Essays on Decision Theory

Thesis by
Hamed Hamze Bajgiran

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2019
Defended May 14, 2019

Hamed Hamze Bajgiran
ORCID: 0000-0002-6246-2783

ACKNOWLEDGEMENTS

This work was only possible because of the constant help and support from Caltech faculty members, friends, colleagues, and family.

First and foremost, I would like to thank my committee members: Jaksa Cvitanic, Federico Echenique, Kota Saito, and Robert Sherman.

Jaksa is the main reason I got interested in the application of mathematics in economics. He is the main reason I got into HSS program at Caltech. Even though our research topics have been different, he has always supported me with his valuable time, funding, and guidance. I am indebted to him for the rest of my life.

Federico and Kota are the main reasons I got interested in decision theory. I learned a lot in many courses I had with them. The whole thesis is the result of numerous discussions with them.

Kota was the first person who taught and made me interested in decision theory. Several parts of this thesis started in his research courses. Chapter two and the application section of chapter one are the results of the two research courses I had with him. His style of teaching made me understand the minutest details of each paper and let me develop new results. His obsession with even the smallest details of each result was the most valuable lesson I learned. I am always grateful to all his support and mentorship.

Federico is the best adviser anyone can ask. I learn a lot in every single conversation I had with him. For Chapter one, I am indebted to his course on lattice theory. Moreover, the influence of his suggestions is all over my thesis. I cannot thank him enough for giving me all his support, mentorship, and guidance.

Robert shows me how to be a fantastic teacher. He is definitely one of the best teachers in my whole life. I would like to express my most profound appreciation for his advice, guidance, and encouragement.

For many insightful discussions, I thank Michael Alvarez, Kim Border, Colin Camerer, Laura Doval, Thomas Palfrey, Charles Plott, Luciano Pomatto, Antonio Rangel, Matt Shum, Pablo Schenone, Omer Tamuz, and Leeat Yariv.

I would like to thank Laura Doval and Marcelo Fernandez for their many advice on my preparation for the job market.

I would like to thank my fellow graduate students, Nickolas Adam-Cohen, Saba Devdariani, Yimeng Li, Vadim Martynov, Tatiana Mayskaya, Alejandro Robinson-Cortes, Han Seo, and Hao Zhao.

Special thanks to my fellow office mate, Vadim Martynov. He was always the first person I explained my ideas to.

I would like to thank our division option manager, Laurel Auchampaugh, for her constant supports throughout my PhD degree. Whenever I had a problem, she was the first person to help me. I really appreciate all her time and supports.

Caltech allowed me to meet some of my best friends. During my time at Caltech, Ramses Mourhatch became one of my closest friends. I cannot thank him enough for always being there for me.

Finally, I should say that whoever I am today and every success in my life is because of my mom. As a single mom, she was the leading supporter of me throughout my life.

ABSTRACT

This thesis introduces some general frameworks for studying problems in decision theory. The purpose of this dissertation is two-fold. First, I develop general mathematical frameworks and tools to explore different decision theoretic phenomena. Second, I apply my developed frameworks and tools in different topics of Microeconomics and Decision Theory.

Chapter 1 introduces a notion of the classifier, to represent the different classes of data revealed through some observations. I present a general model of classification, notion of complexity, and how a complicated classification procedure can be generated through some simpler classification procedures.

My goal is to show how an individual's complex behavior can be derived from some simple underlying heuristics. In this chapter, I model a classifier (as a general model for decision making) that based on observing some data points classifies them into different categories with a set of different labels. The only assumption for my model is that whenever a data point is in two categories, there should be an additional category representing the intersection of the two categories. First, I derive a duality result similar to the duality in convex geometry. Then, using my result, I find all representations of a complex classifier by aggregating simpler forms of classifiers. For example, I show how a complex classifier can be represented by simpler classifiers with only two categories (similar to a single linear classifier in a neural network). Finally, I show an application in the context of dynamic choice behaviors. Notably, I use my model to reinterpret the seminal works by Kreps (1979) and Dekel, Lipman, and Rustichini (2001) on representing preference ordering over menus with a subjective state space. I also show the connection between the notion of the minimal subjective state space in economics with my proposed notion of complexity of a classifier.

In Chapter 2, I provide a general characterization of recursive methods of aggregation and show that recursive aggregation lies behind many seemingly different results in economic theory. Recursivity means that the aggregate outcome of a model over two disjoint groups of features is a weighted average of the outcome of each group separately.

This chapter makes two contributions. The first contribution is to pin down any aggregation procedure that satisfies my definition of recursivity. The result unifies aggregation procedures across many different economic environments, showing that all of them rely on the same basic result. The second contribution is to show different extensions of the result in the context of belief formation, choice theory, and welfare economics.

In the context of belief formation, I model an agent who predicts the true state of nature, based on observing some signals in her information structure. I interpret each subset of signals as an event in her information structure. I show that, as long as the information structure has a finite cardinality, my weighted averaging axiom is the necessary and sufficient condition for the agent to behaves as a Bayesian updater. This result answers the question raised by Shmaya and Yariv (2007), regarding finding a necessary and sufficient condition for a belief formation process to act as a *Bayesian updating rule*.

In the context of choice theory, I consider the standard theory of discrete choice. An agent chooses randomly from a menu. The outcome of my model is the average choice (mean of the distribution of choices) rather than the entire distribution of choices. Average choice is easier to report and obtain than the entire distribution. However, an average choice does not uniquely reveal the underlying distribution of choices. In this context, I show that (1) it is possible to uniquely extract the underlying distribution of choices as long as the average choice satisfies weighted averaging axiom, and (2) there is a close connection between my weighted averaging axiom and the celebrated Luce (or Logit) model of discrete choice.

Chapter 3 is about the aggregation of the preference orderings of individuals over a set of alternatives. The role of an aggregation rule is to associate with each group of individuals another preference ordering of alternatives, representing the group's aggregated preference. I consider the class of aggregation rules satisfying an *extended Pareto* axