* $V \in \mathcal{V}$ *such that* $A \subseteq V \subset V^A$. *Moreover, if Axiom 8 holds, then* $\mathcal{V}$ *is closed under intersection.*

Notice that $\mathcal{V}$ may be closed under intersection even if v does not satisfy monotonicity. It is also evident that Axiom 4 implies Information monotonicity.

More generally, the basic model allows for updating rules that exhibit non-constant returns to information. This is the interpretation I give to the behavior of the function v with respect to the union of signals. In general, given two signals A and B, there are no constraints on the relation between $v(A)$ and $v(B)$ on one side and $v(A \cup B)$ on the other. Consider the following example. Let $S = \{r, y, g, b\}$ and suppose

$$v(\{r\}) = \{r, g\} \quad v(\{y\}) = \{y, g\} \quad v(\{r, y\}) = \{r, y, b\},$$

while all the other signals are fixed points of v. The example clearly satisfied the assumptions in Proposition 2.2.1. However, neither $v(\{r, y\}) \subseteq (v(\{r\}) \cup v(\{y\}))$, nor $(v(\{r\}) \cup v(\{y\})) \subseteq v(\{r, y\})$. I say that the updating rule exhibits increasing returns to information if, for any signals A and B, $(v(A) \cup v(B)) \subset v(A \cup B)$. The idea is that when "objective" information increases (as when we move from the signal $A \cup B$ to either A or B), the amount of information that the decision maker retains increases "more than proportionally": both the signals A and B allow the decision maker to exclude some event that is not excluded when she receives the signal $A \cup B$. Analogously, I say that the updating rule exhibits decreasing returns to information if, for all signals A and B, $v(A \cup B) \subset (v(A) \cup v(B))$. In this case, when moving to a more informative signal, the decision maker may loose some information she was able to retain before (which does not mean she cannot at the same time acquire new information as well). Finally, if, for any signals A and B, $v(A \cup B) = (v(A) \cup v(B))$, I say that there are constant returns to information.

The following axiom reformulates the property of decreasing or constant returns in terms of preferences.

$\mathcal{A}9$ (Non-increasing returns to information) For any $A, B, C \in \Sigma$, if $\forall f, g, h \in L$, $f|_{C^c}^h \sim^A g|_{C^c}^h$ and $f|_{C^c}^h \sim^B g|_{C^c}^h$, then $f|_{C^c}^h \sim^{A \cup B} g|_{C^c}^h$.

The following proposition analyzes the relation between Axiom 8 and Axiom 9.

Proposition 2.5.2. *Given the representation as in Proposition 2.2.1, conditional preferences satisfy Axiom 8 if and only if the updating rule exhibits increasing or constant returns to information. If, in addition, global underconfidence is assumed, then:*

(i) *if conditional preferences satisfy Axiom 9, then $\mathcal{V}$ is closed under union;*

(ii) *if $\mathcal{V}$ is closed under complements and Axiom 8 holds, then the updating rule exhibits constant returns to information.*

Remark 1. Notice that the behavioral hypotheses introduced in this section imply some structural properties on $\mathcal{V}$, but they are not equivalent to them: Information monotonicity implies closure under intersection, while Decreasing returns to information lead to closure under union. The inverse implications do not hold. So in part (ii) it can be easily shown that constant returns to information are not implied if we substitute Axiom 9 for Axiom 8; that is, they are not a consequence of $\mathcal{V}$ being an algebra. It would be nice to find a characterization of the structural properties themselves. However, this does not seem to be possible, unless we directly assume the properties (which can clearly be defined in behavioral terms).

Finally, no more general axiom seems to imply closure under complements. In particular, constant returns to information do not imply that $\mathcal{V}$ is an algebra.

2.6 Conclusions

This chapter has introduced a new framework to look at phenomena of under- and overconfidence. Although quite simple, it allows to link confidence considerations to misperception issues. In particular, I have considered the cases of indistinguishable states and of vividness comparisons. The abstract nature of the framework can be a limit to its applicability, but it has the advantage of providing concepts and tools that can be used in more structured environments.

The emergence of some sort of confidence monotonicity seems to be one of the central questions that can be asked using my framework for under- and overconfidence. Can we claim that the level of confidence depends on how much information a signal provides? It would be interesting to be able to say that a decision maker is, in some sense, less confident for more informative signals than for less informative ones. That is, I would like to have a decision maker who, when she receives a lot of information, tends to discard some of it; on the other hand, when she receives little, she may view the signal as more informative than it actually is. In Section 2.4, I have assumed a weak form of such a monotonicity and I have provided a rationale for such a hypothesis. However, the present chapter does not address the question in a satisfying way. Nevertheless, I think it introduces the formal tools that will make it possible to look for a more complete answer. What we need is to construct a model in which confidence monotonicity emerges as a property of endogenously generated under- and overconfidence.

A first step towards such a model is provided by next Chapter, in which I build a model in which underconfidence and overconfidence are not assumed, but emerge endogenously.

Chapter 3

Optimal Underconfidence and Overconfidence

3.1 Introduction

Imagine a finite set of states of the world and suppose that an individual is uncertain about which of the states is the true one. Her beliefs can be represented with a probability distributions over the states. Suppose now that, before making her decision, she receives a signal. This signal provides a particular kind of information: it limits the set of states that can be realized, but it does not provide any additional information on the probability of the the states that are not excluded. We can therefore represent such signals as sets of states of the world. The individual, however, may misperceive the signal she has received, interpreting it as a set (of states) possibly different from the real one. She then updates her prior as if she has received this second signal.

What I have just described is the framework that I have introduced in Chapter 2. There, I imagine an external observer who can correctly observe the signal, but who does not know what signal the individual has actually perceived. Observing her behavior, he wants to determine the nature of the perceiving mistakes that she is making. Within this framework, it is possible to define notions of underconfidence and overconfidence in terms of relations between true and perceived signals. The individual exhibits underconfidence if the perceived signal is a superset of the true one; she is overconfident when the perceived signal is a subset of the one she has received.

The present chapter introduces a model in which the previous concepts may found an application. The main idea is the following. A decision maker has to choose how to allocate a unit among the different states of the world. If a state is realized, she will win the fraction that she has allocated to that state. For simplicity, assume that she is an expected utility maximizer, with a prior on the probability of the different states. Before making her choice, the decision maker receives a signal of the kind described above. The decision maker's beliefs include the likelihoods of the signals conditional on each state of the world. After receiving the signal, she updates her prior and chooses the allocation that maximizes expected utility.

There is, however, a complication. The decision maker is constrained in the number of different signals she can perceive. The constraint can be interpreted as a boundary on her computational capabilities. Each signal she may receive is interpreted as one of the signals she can actually perceive. The decision maker's beliefs are updated through Bayes rule given the actually perceived signal, not the one she truly receives.

Up to this point the model adds little to the framework in Chapter 2. The fundamental difference is the existence of an ex-ante stage. Ex ante, that is before receiving the signal, the decision maker knows that the number of possible true signals is higher than that of the signals she can distinguish. Given this computational constraint, she wants to choose the set of perceivable signals and the mapping from true signals to perceived ones that maximize her ex-ante expected utility.

The different steps in the model can be represented as in figure 3.1.

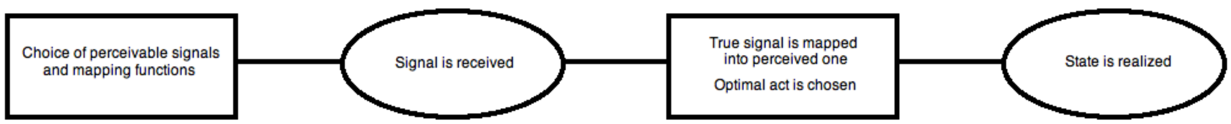

Figure 3.1: Steps

In a model like this, underconfidence and overconfidence may emerge endogenously. Different initial beliefs will determine different optimal sets of perceivable signals and different maps from true signals to perceived ones. In some of the cases, a signal A may be mapped into some signal $A' \supset A$, denoting underconfidence; for different beliefs, the same signal may be perceived as $A'' \subset A$, a case of overconfidence.

The purpose of the present chapter is to analyze this model and determine the properties of the optimal mapping from true to perceived signals. These are some of the questions that may be asked about the model:

- Is it really the case that both underconfidence and overconfidence emerge in the model?

- Is there some common feature characterizing the sets of perceivable signals that can be optimal under some initial belief?

- Does the decision maker's confidence on some signals have implications on her confidence on other signals?

Section 3.2 will introduce my notation and formally describe the model. Some general properties will be proved in Section 3.3. However, because of the mathematical complexity of the model, I will not reach in this chapter a satisfactory general characterization of the mapping. Some properties will be presented as conjectures. In order to get some intuition on these conjectures, it will be useful to study some special cases. This will be done in Sections 3.4 and 3.5. I will first consider the easiest possible case, that with only two states of the world. However, the analysis of this case will provide little insight on the properties of the mapping function. More considerations will be possible in the slightly more complicated case of three states of the world. The computational details of the analysis will be presented separately in Appendix C. Section 3.6 will contain some conclusive considerations.

3.2 The Model

Let $S = \{s_1, ..., s_n\}$ be a finite set of states of the world. Σ, the set of all subsets of S, is the set of events. A decision maker has a prior probability distribution μ over states. I limit the set of acts to all the possible allocation of one unit among the n states. Formally, if the set X of consequences is the set of non-negative real number smaller or equal to 1, an act is any function $f : S \to X$ such that $\sum_{s \in S} f(s) = 1$. Let F be the set of these acts. The decision maker is an expected utility maximizer with von Neumann-Morgenstern utility index $u : X \to \mathbb{R}$, where u is increasing in X and strictly concave.

Before making her choice, the decision maker receives a signal providing information on the true state of the world. More specifically, the signal limits the set of states that can be realized, but it does not provide any additional information on the probability of the the states that are not excluded. Therefore, each signal can be associated to an element of Σ: the true state of the world is one of those contained in the set. Moreover, I assume that any set in Σ describe a possible signal. With this assumption, I can use the same notation to denote both events and signals.

Let's now be more formal on the nature of the signals. The decision maker associates to the signals a likelihood function, which specifies the probability of each signal conditional on each state of the world. I will denote with $p_s(A)$ the probability of signal A conditional on the state s. The interpretation of signals as sets of states, as described above, is equivalent to the following assumptions on the likelihood function:

1. $\forall A \in \Sigma,\ p_s(A) > 0 \implies s \in A$;
2. $\forall A \in \Sigma,\ \forall s, s' \in A,\ p_s(A) = p_{s'}(A)$.

The first assumption says that signals must be truthful: if the signal A can be generated when the true state is s, then A must include s as one of the possible true states. The second assumption assures that no additional information is obtained from the signal other than the restriction of the set of possible states. It implies that, when updating the prior μ through Bayes rule, for any couple of states in A the ratio of the posterior probabilities is the same as the ratio of the priors. In addition, I will also assume that

3. $\forall A \in \Sigma,\ s \in A \implies p_s(A) > 0$,

so that, conditional on the state s, any signal that does not exclude s can be generated.

Given these assumption, the notation can be simplified noting that

$$p_s(A) = \begin{cases} p(A) > 0 & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}$$

To easily follow the rest of the construction of the model, keep in mind Figure 3.1 and its description in the Introduction. In the second stage of the decision process, before the decision maker chooses an act, the received (true) signal is mapped into a perceived signal. Let $\mathcal{V} \subset \Sigma$ be the set of perceivable signals. The true signal is interpreted as one of the sets in $\mathcal{V}$. Let $v : \Sigma \to \mathcal{V}$ be the mapping function. Therefore, after receiving a signal A, the decision maker actually perceives the signal $v(A)$. She then updates her prior, using Bayes rule, as if she had received $v(A)$. After this, she chooses the act that maximizes expected utility.

The set $\mathcal{V}$ and the function v are not exogenously given. Before receiving the signal, in the first stage of the decision process, the decision maker knows that there is a limit in the number of different signals she can perceive. Let $\bar{V}$ denote such upper limit. $\bar{V}$ and the decision maker's beliefs (prior and likelihoods) are the only exogenous parameters in the model. The decision maker knows that, in the second stage, she will behave as described in the previous paragraph. Ex ante, she wants to choose the set $\mathcal{V}$, with at most $\bar{V}$ elements, and the function v that, given her prior μ and likelihoods $p(\cdot)$, maximize her expected utility.

The optimization problem can be expressed formally as follows. In the second stage, after receiving a signal A, the decision maker solves

$$\max_{f \in F} \left\{ \sum_{s \in S} u(f(s))\mu(s|v(A)) \right\} \qquad \text{where} \qquad \mu(s|v(A)) = \frac{p_s(v(A))\mu(s)}{\sum_{s' \in S} p_{s'}(v(A))\mu(s')}$$

Since

$$\mu(s|v(A)) = \begin{cases} \frac{\mu(s)}{\sum_{s' \in v(A)} \mu(s')} & \text{if } s \in v(A) \\ 0 & \text{if } s \notin v(A) \end{cases}$$

the problem can be simplified to

$$\max_{f \in F} \left\{ \sum_{s \in v(A)} u(f(s))\mu(s) \right\}$$

Let $f^*_{(v,A)}$ be the act that solves the problem. Since u is increasing and strictly concave, $f^*_{(v,A)}(s) = 0$ for any $s \notin v(A)$ and $f^*_{(v,A)}(s) > 0$ for any $s \in v(A)$. From an ex-ante point of view, the expected utility that the decision maker gets after receiving the signal A and given the map v is

$$EU(v, A) = \sum_{s \in S} u(f^*_{(v,A)}(s))\mu(s|A)$$

or, equivalently,

$$EU(v,A) = \frac{1}{\sum_{s'\in A}\mu(s')}\sum_{s\in A} u(f^*_{(v,A)}(s))\mu(s)$$

In the first stage, the decision maker chooses a function v such that the cardinality of $v(\Sigma)$ is $\bar{V}$, in order to maximize the ex ante expected utility. Let $\{v_{\bar{V}}\}$ denote the set of functions v satisfying the cardinality requirement. The problem to solve is then

$$\max_{v\in\{v_{\bar{V}}\}} \left\{ \sum_{A\in\Sigma} \left(EU(v,A) \sum_{s\in S} p_s(A)\mu(s) \right) \right\}$$

or, equivalently, simply

$$\max_{v\in\{v_{\bar{V}}\}} \left\{ \sum_{A\in\Sigma} \left(EU(v,A) \sum_{s\in A} p(A)\mu(s) \right) \right\}$$

Although quite intuitive, the model is computationally cumbersome even for small numbers of states. However, some general properties can be proved, as shown in the next section.

3.3 General Results and Conjectures

3.3.1 Properties of optimal sets $\mathcal{V}$

Not all possible sets $\mathcal{V}$ can be optimal for some initial beliefs. The following proposition states a property that can be easily proved.

Proposition 3.3.1. *When $\bar{V} = 1$, the function v such that $v(E) = S$ for all $E \in \Sigma$ is always optimal.*

Proof is straightforward: when updated beliefs are the same no matter the signal received, the optimal thing to do is to consider the signals uninformative; that is, to behave as if S has been received.[1] For generic values of $\bar{V}$, other restriction likely apply. At the present stage I do not have proofs for these properties. However, the analysis of the case with three states of the world, described in Section 3.5 support the following conjectures. The first one can be seen as a generalization of Proposition 3.3.1 for the case of generic $\bar{V}$.

Conjecture 3.3.1. *For any initial beliefs, there exists an optimal set $\mathcal{V}$ of observable signals such that $\bigcup\{E\colon E \in \mathcal{V}\} = S$. If $u'(0) = \infty$, then a set $\mathcal{V}$ can be optimal only if it satisfies this property.*

[1] Such v is the only optimal mapping if the decision maker is sufficiently risk averse, so that she allocates to each state of the world a positive share of the unit.

Is it the case that any $\mathcal{V}$ satisfying this property can be optimal for some initial beliefs? The analysis in Section 3.5 seems to exclude this possibility. On the other hand, it may support an additional conjecture.

Conjecture 3.3.2. *Let $u'(0) = \infty$. Any set $\mathcal{V}$ such that $\bigcup\{E\colon E \in \mathcal{V}\} = S$ can be optimal for some initial beliefs, unless $S \notin \mathcal{V}$ and there exists a state s such that, for any $V \in \mathcal{V}$, $s \in V$.*

In fact, with three states of the world, the only set that apparently cannot be optimal is $\mathcal{V} = \{\{s_1\}, \{s_1, s_2\}, \{s_1, s_3\}\}$.

3.3.2 Properties of optimal functions v

There are other properties of optimal functions v that can be proved, which do not translate immediately into features of the set $\mathcal{V}$. First, it is immediate to see that there always exists an optimal function v mapping any signal into some fixed point of v itself. This property was called Coherence in Chapter 2. Its interpretation is straightforward: any signal in the set $\mathcal{V}$ of perceivable signals can be correctly observed when received.

Proposition 3.3.2. *Given any prior μ, likelihoods p and utility index $u(.)$, there exists an optimal mapping v such that, for any $E \in \Sigma$, $v(E) = v(v(E))$.*

Proof. Consider a mapping v such that there exists a signal $A \in \Sigma$ such that $v(A) \neq v(v(A))$. Consider now the map v' defined as

$$v'(E) = \begin{cases} v(E) & \text{if } E \neq v(A) \\ v(A) & \text{if } E = v(A) \end{cases}$$

Notice that $EU(v', v(A)) \geq EU(v, v(A))$, while $EU(v', E) = EU_2(v, E)$ for any $E \neq v(A)$. Denote with $\mathcal{V}_v$ the set $\mathcal{V}$ corresponding to the function v. It is clear that $\mathcal{V}_{v'} \subseteq \mathcal{V}_v$, so that v' continues to satisfy the constraint on the cardinality of $\mathcal{V}$. This is enough to prove the result. □

Assuming that Conjecture 3.3.1 be true, it is easy to verify that there always exists an optimal v such that, for any received signal E, the corresponding perceived signal $v(E)$ must include some of the states in E. The next proposition formulates this property in more general terms.

Proposition 3.3.3. *For any $E \in \Sigma$, if there exists a set $V \in \mathcal{V}$ such that $V \cap E \neq \emptyset$, then there is an optimal mapping v for which $v(E) \cap E \neq \emptyset$.*[2]

Proof. Consider a mapping v such that there exists a signal A such that $v(A) \cap A = \emptyset$. Let $V \in \mathcal{V}$ be such that $V \cap A \neq \emptyset$. Consider now the map v' defined as

$$v'(E) = \begin{cases} v(E) & \text{if } E \neq A \\ V & \text{if } E = A \end{cases}$$

[2] Notice that the hypothesis would be satisfied if Conjecture 3.3.1 were true.

Notice that $EU(v', A) \geq EU(v, A)$, while $EU(v', E) = EU(v, E)$ for any $E \neq A$. Moreover, $\mathcal{V}_{v'} = \mathcal{V}_v$. This proves the result. □

Following the approach in Chapter 2, we can define the concept of underconfidence and overconfidence. The decision maker is underconfident given the signal E when $v(E) \supset E$; she is globally underconfident when, for any signal E, $v(E) \supseteq E$ and the inclusion is strict for some signal. Similarly, the decision maker exhibits overconfidence given the signal E when $v(E) \subset E$; she is globally underconfident when, for any signal E, $v(E) \subseteq E$ and the inclusion is strict for some signal. An underconfident decision maker does not "trust" the information received: she behaves as if she has observed a signal less informative than the one received; that is, one including a superset of the states. Overconfidence is the opposite situation: the decision maker interprets the signal as more informative than it really is, excluding some states of the world that, according to the received signal, should be considered possible. Clearly, we can also think of a function v such that, for some signal A, $v(A)$ is neither a superset nor a subset of A. In this case, the decision maker is neither under- nor overconfident.

The present model is compatible with both underconfidence and overconfidence. It will be shown in Section 3.4 that both behaviors can emerge endogenously from the optimization problem. Which behavior is observed depends on the decision maker's initial beliefs (prior and likelihoods).

3.3.3 Further observations

It may be interesting to verify whether the optimal functions v have to satisfy some conditions that could be interpreted as fixed patterns underlying the misperception of signals. Unfortunately, it will be shown that the model does not generally support such patterns. Some useful consideration can however be obtained from these negative results.

It has been experimentally shown that people tend to be overconfident in the information they possess when facing difficult tasks; their confidence level declines for easier ones, possibly leading to underconfidence. See, for example, Lichtenstein and Fischhoff 1977 [8], Griffin and Tversky 1992 [10], or Erev, Wallsten, and Budescu 1994 [5]. In the context of the present model, we can consider the choice of the optimal act as a simple task when the signal received is very informative, and as a more difficult one when the signal is less informative. It would be a nice feature of the model if the confidence that the decision maker shows on a signal decreased as the information it conveys increases. There is not a unique way to formalize this idea. One possibility is to require that, given two signals A and A', with $A' \subset A$, and given a state $s \in A'$ such that $s \notin v(A')$, it cannot be the case that $\{s\} \cup (A' \cap v(A')) \subseteq v(A)$. That is, if the decision maker, after receiving the signal A', is so confident that she exclude the state $s \in A'$, then she must be at least as confident when receiving the less informative signal A. We can call this property "Monotone overconfidence".[3]

[3]This property is a weakening of the second point in the definition of Relative vividness, as introduced in Chapter 2.

Does the decision maker in the present model exhibit Monotone overconfidence? In general, this turns out to be false, as shown by the following example.

Example 3.3.1. Suppose that $S = \{s_1, s_2, s_3\}$. As explained in Section 3.5, initial beliefs can be represented with a vector of six parameters: $\mu(s_1)$, $\mu(s_2)$, $p(\{s_1\})$, $p(\{s_2\})$, $p(\{s_3\})$, and $p(\{s_S\})$. Let $u(x) = \sqrt{x}$ and $\bar{V} = 2$, and consider the following values:

$$\mu(s_1) = \frac{2}{11} \quad \mu(s_2) = \frac{2}{11}$$

$$p(\{s_1\}) = \frac{4}{9} \quad p(\{s_2\}) = \frac{1}{6} \quad p(\{s_3\}) = \frac{1}{6} \quad p(\{S\}) = \frac{1}{2}$$

Using the program in Appendix C to compute the optimal function v, we find that the optimal set of observable signals is $\mathcal{V} = \{\{s_1\}, S\}$. Monotone overconfidence would require $v(\{s_1, s_2\} = S$. However, for these specific initial beliefs, this is not the case.

So, although functions v satisfying monotone overconfidence can be optimal for some initial beliefs, this property does not hold in general. This discomforting result can nonetheless teach us something. The pattern of under- and overconfidence commonly observed cannot be the result of the optimization of limited computational capabilities. Other behavioral consideration must be involved in an explanation of the phenomenon.

I now go over some of the properties introduced in the previous chapter. The analysis will give us insight on the nature of the decision maker's perception mistakes. One interesting properties is the one called "Indistinguishability". To define it in the context of the present model, let's say that two states s_1 and s_2 are indistinguishable when the corresponding singleton signals $\{s_1\}$ and $\{s_2\}$ are mapped into the same perceived signal. Now suppose that for some signal A, there exists a state $s \in v(A) \cap A^c$. Indistinguishability requires the existence of a state $s' \in A$ such that s and s' are indistinguishable. The interpretation is that, if the decision maker, after receiving the signal A, still considers some state not in A to be possible, this must be due to the fact that she cannot distinguish that state from some other state included in A. This is a reasonable property if we want to interpret the difference between received and perceived signals as arising not from a computational constraint, as done in this model, but from limited perception capabilities.

It is immediate to show that Indistinguishability is not satisfied by the model. A counterexample can be found analyzing the case with two states of the world s_1 and s_2, and $\bar{V} = 2$, as done in Section 3.4. For some initial beliefs, the set $\mathcal{V} = \{\{s_1\}, \{s_1, s_2\}\}$ may be optimal. In this case, $v(\{s_2\}) = \{s_1, s_2\}$, but $v(\{s_1\}) = \{s_1\}$. So, although $s_1 \in v(\{s_2\})$, s_1 and s_2 are not indistinguishable. We can conclude that the behavior of a decision maker in this model cannot in general be interpreted as the result of limited perception capabilities.

An entire section in Chapter 2 was devoted to studying how the decision maker reacts to signals differing in the amount of information. Monotone confidence, as defined above, is a property included in this category. In that chapter I suggested two other properties. The first was Information monotonicity. Consider two received signals A and A' such that $A' \subset A$; that is, A' is more informative than A. The decision maker's behavior satisfies Information monotonicity if $v(A') \subset v(A)$: no information is lost when moving from a less informative signal to a more informative one.

Information monotonicity is not guaranteed in the present model. We will se in Section 3.5 that, given three states of the world s_1, s_2, and s_3, and given $\bar{V} = 2$, there are initial beliefs such that the set $\mathcal{V} = \{\{s_1, s_2\}\{s_2, s_3\}\}$ is optimal. In this case, $v(\{s_1, s_2\}) = \{s_1, s_2\}$ and $v(\{s_2, s_3\}) = \{s_2, s_3\}$. Clearly, $v(\{s_2\})$ cannot be a subset of both $v(\{s_1, s_2\})$ and $v(\{s_2, s_3\})$, so that Information monotonicity is not satisfied.

We can think of a weaker result, and requiring that a more informative received signal cannot be mapped into a less informative one. Formally, given two signals A and A' with $A' \subset A$, we may want that no optimal v can be such that $v(A) \subset v(A')$.[4] Even this property turns out to be false in general.

Example 3.3.2. Suppose that $S = \{s_1, s_2, s_3\}$. Let $u(x) = \sqrt{x}$ and $\bar{V} = 2$, and consider the following values:

$$\mu(s_1) = \frac{2}{5} \quad \mu(s_2) = \frac{1}{3}$$

$$p(\{s_1\}) = \frac{7}{16} \quad p(\{s_2\}) = \frac{3}{8} \quad p(\{s_3\}) = \frac{3}{8} \quad p(\{S\}) = \frac{1}{2}$$

It turns out that the optimal set of observable signals is $\mathcal{V} = \{\{s_2\}, S\}$ and, in particular, $v(\{s_2, s_3\}) = \{s_2\}$ and $v(\{s_3\}) = S$. This violates the weak monotonicity property.

Another aspect studied in Chapter 2 was the problem of returns of information. The decision maker exhibits increasing returns to information if, for any signals A and B, $(v(A) \cup v(B)) \subset v(A \cup B)$. The idea is that when "objective" information increases (as when we move from the signal $A \cup B$ to either A or B), the amount of information that the decision maker retains increases "more than proportionally": both the signals A and B allow to exclude some event that is not excluded when she receives the signal $A \cup B$. Analogously, the decision maker exhibits decreasing returns to information if, for all signals A and B, $v(A \cup B) \subset (v(A) \cup v(B))$. In this case, when moving to a more informative signal, she may loose some information that she was able to retain before. Finally, if, for any signals A and B, $v(A \cup B) = (v(A) \cup v(B))$, we have constant returns to information.

It is possible to show that no one of these properties can hold in general in the model: for different initial beliefs, the decision maker can exhibit both increasing and decreasing returns.

Example 3.3.3. Suppose again that $S = \{s_1, s_2, s_3\}$. Let $u(x) = \sqrt{x}$ and $\bar{V} = 3$, and consider the following values:

$$\mu(s_1) = \frac{2}{5} \quad \mu(s_2) = \frac{2}{5}$$

[4] Or, should Conjecture 3.3.1 be false, that there always exists an optimal v for which $v(A) \subset v(A')$ does not hold.

$$p(\{s_1\}) = \frac{3}{8} \quad p(\{s_2\}) = \frac{3}{8} \quad p(\{s_3\}) = \frac{3}{8} \quad p(\{S\}) = \frac{1}{2}$$

It turns out that the optimal set of observable signals is $\mathcal{V} = \{\{s_1\}, \{s_2\}, S\}$ and, in particular, $v(\{s_1\}) = \{s_1\}$, $v(\{s_2\}) = \{s_2\}$ and $v(\{s_1, s_2\}) = S$. That is, we observe increasing returns to information.

Example 3.3.4. Let $S = \{s_1, s_2, s_3\}$, $u(x) = \sqrt{(}x$ and $\bar{V} = 3$. Consider the following initial beliefs:

$$\mu(s_1) = \frac{2}{5} \quad \mu(s_2) = \frac{2}{5}$$

$$p(\{s_1\}) = \frac{1}{5} \quad p(\{s_2\}) = \frac{1}{5} \quad p(\{s_3\}) = \frac{1}{5} \quad p(\{S\}) = \frac{1}{5}$$

Computations give $\mathcal{V} = \{\{s_1, s_2\}, \{s_1, s_3\}, \{s_2, s_3\}\}$ as the optimal set of observable signals, while $v(\{s_1, s_2\}) = \{s_1, s_2\}$, $v(\{s_1\}) = \{s_1, s_3\}$, and $v(\{s_2\}) = \{s_2, s_3\}$. Therefore, there are decreasing returns to information.

3.4 The Case of Two States

Assume that $S = \{s_1, s_2\}$, so that $|\Sigma| = 3$. Consider the case $\bar{V} = 2$. At the ex-ante stage, the decision maker has to choose the optimal set $\mathcal{V}$ of perceivable signals and the corresponding optimal map v from true signals to perceived ones. I will assume that her preferences can be described with a von Neumann-Morgenstern utility index $u(x) = \sqrt{x}$.

To simplify the notation, let $\mu(s_1) = \mu$ and $p_{s_1}(S) = p_{s_2}(S) = 1 - \alpha$; it follows that $p_{s_1}(\{s_1\}) = p_{s_2}(\{s_2\}) = \alpha$. The function f_i^A, which gives the optimal share to be allocated to state s_i given the perceived signal A. becomes

$$f_1^{\{s_1\}} = 1 \quad f_1^{\{s_2\}} = 0 \quad f_1^S = \frac{\mu^2}{\mu^2 + (1-\mu)^2}$$

$$f_2^A = 1 - f_1^A, \ \forall A \in |\Sigma|$$

With $\bar{V} = 2$, there are three possible sets $\mathcal{V}$:

$$\mathcal{V}_1 = \{\{s_1\}, S\} \quad \mathcal{V}_2 = \{\{s_2\}, S\} \quad \mathcal{V}_3 = \{\{s_1\}, \{s_2\}\}$$

The first step to solve the for the optimal $\mathcal{V}$ is to determine the optimal function v conditional on each $\mathcal{V}$. In this special case, the task is easy. It is straightforward to realize that

- for $\mathcal{V} = \mathcal{V}_1$, $v(\{s_1\}) = \{s_1\}$ and $v(\{s_2\}) = v(S) = S$;
- for $\mathcal{V} = \mathcal{V}_2$, $v(\{s_2\}) = \{s_2\}$ and $v(\{s_1\}) = v(S) = S$;
- for $\mathcal{V} = \mathcal{V}_3$, $v(\{s_1\}) = \{s_1\}$, $v(\{s_2\}) = \{s_2\}$ and $v(S) = \begin{cases} \{s_1\} & \text{if } \mu \geq \frac{1}{2} \\ \{s_2\} & \text{if } \mu < \frac{1}{2} \end{cases}$

Notice that, for $\mathcal{V} = \mathcal{V}_1$, the decision maker is globally underconfident (she correctly perceives the signals $\{s_1\}$ and S, but she interprets $\{s_2\}$ as S; similarly when $\mathcal{V} = \mathcal{V}_2$.[5] On the other hand, she is globally overconfident for $\mathcal{V} = \mathcal{V}_3$: she correctly perceives $\{s_1\}$ and $\{s_2\}$, but interprets S as either $\{s_1\}$ or $\{s_2\}$.

At this point we can compute the ex-ante expected utility for each $\mathcal{V}$ when the corresponding optimal v is used. Let $U^*_{\mathcal{V}_i}$ be the ex-ante expected utility for $\mathcal{V}_i$. Clearly

- $U^*_{\mathcal{V}_1} = \alpha\mu + \alpha(1-\mu)\sqrt{\frac{(1-\mu)^2}{\mu^2+(1-\mu)^2}} + (1-\alpha)\left[\mu\sqrt{\frac{\mu^2}{\mu^2+(1-\mu)^2}} + (1-\mu)\sqrt{\frac{(1-\mu)^2}{\mu^2+(1-\mu)^2}}\right]$
- $U^*_{\mathcal{V}_2} = \alpha\mu\sqrt{\frac{\mu^2}{\mu^2+(1-\mu)^2}} + \alpha(1-\mu) + (1-\alpha)\left[\mu\sqrt{\frac{\mu^2}{\mu^2+(1-\mu)^2}} + (1-\mu)\sqrt{\frac{(1-\mu)^2}{\mu^2+(1-\mu)^2}}\right]$
- $U^*_{\mathcal{V}_3} = \alpha\mu + \alpha(1-\mu) + (1-\alpha)\max\{\mu, 1-\mu\}$

Some simple algebra allows to determine the optimal $\mathcal{V}$ as a function of the decision maker's initial beliefs:

- $U^*_{\mathcal{V}_1} \geq U^*_{\mathcal{V}_2} \iff \mu \geq \frac{1}{2}$,
- when $\mu \geq \frac{1}{2}$, $U^*_{\mathcal{V}_3} \geq U^*_{\mathcal{V}_1} \iff \alpha \geq h(\mu)$,
- when $\mu \leq \frac{1}{2}$, $U^*_{\mathcal{V}_3} \geq U^*_{\mathcal{V}_2} \iff \alpha \geq h(1-\mu)$,

where

$$h(\mu) = \frac{\mu^2 + (1-\mu)^2 - \mu\sqrt{\mu^2 + (1-\mu)^2}}{\mu^2 - 2\mu + 1}.$$

This result can be represented graphically as in Figure 3.2.

The special case with two states of the world shows that both underconfidence and overconfidence can emerge endogenously, depending on the decision maker's initial beliefs. However, the case is too simple to provide more insight into the properties of the model. For example, all the possible sets $\mathcal{V}$ are such that the union of their elements in the set S. Therefore, it is not possible to study the conjectured sub-optimality of the sets for which such a property does not hold. Moreover, the small number of sets $\mathcal{V}$ does not allow to make observations on the conjectured monotonicity properties of confidence. This is why it is interesting to study a slightly more complicated case, that with three possible states.

3.5 The Case of Three States

Given the conjectural status of some of the properties discussed in Section 3.3, it is useful to look for evidence in support of the conjecture through the analysis of special cases. We have see that a model with two states of the world does not allow to address these issues. More insight can be gained from studying a model with three states.

[5] Moreover, in these cases, the decision maker does not satisfy Indistinguishability, as defined in Section 3.3.

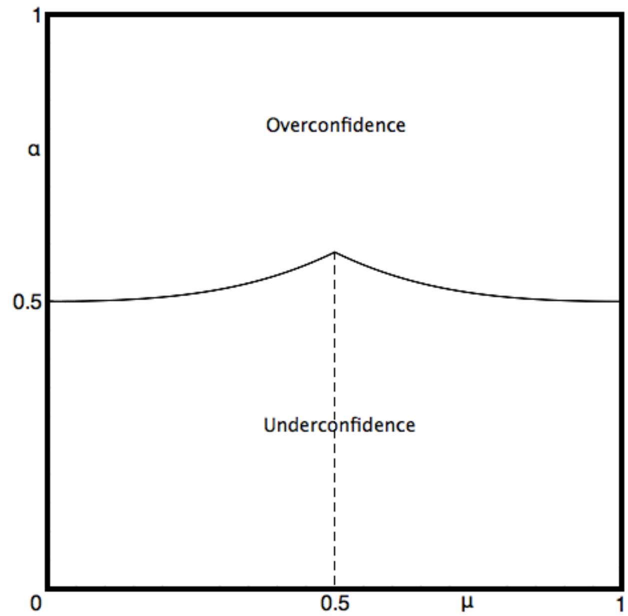

Figure 3.2: Under- and Overconfidence with Two States

Unfortunately, the analysis in this case is computationally much more complicated than in the model with two states. First, to describe initial beliefs six parameters are needed: $\mu(s_1)$, $\mu(s_2)$, $p(\{s_1\})$, $p(\{s_2\})$, $p(\{s_3\})$, and $p(\{s_S\})$.[6] Moreover, $|\Sigma| = 7$, implying that different values of $\bar{V}$ must be considered (two, three, four, five, and six).

To overcome this difficulties, I have used Matlab to solve the problem numerically. The main steps of the program are the following:

- fixing specific parameters values, the optimal act is computed for any possible signal observed (this step can be easily solved analytically);
- for any set $\mathcal{V}$ of observable signals and for any signal received, the program finds the corresponding signal in $\mathcal{V}$ maximizing (ex-post) expected utility;
- it is then possible to compute, for each $\mathcal{V}$, the corresponding ex-ante expected utility; the program selects the set $\mathcal{V}$ for which utility is maximized.

In this way, the program finds the optimal $\mathcal{V}$ for a specific value of the decision maker's initial beliefs. A more general problem is to determine whether a given $\mathcal{V}$ can be optimal for some values of the parameters. To do this, I have constructed a

[6]The remaining values can be derived as follows:
$\mu(s_3) = 1 - \mu(s_1) - \mu(s_2)$,
$p(\{s_1, s_2\}) = \frac{1-p(\{s_1\})-p(\{s_2\})+p(\{s_3\})-p(\{s_S\})}{2}$,
$p(\{s_1, s_3\}) = \frac{1-p(\{s_1\})+p(\{s_2\})-p(\{s_3\})-p(\{s_S\})}{2}$,
$p(\{s_2, s_3\}) = \frac{1+p(\{s_1\})-p(\{s_2\})-p(\{s_3\})-p(\{s_S\})}{2}$.

function that, after computing the optimal $\mathcal{V}$, assigns the value 0 to the one under consideration, and 1 to all the others. Minimizing this function over the possible values of the parameters, it is possible to see whether that particular set $\mathcal{V}$ can ever be optimal: this is case when the minimum of the function is 0. A detailed description of the program can be found in Appendix C.

Using this algorithm, I find the following results:

- only sets $\mathcal{V}$ such that $\bigcup\{E \colon E \in \mathcal{V}\} = S$ can be optimal for some initial beliefs;
- all the sets $\mathcal{V}$ satisfying this property are optimal for some values of the parameters, except $\{\{s_1\}, \{s_1, s_2\}, \{s_1, s_3\}\}$.

Appendix C gives example of initial beliefs that make each set $\mathcal{V}$ optimal. The results support the conjectures in Section 3.3.

However, there is a caveat. The algorithm use for the maximization problem looks for local maxima only. Although I used a grid of initial values for the six parameters, the possibility of having missed some optimal solutions cannot be excluded.

3.6 Conclusions

This chapter has introduces a model in which a decision maker is aware that she cannot extract all the information from the observation of events. She knows that the number of possible signals is larger than the number of different pieces of information she can handle. At an ex-ante stage, she therefore wants to optimally map "true" signals into "perceived" one, so that to maximize her ex-ante expected utility.

The analysis of optimal mapping performed in this chapter is clearly partial. Few results have been established, while the main properties are just presented as conjectures. Although intuitively simple, the model is mathematically quite complex. Of course, better results can be obtained with a deeper analysis. In particular, Conjectures 3.3.1 and 3.3.2 are clearly provable (or disprovable).

What emerges, however, is the presence of very limited structure. An optimal mapping does not necessarily exhibit any kind of "confidence monotonicity" or even "information monotonicity" (see Subsection 3.3.3). However, this does not exclude the possibility of the emergence of more structure once the set of possible initial beliefs is reduced. I am thinking in particular to restriction on the values of the likelihood functions p, which may be reasonable in particular contexts. Studying whether there exist relations between restrictions on p and structural properties of the optimal mapping may be an interesting area of future research.

Appendix A

Proofs for Chapter 1

A.1 Proof of Theorem 1.4.1

The following simple lemma has a crucial role in the proof.

Lemma A.1.1. *Consider $x, y \in (0,1)^n$ with $x(n) \neq y(n)$ and $\{\alpha_i\}_{i=1}^{n-1}$, where $\alpha_i > 0$ and $\sum_{i=1}^{n-1} \alpha_i < 1$. There exist scalars $\{k_i\}_{i=0}^{n}$ such that*

$$\prod_{i=1}^{n-1} x_i^{\alpha_i} x_n^{1-\sum \alpha_i} \gtreqless \prod_{i=1}^{n-1} y_i^{\alpha_i} y_n^{1-\sum \alpha_i} \iff k_0 + \sum_{i=1}^{n-1} k_i \alpha_i \gtreqless 0.$$

Proof. Simple algebra shows that the relation is satisfied for any $(k_0, ..., k_{n-1})$ such that

$$\frac{k_i}{k_0} = \frac{\ln x_i - \ln x_n - \ln y_i + \ln y_n}{\ln x_n - \ln y_n}, i = 1, ..., n-1.$$

□

What the lemma says is that, given a likelihood function (which is represented by the vectors x and y), there exists a unique hyperplane separating the limit frequencies for which beliefs converge to $(1,0)$ and those for which beliefs converge to $(0,1)$. Therefore, any likelihood function can be associated with a hyperplane.

Consider the case $\mu \in \Delta(\{\mathcal{L}_1, \mathcal{L}_2\})$. Assume that there exist two limit frequencies $r^1, r^2 \in int\Delta(S)$ such that

$$r_t \to r^1 \Rightarrow \mu_t \to \delta_{\mathcal{L}_1},$$

$$r_t \to r^2 \Rightarrow \mu_t \to \delta_{\mathcal{L}_2}.$$

If this is not true, then there is nothing to prove.

Consider any recursive updating rule satisfying the conditions of the theorem. Suppose that $\mu = (1,0)$ or $\mu = (0,1)$ can be reached in finite time with positive probability; in what follows I consider the first case, but the argument clearly applies to the other, too. There are two cases: either there exists $s \in S$ such that $f(s,(1,0)) \neq (1,0)$, or no such outcome exists. The second case can be easily dismissed. In fact, given any limit frequencies, the probability of observing the sequence of outcomes leading to $\mu = (1,0)$ is positive; but since there must exist limit frequencies with different limit beliefs, we are led to a contradiction. In order to get rid of the first case, consider a limit frequency for which we expect convergence to $\mu = (1,0)$. Clearly, as μ approaches $(1,0)$, the probability of reaching $(1,0)$ in finite time must go to zero. Otherwise, there would not be a t such that for all $\epsilon, \delta > 0$, $prob(\mu_{t'} \notin (1, 1-\epsilon)) < \delta, \forall t' > t$. Therefore, if the only thing we are interested in is convergence, we can disregard the possibility of reaching $\mu = (1,0)$ in finite time: if an updating mechanism exists for which μ converges to $(1,0)$, then convergence has to be possible for some mechanism in which $(1,0)$ is not reached in finite time. So, if we prove that no mechanism of this last type exists, we can conclude that the appropriate convergence cannot be obtained with any recursive mechanism.

Once we exclude the previous cases, μ_{t+1} can be obtained from μ_t multiplying each component μ_i by a positive scalar. Formally, there exists a function $v : (S, \Delta(S)) \to \mathbb{R}^2_{++}$ such that

$$f(s,\mu) = \left(\frac{v(s,\mu)_i \mu_i}{\sum_i v(s,\mu)_i \mu_i} \right)^2_{i=1}.$$

Moreover, we can always use the normalization $\sum_i v(s,\mu)_i = 1$ (so that each f is associated to a unique v).

The following lemma shows the consequences for the function v of assuming that f is well-behaved. It turns out that v has a limit when beliefs approach certainty. Although it would be nice to get rid of this hypothesis, it's still unclear whether the theorem holds in the more general case.

Lemma A.1.2. *If f is well-behaved, then $v(., \mu)$ converges to a limit function for $\mu \to (1,0)$ and for $\mu \to (0,1)$.*

Proof. Convergence of $v(., \mu)$ means convergence of $v(s, \mu)$ for any $s \in S$. Let's consider a generic s. To simplify notation, I will drop the argument s in the proof of this lemma. Therefore, $f(\mu)$ and $v(\mu)$ must be interpreted as $f(s,\mu)$ and $v(s,\mu)$ respectively. In addition, I will redefine $\mu \in [0,1]$ as the probability associated to $\mathcal{L}_1$. Define $z(\mu) := \frac{v_1(\mu)}{v_2(\mu)}$. Then we can write

$$f(\mu) = \frac{z(\mu)\mu}{z(\mu)\mu + (1-\mu)}$$

or, equivalently,

$$z(\mu) = \frac{\frac{1-\mu}{\mu} f(\mu)}{1 - f(\mu)}.$$

Let's consider the case in which $\mu \to 1$; to address the other case, just take the limit of $z(\mu)^{-1}$ as $\mu \to 1$. Computing the limit, we get

$$\lim_{\mu \to 1} z(\mu) = \lim_{\mu \to 1} \frac{\frac{1-\mu}{\mu} f(\mu)}{1 - f(\mu)} =$$

$$= \lim_{\mu \to 1} \frac{\frac{f(\mu)}{\mu^2} - f'(\mu)\frac{1-\mu}{\mu}}{f'(\mu)} = \lim_{\mu \to 1} f'(\mu)^{-1}.$$

Since f is well-behaved, this limit exists (not necessarily finite). □

By Lemma A.1.2, as $\mu \to (1,0)$ or $\mu \to (0,1)$, $v(.,\mu)$ converges to a limit function. Both limits are, according to the definition in Section 3, likelihood functions. As we already know, such functions can be associated to hyperplanes separating the half-spaces of limit frequencies for which μ_t converges to $(1,0)$ or $(0,1)$ if updating is characterized by the likelihood function itself. We now need to prove the following results:

- Consider a limit frequency such that beliefs converge to $(1,0)$. I claim that the same convergence result must hold if updating is performed using the limit likelihood function for $\mu \to (1,0)$.
- Consider any limit frequency belonging to the half-space where beliefs converge to $(1,0)$ according to the limit likelihood function for $\mu \to (1,0)$. Clearly, there exists $\epsilon > 0$ such that for any $\epsilon' < \epsilon$, the limit frequency lays on the same "side" with respect to the hyperplanes associated to the functions $v(.,\mu)$ for $\mu = (1-\epsilon', \epsilon')$. I then claim that, for μ_0 sufficiently close to $(1,0)$, beliefs converge to $(1,0)$ with positive probability.

To prove the first point, suppose that the convergence result does not hold for the limit likelihood function. Then, since beliefs converge to $(1,0)$, there must be a t such that, for any $t' > t$, beliefs are updated using likelihood functions that, each by themselves, would lead to convergence to $(0,1)$.[1] I want to show that this is not possible. The following lemma provides this result.

Lemma A.1.3. *Consider a sequence of likelihood functions $\{g_t\}_{t=1}^{\infty}$ and suppose updating is performed at each period t using the likelihood function g_t. Let $p \in \Delta(S)$ be a limit frequency such that, for any t, beliefs μ_n converge to $(0,1)$ ($(1,0)$) for $n \to \infty$ when updating is performed using the likelihood function g_t in every period. Then, when we use, at each period t, the likelihood function g_t, beliefs converge to $(0,1)$ ($(1,0)$).*

Proof. Let $i(s_t) : \Omega \to \{1,...,S\}$ be the random variable whose value is the index of the observation at time t; that is, $i(s_t)$ says which element of S is observed in period t. Since the experiments are not identical nor independent, $i(s_t)$ depends on the history $(s_1,...,s_{t-1})$. Let $(x_{i,t}, y_{i,t})_{i=1}^{S} : S \to \mathbb{R}^2_{++}$ be the likelihood function in period t. The decision maker does not remember past observations; therefore, the likelihood function can only depend on μ_{t-1}. Studying beliefs convergence means to look at the limit of

$$\prod_{t=1}^{n} \left(\left(\frac{y_{i(s_t),t}}{x_{i(s_t),t}} \right) \Bigg| (s_1,...,s_{t-1}), \prod_{h=0}^{t-1} \left(\frac{y_{i(s_h),h}}{x_{i(s_h),h}} \right) \mu_0 \right),$$

[1] I am implicitly excluding the case in which the limit frequency belongs to the hyperplane corresponding to the limit likelihood function. This can be safely done since the theorem deals with limit frequencies in open sets.

where I define $y_{i(s_0),0}/x_{i(s_0),0} \equiv 1$. By the assumption in Theorem 1.4.1, the random variables $i(s_t)$ are independent in the limit, that is

$$\prod_{t=1}^{n}\left(\left(\frac{y_{i(s_t)}}{x_{i(s_t)}}\right)\Bigg|(s_1,...,s_{t-1})\right)\to\left(\prod_{j=1}^{S}\left(\frac{y_j}{x_j}\right)^{p_j}\right)^{n},\quad n\to\infty.$$

By hypothesis we know that beliefs converge for any fixed likelihood function, that is

$$\left(\prod_{j=1}^{S}\left(\frac{y_j}{x_j}\right)^{p_j}\right)^{n}\to 0\qquad \text{i.e.}\ \prod_{j=1}^{S}\left(\frac{y_j}{x_j}\right)^{p_j}<1.$$

Even in the case of non-constant likelihood functions, nothing changes for the random variables $i(s_t)$: they only depend on past observations, not on the decision maker's beliefs. Therefore

$$\prod_{t=1}^{n}\left(\left(\frac{y_{i(s_t),t}}{x_{i(s_t),t}}\right)\Bigg|(s_1,...,s_{t-1}),\prod_{h=0}^{t-1}\left(\frac{y_{i(s_h),h}}{x_{i(s_h),h}}\right)\mu_0\right)\to\prod_{t=1}^{n}\prod_{j=1}^{S}\left(\frac{y_{j,t}}{x_{j,t}}\Bigg|\prod_{h=0}^{t-1}\left(\frac{y_{i(s_h),h}}{x_{i(s_h),h}}\right)\mu_0\right)^{p_j}.$$

Notice that not only is the the second product on the right hand side smaller than 1 for any t, but also its limit for $t\to\infty$ is strictly smaller than 1. In fact, it corresponds to the limit likelihood function, for which we assume convergence of the decision maker's beliefs. So the right hand side goes to 0 as $n\to\infty$, that is, beliefs converge to the same limit implied by the limit likelihood function. □

To address the second point, let's first introduce a slightly different notation. I will denote with $\left(\frac{y_{j,(\mu)}}{x_{j,(\mu)}}\right)_{j=1}^{S}$ the ratios that summarize the likelihood function associated to beliefs μ. Since likelihood functions converge as $\mu\to(1,0)$, each ratio converges to $\left(\frac{y_{j,(1,0)}}{x_{j,(1,0)}}\right)_{j=1}^{S}$, so that,

$$\forall\epsilon>0,\exists\delta>0\ s.th.\ |\mu-(1,0)|<\delta\Longrightarrow\max_{j\in S}\left|\frac{y_{j,(\mu)}}{x_{j,(\mu)}}-\frac{y_{j,(1,0)}}{x_{j,(1,0)}}\right|<\epsilon.$$

This means that, for any $\alpha\in(0,1)$, we can pick μ_0 sufficiently close to 1 to make the probability of $|\mu_1-(1,0)|<\alpha$ arbitrarily close to the probability that would be obtained from the limit likelihood function. Moreover, we can always find μ_0 to make $|\mu_1-(1,0)|$ arbitrarily small. Therefore, for any finite n, we can choose μ_0 such that, with probability 1, every μ_t, for $t=1,...,n$, is associated with a likelihood function that would lead to convergence to $(1,0)$. Let's call n_{μ_0} the maximum n such that, when starting from beliefs μ_0, with probability 1 $v(.,\mu_t)$ would lead to convergence to $(1,0)$ for any $t\leq n$. Define beliefs $\mu^*=(1-\epsilon^*,\epsilon^*)$ such that, for any $\mu=(1-\epsilon,\epsilon)$ with $\epsilon<\epsilon^*$, the associated likelihood functions would lead to convergence to $(1,0)$ and there is no $\epsilon^{**}>\epsilon^*$ with the same property.

By Lemma A.1.3 we know that if, in any period, updating is performed using a likelihood function that would lead to convergence to $(1,0)$, then beliefs converge to

$(1,0)$. Therefore, for any μ_0, $\bar{\mu}$ and $p \in (0,1)$, there exists a smallest $N_{\mu_0,\bar{\mu},p}$ such that, with probability p, $\mu_t > \bar{\mu}$ for any $t > N_{\mu_0,\bar{\mu},p}$.[2]

Now fix $\bar{\mu} = \mu^*$. If we can prove that, for some $\mu_0 > \mu^*$ and some p, $n_{\mu_0} > N_{\mu_0,\mu^*,p}$, then the proof of the second point is completed. But n_{μ_0} can be taken arbitrarily large if we choose μ_0 sufficiently close to $(1,0)$, while $\lim_{\mu_0 \to (1,0)} N_{\mu_0,\mu^*,p}$ is clearly bounded.

From the two points above, it follows that, in order to have convergence to the true parameter whenever the limit frequency falls in $int(\mathcal{L}_1 \setminus \mathcal{L}_2)$ or $int(\mathcal{L}_2 \setminus \mathcal{L}_1)$, a necessary condition is the existence of a hyperplane separating the two sets. However, if $S > 3$, this may not be the case, as is clear from the example described by Figure A.1, where $S = 4$.

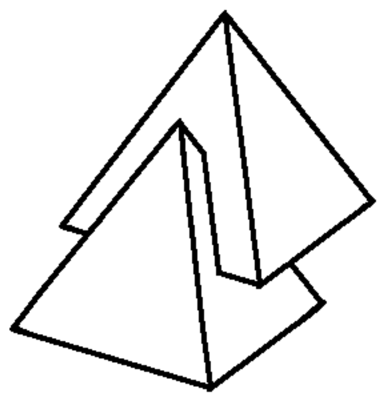

Figure A.1: Three-dimensional Sets

Whatever hyperplane we choose, there will be frequencies in $int(\mathcal{L}_1 \setminus \mathcal{L}_2)$ for which beliefs converge to $\delta_{\mathcal{L}_2}$ and/or frequencies in $int(\mathcal{L}_2 \setminus \mathcal{L}_1)$ for which beliefs converge to $\delta_{\mathcal{L}_1}$. This completes the proof.

A.2 Proof of Theorem 1.4.2

Lemma A.2.1. *Let $A, B \subset \mathbb{R}^2$ be two closed and convex sets such that $B = A + k = \{x + k \in \mathbb{R}^2 : x \in A\}$ for some $k \in \mathbb{R}^2$. There exist disjoint convex sets A' and B' with $A \setminus B \subset A'$ and $B \setminus A \subset B'$.*

Proof. Suppose there exist $x \in A \cap B$, $y_1, y_2 \in A \setminus B$, $z_1, z_2 \in B \setminus A$, $\beta, \gamma \in (0,1)$ such that $x = \beta y_1 + (1-\beta) y_2 = \gamma z_1 + (1-\gamma) z_2$. By construction, $y_1 + k, y_2 + k \in B$ and $z_1 - k, z_2 - k \in A$. Since A and B are convex, $co(\{y_1, y_2, z_1 - k, z_2 - k\}) \subset A$ and $co(\{z_1, z_2, y_1 + k, y_2 + k\}) \subset B$. Notice that these two sets are convex quadrangles, they have nonempty intersection (they both contain x) and one is the translation of the other. It is easy to verify that these conditions imply that, for some $i \in (\{1,2\})$, either $y_i \in B$ or $z_i \in A$. But this is a contradiction. Therefore no $x \in A \cap B$ can be a

[2] Since the set S is finite, $\mu_t(.) \propto v(., \mu_{t-1})\mu_{t-1}$ converges in probability if and only if it converges almost surely.

convex combination of points in both $A \setminus B$ and $B \setminus A$. Define P_A (P_B) as the set of $x \in A \cap B$ that are convex combinations of points in $A \setminus B$ ($B \setminus A$) and consider the sets $A' = (A \setminus B) \cup P_A$ and $B' = (B \setminus A) \cup P_B$. By the previous argument, $A' \cap B' = \emptyset$; moreover, they are convex, since clearly $A' = co(A \setminus B)$ and $B' = co(B \setminus A)$. □

By Lemma A.2.1, there always exists a straight line strictly separating $int(\mathcal{L}_1 \setminus \mathcal{L}_2)$ and $int(\mathcal{L}_2 \setminus \mathcal{L}_1)$. Then, the existence of a likelihood function for which the hyperplane separates the limit frequencies leading to convergence to $\delta_{\mathcal{L}_1}$ and those leading to convergence to $\delta_{\mathcal{L}_2}$ immediately follows form the next lemma.

Lemma A.2.2. *Consider the simplex $\Delta(S)$, with $S = 3$ There exists a straight line $k_0 + \sum_{i=1}^{2} k_i \alpha_i = 0$ strictly separating $int(\mathcal{L}_1 \setminus \mathcal{L}_2)$ and $int(\mathcal{L}_2 \setminus \mathcal{L}_1)$ such that there are $x = (x_1, x_2, 1 - x(1) - x(2))$ and $y = (y_1, y_2, 1 - y(1) - y(2))$ belonging to $\Delta(S)$ that satisfy the result in Lemma A.1, with $k_0 + \sum_{i=1}^{2} k_i x_i > 0$ and $k_0 + \sum_{i=1}^{2} k_i y_i < 0$.*[3]

Proof. The particular shape of the sets under consideration (see Figure 1.3) implies that we can focus on straight line parallel to one of the sides of the simplex: if $int(\mathcal{L}_1 \setminus \mathcal{L}_2)$ and $int(\mathcal{L}_2 \setminus \mathcal{L}_1)$ are strictly separated by a straight line, then they are also strictly separated by a straight line parallel to one of the sides. Moreover, we can always relabel the elements of S such that the line is horizontal. So we can adopt the normalization $k_0 = 1$, $k_1 = 0$ and $k_2 \in (-\infty, -1)$. We are therefore considering straight lines as in Figure A.2.

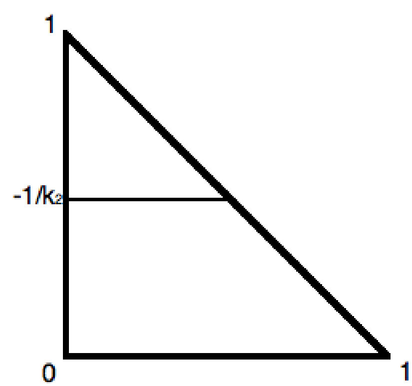

Figure A.2: Horizontal Separating Line

We already know that the implication in Lemma A.1 holds if we choose x and y such that $\frac{k_j}{k0} = \frac{\ln x_j - \ln(1-\sum x_i) - \ln y_j + \ln(1-\sum y_i)}{\ln(1-\sum x_i) - \ln(1-\sum y_i)}$, where the denominator is nonzero by construction. Notice that, for some K, that can be assumed positive without loss of generality,

$$k_0 + \sum_{i=1}^{2} k_i x_i =$$

[3]I think it would be possible to prove that such a likelihood function exists for any straight line strictly separating $int(\mathcal{L}_1 \setminus \mathcal{L}_2)$ and $int(\mathcal{L}_2 \setminus \mathcal{L}_1)$. I also believe that the result still holds for a generic finite set S and for generic convex subsets of $\Delta(S)$. These generalizations, however, are not necessary for the proof of the theorem.

$$= K[\ln(1-\sum x_i)-\ln(1-\sum y_i)]+\sum K[\ln x_i-\ln(1-\sum x_j)-\ln y_i+\ln(1-\sum y_j)]x_i =$$

$$= K\{[\ln(1-\sum x_i)-\ln(1-\sum y_i)](1-\sum x_i)+\sum(\ln x_i-\ln y_i)x_i\} > 0.$$

where the inequality is immediate, provided $x \neq y$. Similarly, $k_0 + \sum_{i=1}^{S-1} k_i y_i < 0$. We still need to prove that we can find x and y that satisfy the equalities above and belong to $\Delta(S)$. Simple algebra provides the functional relation between x and y. In fact,

$$\frac{y_j}{(1-\sum y_i)^{\left(1+\frac{k_j}{k_0}\right)}} = \frac{x_j}{(1-\sum x_i)^{\left(1+\frac{k_j}{k_0}\right)}}, \quad j = 1, 2, \tag{A.1}$$

where by construction $\sum y_i \neq \sum x_i$.

Consider $x \in int(\Delta(S))$ and assume we are able to find y satisfying equations (A.1) and such that $y_1 > 0$. Then, taking the equation for $i = 1$, it must be the case that $1 - y_1 - y_2 > 0$; this in turn implies, from the equation with $i = 2$, that $y_2 > 0$, that is, $y \in int(\Delta(S))$. Also notice that $x \neq y$ implies $x_1 \neq y_1$. So we simply have to show that for any straight line, that is for any $k_2 \in (-\infty, -1)$, there exists $x \in int(\Delta(S))$ such that there is y satisfying the equations above, with $y_1 > 0$ and $y_1 \neq x_1$.

Let's take $x_2 = -\frac{1}{k_2} + \epsilon$, for $\epsilon > 0$. From the system of equations above we get

$$(1 + k_2(1-\epsilon))y_1^{-k_2} - k_2 x_1 y_1^{-k_2-1} - (1 - k_2\epsilon)x_1^{-k_2} = 0.$$

Since we want $x_2 < 1$, clearly $1 + k_2(1-\epsilon) < 0$, while $1 - k_2\epsilon > 0$. To simplify notation, denote the first term as a and the second as b. It is also immediate to see that any solution must have the form $y_1 = hx_1$, for some $h > 0$. Therefore all we need is to find positive solutions to the equation

$$ah^{-k_2} - k_2 h^{-k_2-1} - b = 0.$$

First of all, notice that $h = 1$ is always a solution, and it is the only solution if $\epsilon = 0$. For $\epsilon > 0$, we want to prove that there exists another solution. The task is nontrivial since k_2 can be any real number in $(-\infty, -1)$. I will prove the result for k_2 rational. Given the smooth nature of the equation, the result generalizes to the entire interval by a continuity argument. Denote $-k_2 = \frac{\alpha}{\beta}$, where $\alpha, \beta \in \mathbb{N}_{++}$ and $\alpha > \beta$. With an obvious change of variables, we obtain the equation

$$az^{\alpha} + \frac{\alpha}{\beta}z^{\alpha-\beta} - b = 0.$$

I am going to use Sturm's theorem (see, for example, [16]) to show that such an equation always has two and only two positive solutions. First we have to construct a particular sequence of polynomials called Sturm chain. The first element is the polynomial $az^{\alpha} + \frac{\alpha}{\beta}z^{\alpha-\beta} - b$ itself; the second is its derivative. Every other element is computed as the reminder of the polynomial division between the two polynomials immediately above, with its sign changed. Notice that the degree of the polynomials in the chain is strictly decreasing. The chain ends when the last polynomial has degree 1. To find the number of distinct positive real roots of our equation, we have to determine the signs of each polynomial in the chain corresponding to $z = 0$ and $z \to +\infty$ and to see how many times signs flip in these two cases. Sturm's theorem states that the number of solutions is equal to the difference between the numbers of sign flips for $z = 0$ and $z \to +\infty$.

In order to find the number of solutions, it is not really necessary to compute the Sturm chain. All we need are the corresponding sequences of the exponents appearing in the polynomials and the signs of the coefficients. I am now going to show how these sequences can be computed. For the moment, I will build generic sequences, disregarding the specific values of α and β for which some exponent turns out to be 0. I will show later how to handle these cases. So in what follows, any exponent that is not explicitly 0 is assumed to be positive. On the other hand, I will indicate as special cases the values of α and β that change the functional form of the exponents with respect to the general case. How the chain evolves after these special cases will be examined later.

The first three elements of the sequences are easy to compute.

Element	General case		Special cases	
	Exponents	Coefficients' signs	Exponents	Coefficients' signs
1	α, $\alpha-\beta$, 1	- + -		
2	$\alpha-1$, $\alpha-\beta-1$	- +		
3	$\alpha-\beta$, 0	- +	$\beta = 1$: $\alpha-2$, 0	$\beta = 1$: - +

From this point on, signs and exponents can be computed following simple rules, whose proof is straightforward and omitted. If the element number is even, the sign of the first term is the opposite to that of the second term of the one to the last element; the sign of the second term is the product of the signs of the terms of the last element times the sign of the first term of the one to the last element. For elements whose number is odd, the rules for determining the signs of the two terms are switched. The rule for the exponents follows a cyclical scheme, shown in the table below.

Element	General case	Special cases
2	g_1, g_2	
3	g_2+1, 0	$g_1 = g_2 + 1$: g_2, 0
4	g_2, g_3 (depending on the relation between g_1 and g_2)	$g_2 = g_3$: g_2
5	g_3+1, 0	$g_2 = g_3 + 1$: g_3, 0
6	g_3, g_4 (depending on the relation between g_2 and g_3)	$g_3 = g_4$: g_3

The cyclicality of the exponents is apparent; the specific form of the second exponent in the even elements is not important for the result. Knowing these rules, we can easily find the actual sequences.[4]

<table>
<tr><th rowspan="2">El.</th><th colspan="2">General case</th><th colspan="2">Special cases</th></tr>
<tr><th>Exponents</th><th>Signs</th><th>Exponents</th><th>Signs</th></tr>
<tr><td rowspan="2">4</td><td rowspan="2">$\alpha-\beta-1$,
$(i-1)\beta-(i-2)\alpha-1$
for $\alpha\in\left[\frac{i}{i-1}\beta, \frac{i-1}{i-2}\beta\right)$</td><td rowspan="2">- +</td><td colspan="2">$\alpha=\frac{i}{i-1}\beta$</td></tr>
<tr><td>$\alpha-\beta-1$</td><td>-</td></tr>
<tr><td rowspan="2">5</td><td rowspan="2">$(i-1)\beta-(i-2)\alpha$, 0</td><td rowspan="2">+ -</td><td colspan="2">$\alpha=\frac{i}{i-1}\beta+\frac{1}{i-1}$</td></tr>
<tr><td>$(i-1)\beta-(i-2)\alpha-1$, 0</td><td>+ -</td></tr>
<tr><td rowspan="2">6</td><td rowspan="2">$(i-1)\beta-(i-2)\alpha-1$,
$[(i-2)j+(i-1)]\alpha-[(i-1)j+i]\beta-1$
for $\alpha\in\left(\frac{(j+1)i-j}{(j+1)i-(1+2j)}\beta, \frac{(j+2)i-(j+1)}{(j+2)i-(3+2j)}\beta\right]$</td><td rowspan="2">- +</td><td colspan="2">$\alpha=\frac{(j+2)i-(j+1)}{(j+2)i-(3+2j)}\beta$</td></tr>
<tr><td>$(i-1)\beta-(i-2)\alpha-1$</td><td>+</td></tr>
<tr><td rowspan="2">7</td><td rowspan="2">$[(i-2)j+(i-1)]\alpha-[(i-1)j+i]\beta$, 0</td><td rowspan="2">+ -</td><td colspan="2">$\alpha=\frac{(j+2)i-(j+1)}{(j+2)i-(3+2j)}\beta+\frac{1}{(j+2)i-(3+2j)}$</td></tr>
<tr><td>$[(i-2)j+(i-1)]\alpha+$
$-[(i-1)j+i]\beta-1$, 0</td><td>+ -</td></tr>
</table>

Notice the cyclicality of the sequence of the signs: the signs at elements 6 and 7 are the same as at elements 2 and 3. The same cyclicality (with cycles of length 4 starting from element 4) is therefore inherited by the sequence of the signs of the polynomials for $z=0$ and $z\rightarrow=\infty$ and, in turn, by the sequence of the difference between the numbers of sign flips.

Element	1	2	3	4	5	6	7
0	-	0	+	0	-	0	+
$+\infty$	-	-	-	-	+	-	-
Difference	0	0	1	1	1	0	1

We now have to consider all the cases in which the sequence ends. It is easy to see that this can happen only when:

- we are in one of the special cases considered above; we will see that in these cases the sequence ends after at most two additional steps;
- the second term of an even element has exponent 0; the sequence ends after two additional steps;
- an even element has degree 0.

[4]The signs for the special cases of elements 4 and 6 are more difficult to compute. However, we will see that there is a shortcut to get the results.

For each of these cases, we need to check whether the difference between the numbers of sign flips turns out to be 2. Let's look at the table with the rule for computing the exponents and let's start with the cases of the type $g_2 = g_3$ ($g_2 > 0$) in an even element of the sequence. In this case, in the next step, the power of the polynomial is 0 and the sequence ends. Obviously, the signs for $z = 0$ and $z \to +\infty$ must be the same. So we have a situation like

Element	Even	Odd
0	0	s_2
$+\infty$	s_1	s_2

To determine the values of s_1 and s_2, notice that at any odd element the sign for $z = 0$ is the opposite than that for $z \to +\infty$, and the difference between signs flips is 1. Moreover, we know that at least one positive solution exists, that is $z = 1$. The only way the final difference between sign flips can be at least one is that s_1 be the same as the corresponding sign for the element immediately above and $s_2 = s_1$. But in this case the number of sign flips turns out to be 2.

Let's consider now the cases of the sort $g_2 = g_3 + 1$, which are special for odd elements. It is easy verifiable that we can only have the following sub-cases:

- if $g_2 > 2$, the next element has coefficients $(1, 0)$ and the same signs as the element immediately above, so that the difference between sign flips is still 1; the sequence ends after another step, where, again by the fact that a solution exists, the sign must be the opposite of the first sign of the last step; this implies the existence of two solutions;

- if $g_2 = 2$, the sequence ends at the next step; the existence of two solutions follows by the same argument.

When the second term of an even element has exponent 0, the following holds:

- the sign for $z = 0$ becomes +; notice that this implies that the difference between sign flips is 1;

- the next element has exponents $(1, 0)$ and the signs are still computed according to the general rule; this does not change the difference;

- the usual argument determines the sign at the final step, proving the existence of two solutions.

It remains to analyze the cases in which an even element has degree 0. Again, the usual argument gives the sign and implies the existence of two solutions.

□

This completes the proof of Theorem 1.4.2 for the case $S = 3$. It can be easily adapted to the case $S = 2$.

Remark 2. It is easy to verify that the solution $h \neq 1$ is such that $h \to 1$ for $\epsilon \to 0$. This implies that, if $k_0 + \sum_{i=1}^{2} k_i p_i > 0$ for any $p \in int(\mathcal{L}_1 \setminus \mathcal{L}_2)$ and $k_0 + \sum_{i=1}^{2} k_i q_i < 0$ for any $q \in int(\mathcal{L}_2 \setminus \mathcal{L}_1)$, and if $int(\mathcal{L}_1) \cap int(\mathcal{L}_2) \neq \emptyset$, then x and y can be chosen such that $x \in \mathcal{L}_1$ and $y \in \mathcal{L}_2$. The same does not follow when the sets are disjoint.

Remark 3. It is highly plausible that Theorem 1.4.2 holds for general convex subsets of the simplex. To prove it, however, we need to consider all possible straight lines, not just the horizontal ones; that is, we cannot assume $k_1 = 0$ and $k_2 \in (-\infty, -1)$. Sturm's theorem can still be applied, but the proof will certainly be more cumbersome.

Moreover, any attempt to prove that the likelihood function can always be chosen such that $x \in \mathcal{L}_1$ and $y \in \mathcal{L}_2$ clearly requires the use of generic straight lines. The idea is that, for any two subsets, there should exists a straight line such that one of the corresponding likelihood function satisfies the requirement.

A.3 Proof of Lemma 1.4.1

To see why Lemma 1.4.1 must be true, imagine two overlapping sets $\mathcal{L}_1$ and $\mathcal{L}_2$ such that $int(\mathcal{L}_1 \backslash \mathcal{L}_2)$ and $int(\mathcal{L}_2 \backslash \mathcal{L}_1)$ can be strictly separated only by a horizontal straight line. Fix $x \in \mathcal{L}_1$. Obviously, there must exist $\alpha \in (1, \infty)$ such that $x_2 = 1 - \alpha x_1$. If we take $y \in \Delta(S)$ satisfying equations (A.1), it is easy to see that $y_2 = 1 - \alpha y_1$.[5] This means that any couple of points (x, y) satisfying the equations must lay on the same straight line going through the point $(0, 1)$. This is enough to prove the lemma. In fact, as can be seen in Figure A.3, we can easily construct $\mathcal{L}_1$ and $\mathcal{L}_2$ for which such x and y cannot be in the same "position" within $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively.

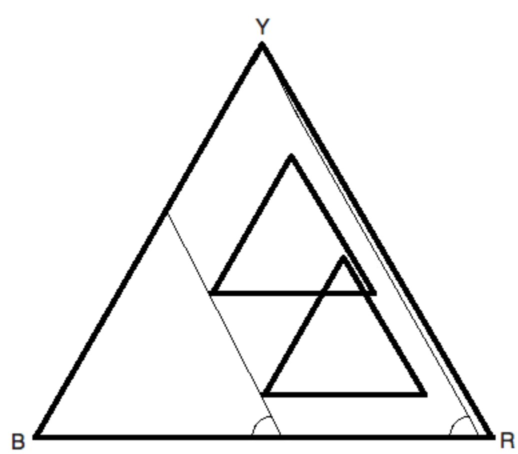

Figure A.3: A Counterexample

[5]Just take the equation for $j = 1$ in the special case of a horizontal line.

Appendix B

Proofs for Chapter 2

B.1 Proof of Proposition 2.3.1

Let's prove that the representation implies the axiom. First, $\forall E \in \Sigma$, $E \subseteq v(E)$, as required by Axiom 4. Now, consider $s \notin E$ that is also $\succsim^E$-null, and suppose that there exists $s' \in E$ such that s and s' are indistinguishable. I will denote indistinguishability with the notation $\succsim^{\{s\}} = \succsim^{\{s'\}}$. Since $E \subseteq V^E$, we have $s' \in V^E$. Similarly, $s \in V^{\{s\}}$ and $s' \in V^{\{s'\}}$. Since $\succsim^{\{s\}} = \succsim^{\{s'\}}$, $V^{\{s\}} = V^{\{s'\}}$, so that $s \in V^{\{s'\}}$. Since $\mathcal{V}$ is closed under intersections, $(V^{\{s'\}} \cap V^E) \in \mathcal{V}$. We know that $s' \in (V^{\{s'\}} \cap V^E)$, but $(V^{\{s'\}} \cap V^E) \subset V^{s'}$, because $s \notin V^E$. This contradicts point (iii).

We need now to show that, if there exists $s \notin E$ that is also $\succsim^E$-non-null, there must exist an indistinguishable $s' \in E$. Before doing this, let's prove a preliminary result. Given two states s_1 and s_2, either $V^{\{s_1\}} = V^{\{s_2\}}$ or the two are disjoint. Suppose $V^{\{s_1\}} \cap V^{\{s_2\}} \neq \emptyset$ and consider the following two cases:

- $s_1 \notin V^{\{s_2\}}$. Since $\mathcal{V}$ is an algebra, $V^{\{s_1\}} \setminus (V^{\{s_1\}} \cap V^{\{s_2\}}) \in \mathcal{V}$. Moreover, $s_1 \in V^{\{s_1\}} \setminus (V^{\{s_1\}} \cap V^{\{s_2\}}) \subset V^{\{s_1\}}$. This contradicts point (iii). The argument is the same is we invert s_1 and s_2.

- $s_1 \in V^{\{s_2\}}$ and $s_2 \in V^{\{s_1\}}$. Since $s_1 \in (V^{\{s_1\}} \cap V^{\{s_2\}})$, point (iii) implies that $V^{\{s_1\}} \subseteq V^{\{s_2\}}$. Similarly, we must have $V^{\{s_2\}} \subseteq V^{\{s_1\}}$ and so $V^{\{s_1\}} = V^{\{s_2\}}$.

Now, let $s \notin E$ be $\succsim^E$-non-null and suppose that, for any $s' \in E$, $\succsim^{\{s\}} = \succsim^{\{s'\}}$. The representation requires that $V^{\{s\}} \neq V^{\{s'\}}$. We have already shown that this implies that $V^{\{s\}} \cap V^{\{s'\}} = \emptyset$. So, for any $s' \in E$, s' is $\succsim^{\{s\}}$-null. Therefore E is $\succsim^{\{s\}}$-null; that is, $E \cap V^{\{s\}} = \emptyset$. Clearly, $S \in \mathcal{V}$ and, since $\mathcal{V}$ is an algebra, $S \setminus V^{\{s\}} \in \mathcal{V}$. Being $E \subseteq S \setminus V^{\{s\}}$, point (iii) implies that $V^E \subseteq S \setminus V^{\{s\}}$. Since $s \in V^{\{s\}}$, we have that s is $\succsim^E$-null, which contradicts the hypothesis.

It remains to be proved that the axiom implies the representation. Points (i) and (ii) are the same as in Proposition 2.2.1. So I am going to show that Axiom 4 implies both Axioms 2 and 3. Axiom 3 obviously holds: given a signal E any $s \in E$ is $\succsim^E$-non-null. Suppose Axiom 2 is not satisfied. Then, for some signal E, $v(v(E)) \neq v(E)$. By Axiom 4, $v(E) \subseteq v(v(E))$. Given $s \in v(v(E)) \setminus v(E)$, there exists $s' \in v(E)$ such that $\succsim^{\{s\}} = \succsim^{\{s'\}}$. But there also exists $s'' \in E$ such that $\succsim^{\{s'\}} = \succsim^{\{s''\}}$ (this may even be s' itself). Therefore s is indistinguishable from some state in E, and Axiom 4 implies that $s \in v(E)$, which contradicts the hypothesis.

For point (iii), we already know that $E \subseteq v(E)$. Suppose now that there exists a signal F which is a fixed point of v and such that $E \subseteq F \subset v(E)$. Since $v(F) = F$, by Axiom 4 there is no $s \in v(E) \setminus F$ such that s is indistinguishable from some state in F. Given that $E \subseteq F$, again by Axiom 4 s cannot be contained in $v(E)$, contradicting the hypothesis.

The last thing to prove is that the fixed points of v constitute an algebra. Let E and F be two fixed points of v. Then there is no $s \notin E$ that is indistinguishable from a state in E, and similarly there is no $s' \notin F$ that is indistinguishable from some state in F. Therefore, no state in $(E \cap F)^c$ can be indistinguishable from some state in $E \cap F$. So $v(E \cap F) = E \cap F$ and the set of fixed points of v is closed under intersections. Closure under complements is obvious.

B.2 Proof of Proposition 2.4.1

Let's first prove that the representation implies the axioms. Axioms 2 and 3 are obviously implied. To prove Axiom 6, let $s \in V^A \setminus A$ and suppose that there is no $s' \in A$ such that $\succsim^{\{s'\}} = \succsim^{\{s\}}$, that is, such that $V^{\{s'\}} = V^{\{s\}}$. Properties (iv)1 and (iv)3 and the closure of $\mathcal{V}$ under intersection imply that $V^{\{s\}}$ is the smallest element of $\mathcal{V}$ that contains s. In fact, by (iv)1, $V^{\{s\}}$ must contain s; it cannot be a superset of some other element of $\mathcal{V}$ containing s, because of (iv)3. And there cannot be multiple sets satisfying these properties, because their intersection would still belong to $\mathcal{V}$, leading to a contradiction. It is therefore obvious that, if $V^{\{s\}} = V^{\{s'\}}$, then both s and s' are in $V^{\{s\}}$. But I need the inverse result: I want to show that, if $V^{\{s\}} \neq V^{\{s'\}}$, then $s' \notin V^{\{s\}}$. Suppose that this is not true, that is, $s' \in V^{\{s\}}$ and consider the following two cases:

- we cannot have $V^{\{s\}} \subset V^{\{s'\}}$, because $V^{\{s'\}}$ is the smallest element of $\mathcal{V}$ containing s';
- if $V^{\{s\}} \setminus V^{\{s'\}} \neq \emptyset$, (iii) implies that $V^{\{s\}} \setminus V^{\{s'\}} \in \mathcal{V}$. Since $V^{\{s\}} \setminus V^{\{s'\}} \subset V^{\{s\}}$, it cannot include the state s. Therefore $s \in V^{\{s'\}}$. By (iii), $V^{\{s\}} \cap V^{\{s'\}} \in \mathcal{V}$. Clearly, $s \in V^{\{s\}} \cap V^{\{s'\}}$ and $V^{\{s\}} \cap V^{\{s'\}} \subset V^{\{s\}}$. But such a set cannot exist.

So we have proved that, for any $s' \in A$, $s' \notin V^{\{s\}}$. But $V^A \setminus V^{\{s\}} \in \mathcal{V}$, violating (iv)3.

Axiom 7 is easily proved. Consider two indistinguishable states s and s' such that $s \in V^A$, but $s' \notin V^A$. it has already been proved that $s, s' \in V^{\{s\}}$. So $V^{\{s\}} \cap V^A \subset V^{\{s\}}$. By (iii), $V^{\{s\}} \cap V^A \in \mathcal{V}$. This contradicts the fact that $V^{\{s\}}$ is the smallest element of $\mathcal{V}$ containing s, as shown above.

Now I want to show that the axioms imply the representation. Properties (i) and (ii) immediately follow from Proposition 2.2.1. Let's now prove that any signal A is $\succsim^A$-non-null (property (iv)1). Suppose not, so that $A \cap V = \emptyset$. By Axiom 6, for any state in V there exists an indistinguishable state in A. But Axiom 7 requires that the last state is conditionally non-null, which contradicts the initial hypothesis.

To prove that the set of fixed points of v is a semialgebra, we need to show that

- $\forall A, B \in \mathcal{V},\ A \cap B \in \mathcal{V}$;
- $\forall A \in \mathcal{V}$, there exists a partition of $S \setminus A$ whose elements belong to $\mathcal{V}$.

Let's first prove closure under intersections. Consider $A, B \in \mathcal{V}$ such that $A \cap B \neq \emptyset$. Since by assumption $v(A) = A$, Axiom 7 implies that no state in A^c is indistinguishable from some state in A. The same is true for B. Therefore, by Axiom 6, $v(A \cap B) \subseteq A \cap B$. Suppose now that there exists $s \in A \cap B$ such that $s \notin v(A \cap B)$. By Axiom 5, s must be $\succsim^A$-null, that is $s \notin v(A)$, but this is impossible. So $A \cap B \in \mathcal{V}$.

Before moving to the second part, it is useful to prove the other property of $\mathcal{V}$, that is, closure with respect to set difference. Take $A, B \in \mathcal{V}$ and suppose there exists s such that $s \in v(A \setminus B)$ and $s \notin A \setminus B$. By Axiom 6, there exists $s' \in A \setminus B$ with $\succsim^{\{s'\}} = \succsim^{\{s\}}$. Since by assumption $v(A) = A$, Axiom 7 implies that $s \in A$. By assumption, $s \notin A \cap B$. On the other hand, since $A \cap B \in \mathcal{V}$, using Axioms 6 and 7 we can immediately show that no state in $A \setminus B$ is indistinguishable from some state in $A \cap B$, and therefore $v(A \setminus B) \cap (A \cap B) = \emptyset$. We have reached a contradiction, so we can conclude that $v(A \setminus B) \subseteq A \setminus B$.

To prove the other direction, let's first show that, for any $A, B \in \Sigma$, $v(A \cup B) \subseteq v(A) \cup v(B)$.[1] Take $s \notin v(A) \cup v(B)$ and consider two exhaustive cases.

1. $s \in A$ or $s \in B$. Then by Axiom 5 $s \notin v(A \cup B)$.
2. $s \notin A \cup B$. We can consider two sub-cases. If there is no $z \in A \cup B$ such that $\succsim^{\{s\}} = \succsim^{\{z\}}$, then Axiom 6 implies that $s \notin v(A \cup B)$. On the other hand, if such a z exists, by Axiom 7 it must be the case that $z \notin A \cup B$. Notice that Axiom 7 also implies that $s \in v(A \cup B)$ if and only if $z \in v(A \cup B)$. But z falls in case 1 above.

Take now $A, B \in \mathcal{V}$. By the previous result, we have $v(A) \subseteq v(A \setminus B) \cup v(A \cap B)$. By assumption, $v(A) = A$; moreover, $v(A \cap B) = A \cap B$, because $\mathcal{V}$ is closed under intersection. So, $A \setminus B \subseteq v(A \setminus B)$.

To complete the proof of property (iii), let $A \in \mathcal{V}$. We need to show that, for any $s \in A^c$, there exists an element of $\mathcal{V}$ that contains s and is disjoint from A. Closure under intersection then provides the result. It is easy to see that the following two properties are sufficient for the proof:

- $s \in V^{\{s\}}$;
- $\forall A, B \in \mathcal{V},\ A \setminus B \in \mathcal{V}$.

But the first result is a special case of (iv)1, which we have already proved, while the second is exactly what has been shown above.

Suppose now that property (iv)2 is not satisfied. So there exists $V \in \mathcal{V}$ such that $A \cap V^A \subset A \cap V$. Take $s \in A \cap (V \setminus V^A)$ and consider the following two cases:

- suppose $V^A \subset V$; then Axiom 5(i) requires that $v(V^A \cup \{s\}) = V^A$; but $V^A \cup \{s\} \subseteq V$, so Axiom 5(ii) implies that $s \notin v(V) = V$, which contradicts the assumption;

[1] For an interpretation of this property see Section 2.5.

- if there exists $s' \in V^A \setminus V$, then by Axiom 6 there is a state $s'' \in A \cap V^A$ such that s' and s'' are indistinguishable; since $s'' \in V$ and $v(V) = V$, Axiom 7 implies $s' \in V$, which is false.

Proving property (iv)3 is easy. Take $s \in V^A \setminus A$. By Axioms 6 and 7, there exists a state $s' \in A \cap V^A$ indistinguishable from s. But $A \cap V^A = A \cap V \subseteq V = v(V)$, so that Axiom 7 implies $s \in V$.

Remark 4. Unlike in Proposition 2.3.1, Axiom 2 cannot be deduced from the others.On one hand, it is possible to exclude the existence of some $E \in \Sigma$ such that $\exists s \in v(v(E)) \setminus v(E)$. In fact, Axiom 6 would imply that there exists $s' \in v(E)$ such that $\succsim^{\{s'\}} = \succsim^{\{s\}}$. The state s' is $\succsim^E$-non-null. Therefore, by Axiom 7, s must be $\succsim^E$-non-null, too, leading to a contradiction. On the other hand, the axioms are not sufficient to exclude the existence of some $E \in \Sigma$ such that $\exists s \in v(E) \setminus v(v(E))$.

B.3 Proof of Corollary 2.4.1

We need to show that, for any $V \in \mathcal{V}$ and for any $s \in V$, there exists $A \in \Sigma$ such that $s \in A$, $A \subseteq V$, $v(A) = V$, and, for any B with $v(B) = V$, $B \not\subset A$. The proof can be divided in three parts:

1. for any A with $v(A) = V$, $A \cap V \neq \emptyset$;
2. let A be a signal such that $s \in A$ and $v(A) = V$; then there exists $A' \in A \cap V$ such that $v(A') = V$;
3. take any $s \in V$ and let A be any subset of V such that $s \in A$ and $v(A) = V$; suppose there is a $B \subset A$ such that $s \notin B$ and $v(B) = V$; then there exists $A' \subset A$ such that $s \in A'$, $v(A') = V$, and $B \not\subset A'$.

It is straightforward to see how these three results establish the result. Part 1 is property (iv)1 in Proposition 2.4.1. Before moving to part 2, we need a preliminary result. I will prove that, for any signal A, $v(A \cap V^A) = V^A$. If $V^A \subset A$, then there is nothing to prove: $v(V^A) = V^A$ by Axiom 2. Suppose now that $V^A \setminus A \neq \emptyset$. For any $s \in V^A \setminus A$, Axiom 6 requires the existence of $s' \in A$ such that $\succsim^{\{s'\}} = \succsim^{\{s\}}$. On the other hand, since $v(V^A) = V^A$, Axiom 7 implies that no state in $(V^A)^c$ is indistinguishable from some state in V^A. Therefore, the s' above must belong to $A \cap V^A$. Consider now the signal $A \cap V^A$. I claim that $A \cap V^A \subseteq v(A \cap V^A)$. Suppose not; that is, suppose there exists $z \in A \cap V^A$ such that $z \notin v(A \cap V^A)$. By Axiom 5(ii), z must also be $\succsim^A$-null. But this is false. So we have proved that any state in $V^A \setminus A$ is indistinguishable from some state in $A \cap V^A$ and that any state in $A \cap V^A$ is $\succsim^{A \cap V^A}$-non-null. Axiom 7 then implies that $V^A \subseteq v(A \cap V^A)$. By Axiom 6, any state in $v(A \cap V^A) \setminus (A \cap V^A)$ must be indistinguishable from some state in $A \cap V^A$. But we have already proved that no state in $(V^A)^c$ can satisfy this requirement. Therefore, $v(A \cap V^A) = V^A$. So part 2 is satisfied taking $A' = A \cap V$.

Part 3 is easily proved. Given that $B \subset v(B)$, Axioms 6 and 7 imply that $v(B)$ (that is to say V) is the set of all the states that are indistinguishable from some state in B. In particular, since $s \in v(B)$, there must be a state $s' \in B$ such that $\succsim^{\{s'\}} = \succsim^{\{s\}}$. Now take the signal $A' = (B \setminus \{s'\}) \cup \{s\}$. The set of indistinguishable states must obviously be the same as for B, that is, the set V. If we can prove that $A' \subset v(A')$, then Axioms 6 and 7 imply that $v(A') = V$. But this is immediately implied by Axiom 5: any state in $A' \setminus v(A)$ should also be $\succsim^V$-null, and this is clearly impossible.

B.4 Proof of Proposition 2.5.1

Information monotonicity an be restated as an assumption on the monotonicity of the function v: if $A \subset B$, then $v(A) \subseteq v(B)$. It is straightforward to see that Information monotonicity implies (iii). If not, then we could have $A \in \Sigma$ and $V \in \mathcal{V}$ such that $A \subseteq V \subset v(A)$; since $v(V) = V$, we have a contradiction. The other direction is equally trivial. Without Information monotonicity, there exist $A, B \in \Sigma$ such that $A \subset B$ and $v(B) \subset v(A)$. By global underconfidence we then have $A \subset v(B) \subset v(A)$. Since $v(B) \in \mathcal{V}$, we get a contradiction.

Consider now $V_1, V_2 \in \mathcal{V}$. By Information monotonicity $v(V_1 \cap V_2) \subseteq v(V_1) = V_1$ and $v(V_1 \cap V_2) \subseteq v(V_2) = V_2$, that is $v(V_1 \cap V_2) \subseteq (V_1 \cap V_2)$. Underconfidence then implies $v(V_1 \cap V_2) = (V_1 \cap V_2)$, meaning that $(V_1 \cap V_2) \in \mathcal{V}$.

B.5 Proof of Proposition 2.5.2

The first part is obvious. To prove (i), consider $V_1, V_2 \in \mathcal{V}$. Axiom 9 requires that $v(V_1 \cup V_2) \subseteq (V_1 \cup V_2)$. Underconfidence then implies $v(V_1 \cup V_2) = (V_1 \cup V_2)$, meaning that $(V_1 \cup V_2) \in \mathcal{V}$.

For (ii) we just need to prove that, for all signals A and B, $v(\{A, B\}) \subseteq (v(A) \cup v(B))$. By Proposition 2.5.1, we know that $\mathcal{V} \cup \emptyset$ is an algebra. By Axiom 8, for any A and B, $v(A)$ and $v(B)$ are the smallest elements of $\mathcal{V}$ such that $A \subseteq v(A)$ and $B \subseteq v(B)$. Therefore, $(A \cup B) \subseteq (v(A) \cup v(B))$ and, since $(v(A) \cup v(B)) \in \mathcal{V}$, it follows from from Axiom 8 that $v(\{A, B\}) \subseteq (v(A) \cup v(B))$.

Appendix C

Matlab Code and Numerical Examples for Chapter 3

C.1 Matlab Code

This section describes the Matlab code I used to study the model in the case of three states of the world. I also assumed $u(x) = \sqrt{x}$. First, I want to compute the optimal function v for specific values of the decision maker's initial beliefs, given a constraint on the number of observable signals. In the code, I use k to devote $\bar{V}$ and the following intuitive notation for the initial beliefs: $mu\_1 = \mu(s_1)$, $mu\_2 = \mu(s_1)$, $p\_1 = p(\{s_1\})$, $p\_2 = p(\{s_2\})$, $p\_3 = p(\{s_3\})$, and $p\_S = p(S)$. The prior on s_3 and the other likelihoods $p(\{s_1, s_2\})$, $p(\{s_1, s_3\})$, and $p(\{s_2, s_3\})$ are obtained from y as

```
mu_3=1-mu_1-mu_2;
p_12=(1-p_1-p_2+p_3-p_S)/2;
p_13=(1-p_1+p_2-p_3-p_S)/2;
p_23=(1+p_1-p_2-p_3-p_S)/2;
```

The first thing to do is to compute the optimal act, given the observed signal. To do that, I assign a number to each signal: $\{s_1\} = 1$, $\{s_2\} = 2$, $\{s_3\} = 3$, $\{s_1, s_2\} = 4$, $\{s_1, s_3\} = 5$, $\{s_2, s_3\} = 6$, and $S = 7$; I then creates a 7×3 matrix opt_fs associating to each of the seven signals the corresponding optimal allocation of the unit among the three states. For a generic signal A, the optimal act is such that

$$f(s) = \begin{cases} \frac{\mu(s)^2}{\sum_{s' \in A} \mu(s')^2} & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}$$

The next step is to create a list of all the possible vector of observable signals. I therefore create a $\frac{7!}{k!(7-k)!} \times k$ matrix V containing all the combinations of numbers from 1 to 7 (where numbers correspond to signals as explained above). Then, given any vector of observable signals and any signal received, I compute the associated observed signal maximizing expected utility. In the code, C is the matrix of conditional priors given the signal received. The matrix P gives, for any $\mathcal{V}$, the observed signal corresponding to each of the seven signals that the decision maker may receive.

```
C=[1 0 0; 0 1 0; 0 0 1; mu_1/(mu_1+mu_2) mu_2/(mu_1+mu_2) 0;
   mu_1/(mu_1+mu_3) 0 mu_3/(mu_1+mu_3); 0 mu_2/(mu_2+mu_3) mu_3/(mu_2+mu_3); mu_1 mu_2 mu_3];
P=zeros(factorial(7)/(factorial(k)*factorial(7-k)),7);
for t=1:factorial(7)/(factorial(k)*factorial(7-k))
    for u=1:7
        utility=-1;
        for q=1:k
            utility_provv=C(u,1)*sqrt(opt_fs(V(t,q),1)) + C(u,2)*sqrt(opt_fs(V(t,q),2))
                          + C(u,3)*sqrt(opt_fs(V(t,q),3));
            if utility_provv > utility
                utility=utility_provv;
                perc=V(t,q);
            else
            end
        end
        P(t,u)=perc;
    end
end
```

Given this result, it is possible to compute the ex-ante expected utility for each possible set $\mathcal{V}$ and find the set that, given the decision maker's initial beliefs, maximizes it. This is called *bestV* in the code below.

```
p=[p_1 p_2 p_3 p_12 p_13 p_23 p_S];
exante=-1;
for i=1:factorial(7)/(factorial(k)*factorial(7-k))
    exante_provv= p(1)*(mu_1*sqrt(opt_fs(P(i,1),1))) + p(2)*(mu_2*sqrt(opt_fs(P(i,2),2)))
                  + p(3)*(mu_3*sqrt(opt_fs(P(i,3),3)))+ p(4)*(mu_1*sqrt(opt_fs(P(i,4),1))
                  + mu_2*sqrt(opt_fs(P(i,4),2))) + p(5)*(mu_1*sqrt(opt_fs(P(i,5),1))
                  + mu_3*sqrt(opt_fs(P(i,5),3))) + p(6)*(mu_2*sqrt(opt_fs(P(i,6),2))
                  + mu_3*sqrt(opt_fs(P(i,6),3))) + p(7)*(mu_1*sqrt(opt_fs(P(i,7),1))
                  + mu_2*sqrt(opt_fs(P(i,7),2)) + mu_3*sqrt(opt_fs(P(i,7),3)));
    if exante_provv>=exante
        best_V=i;
        exante=exante_provv;
    else
    end
end
```

A more general problem is to try to determine whether a given set $\mathcal{V}$ can ever be optimal for some initial beliefs. I use the following approach. All the above computations are included into a function file

```
function [best_V2] = bestV2(k,y,vector)
mu_1=y(1); mu_2=y(2);
p_1=y(3); p_2=y(4); p_3=y(5); p_S=y(6);
```

where y is the vector of initial beliefs, while the variable *vector* denotes a specific set $\mathcal{V}$ whose optimality I want to check. At the end of the above calculations, I make the function assume value 0 if the optimal $\mathcal{V}$ is the one I am considering, and the value 1 otherwise.

```
if best_V==vector
    best_V2=0;
else
    best_V2=1;
end
```

The function is then minimized over the possible initial beliefs, checking the optimality of every possible set $\mathcal{V}$.

```
A=[-1 0 0 0 0 0; 0 -1 0 0 0 0; 0 0 -1 0 0 0; 0 0 0 -1 0 0; 0 0 0 0 -1 0;
      0 0 0 0 0 -1; 1 1 0 0 0 0; 0 0 1 1 -1 1; 0 0 1 -1 1 1; 0 0 -1 1 1 1];
b=[0 0 0 0 0 0 1 1 1 1];
optimal=ones(factorial(7)/(factorial(k)*factorial(7-k)),1);
param_values=zeros(factorial(7)/(factorial(k)*factorial(7-k)),6);
for vector=1:factorial(7)/(factorial(k)*factorial(7-k))
    f=@(y)bestV2(k,y,vector);
    [y,fval]=fmincon(f,[2/5,2/5,1/5,1/5,1/5,1/5],A,b,[],[],[],[],[],options);
    if fval < optimal(vector)
        optimal(vector)=fval;
        for i=1:6
            param_values(vector,i)=y(i);
        end
    else
    end
end
```

Since the *fmincon* algorithm only looks for local minima, I created a cycle to consider a grid of values of the parameters vector y.

C.2 Optimal Sets $\mathcal{V}$ for the Case of Three States

For the case of three states of the world, the table below gives examples of initial beliefs that make the different sets $\mathcal{V}$ optimal. Only sets for which $\bigcup\{E \colon E \in \mathcal{V}\} = S$ are included. No cases of optimality have been found for the remaining sets. Among the sets in the table, only $\{\{s_1\}, \{s_1, s_2\}, \{s_1, s_3\}\}$ is apparently never optimal.

$\bar{V}$	$\mathcal{V}$	Initial beliefs					
		$\mu(s_1)$	$\mu(s_2)$	$p(\{s_1\})$	$p(\{s_2\})$	$p(\{s_3\})$	$p(S)$
2	$\{\{s_1\},\{s_2,s_3\}\}$	$1/3$	$1/3$	$1/2$	$1/4$	$1/4$	$1/10$
	$\{\{s_1\},S\}$	$1/3$	$1/3$	$1/3$	$1/4$	$1/4$	$1/3$
	$\{\{s_1,s_2\},\{s_1,s_3\}\}$	$1/5$	$1/4$	$1/20$	$1/10$	$1/10$	$1/30$
	$\{\{s_1,s_2\},S\}$	$3/8$	$2/8$	$1/30$	$1/10$	$1/10$	$2/3$
3	$\{\{s_1\},\{s_2\},\{s_3\}\}$	$1/3$	$1/3$	$2/3$	$2/3$	$2/3$	$1/4$
	$\{\{s_1\},\{s_2\},\{s_1,s_3\}\}$	$1/10$	$1/5$	$1/4$	$1/2$	$1/3$	$1/5$
	$\{\{s_1\},\{s_2\},S\}$	$1/3$	$1/3$	$1/3$	$1/3$	$1/6$	$1/2$
	$\{\{s_1\},\{s_1,s_2\},\{s_1,s_3\}\}$	-	-	-	-	-	-
	$\{\{s_1\},\{s_1,s_2\},\{s_2,s_3\}\}$	$2/5$	$2/5$	$1/2$	$1/3$	$1/3$	$1/10$
	$\{\{s_1\},\{s_1,s_2\},S\}$	$1/20$	$1/20$	$1/10$	$1/10$	$1/10$	$1/10$
	$\{\{s_1\},\{s_2,s_3\},S\}$	$1/2$	$1/3$	$1/3$	$1/3$	$1/3$	$1/3$
	$\{\{s_1,s_2\},\{s_1,s_3\},\{s_2,s_3\}\}$	$1/3$	$1/3$	$1/6$	$1/6$	$1/6$	$1/6$
	$\{\{s_1,s_2\},\{s_1,s_3\},S\}$	$1/6$	$1/2$	$1/6$	$1/3$	$1/3$	$1/6$
4	$\{\{s_1\},\{s_2\},\{s_3\},\{s_1,s_2\}\}$	$1/3$	$1/3$	$1/3$	$1/3$	$1/2$	$1/10$
	$\{\{s_1\},\{s_2\},\{s_3\},S\}$	$1/3$	$1/3$	$1/2$	$1/2$	$1/2$	$1/3$
	$\{\{s_1\},\{s_2\},\{s_1,s_2\},\{s_1,s_3\}\}$	$1/10$	$1/10$	$1/6$	$1/4$	$1/10$	$1/10$
	$\{\{s_1\},\{s_2\},\{s_1,s_2\},S\}$	$1/10$	$1/10$	$1/6$	$1/6$	$1/3$	$1/3$
	$\{\{s_1\},\{s_2\},\{s_1,s_3\},\{s_2,s_3\}\}$	$2/5$	$2/5$	$1/2$	$1/2$	$1/6$	$1/20$
	$\{\{s_1\},\{s_2\},\{s_1,s_3\},S\}$	$1/5$	$2/5$	$1/2$	$1/2$	$1/6$	$1/3$
	$\{\{s_1\},\{s_1,s_2\},\{s_1,s_3\},\{s_2,s_3\}\}$	$1/2$	$1/4$	$1/3$	$1/10$	$1/10$	$1/10$
	$\{\{s_1\},\{s_1,s_2\},\{s_1,s_3\},S\}$	$1/5$	$1/3$	$1/6$	$1/6$	$1/6$	$1/3$
	$\{\{s_1\},\{s_1,s_2\},\{s_2,s_3\},S\}$	$1/2$	$1/4$	$1/3$	$1/10$	$1/10$	$1/3$
	$\{\{s_1,s_2\},\{s_1,s_3\},\{s_2,s_3\},S\}$	$1/3$	$1/3$	$1/10$	$1/10$	$1/10$	$1/3$
5	$\{\{s_1\},\{s_2\},\{s_3\},\{s_1,s_2\},\{s_1,s_3\}\}$	$1/5$	$1/5$	$1/5$	$1/2$	$1/2$	$1/6$
	$\{\{s_1\},\{s_2\},\{s_3\},\{s_1,s_2\},S\}$	$1/5$	$1/5$	$1/5$	$1/3$	$1/3$	$1/3$
	$\{\{s_1\},\{s_2\},\{s_1,s_2\},\{s_1,s_3\},\{s_2,s_3\}\}$	$1/5$	$1/5$	$1/5$	$1/2$	$1/6$	$1/6$
	$\{\{s_1\},\{s_2\},\{s_1,s_2\},\{s_1,s_3\},S\}$	$1/5$	$1/5$	$1/5$	$1/2$	$1/6$	$1/3$
	$\{\{s_1\},\{s_2\},\{s_1,s_3\},\{s_2,s_3\},S\}$	$2/5$	$1/5$	$1/5$	$1/3$	$1/6$	$1/3$
	$\{\{s_1\},\{s_1,s_2\},\{s_1,s_3\},\{s_2,s_3\},S\}$	$1/5$	$2/5$	$1/3$	$1/6$	$1/6$	$1/3$
6	$\{\{s_1\},\{s_2\},\{s_3\},\{s_1,s_2\},\{s_1,s_3\},\{s_2,s_3\}\}$	$1/3$	$1/3$	$1/3$	$1/3$	$1/3$	$1/10$
	$\{\{s_1\},\{s_2\},\{s_3\},\{s_1,s_2\},\{s_1,s_3\},S\}$	$1/3$	$1/3$	$1/4$	$1/3$	$1/2$	$1/4$
	$\{\{s_1\},\{s_2\},\{s_1,s_2\},\{s_1,s_3\},\{s_2,s_3\},S\}$	$1/3$	$1/3$	$1/4$	$1/4$	$1/10$	$1/3$

Bibliography

[1] Bruno de Finetti. La prevision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937.

[2] Jürgen Eichberger, Simon Grant, and David Kelsey. Updating choquet beliefs. *Journal of Mathematical Economics*, 43(7-8):888–899, 2007.

[3] Larry G. Epstein and Martin Schneider. IID: Independently and indistinguishably distributed. *Journal of Economic Theory*, 113(1):32–50, 2003.

[4] Larry G. Epstein and Kyoungwon Seo. Symmetry of evidence without evidence of symmetry. *Theoretical Economics*, 5(3):313–368, 2010.

[5] Ido Erev, Thomas S. Wallesten, and David V. Budescu. Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review*, 101(3):519–527, 1994.

[6] Pablo I. Fierens and Terrence L. Fine. Toward a chaotic probability model for frequentist probability: The univariate case. In *ISIPTA '03 Proceedings*, pages 245–259, 2003.

[7] P.I. Fierens, L.C. Rêgo, and T.L. Fine. A frequentist understanding of sets of measures. *Journal of Statistical Planning and Inference*, 139:1879–1892, 2009.

[8] Baruch Fischhoff, Paul Slovich, and Sarah Lichtenstein. Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4):552–564, 1077.

[9] Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Economic Theory*, 18(2):141–153, 1989.

[10] Dale Griffin and Amos Tversky. The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, 24:411–435, 1992.

[11] Eran Hanany and Peter Klibanoff. Updating preferences with multiple priors. *Theoretical Economics*, 2(3):261–298, 2007.

[12] Lawrence D. Phillips and Ward Edwards. Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72(3):346–354, 1966.

[13] Cesaltina P. Pires. A rule for updating ambiguous beliefs. *Theory and Decision*, 53:137–152, 2002.

[14] Eran Shmaya and Leeat Yariv. Foundations for bayesian updating. California Institute of Tech- nology working paper, 2008.

[15] Tan Wang. Conditional preferences and updating. *Journal of Economic Theory*, 108:286–321, 2003.

[16] Chee Yap. *Fundamental Problems in Algorithmc Algebra.* Oxford University Press, 2000.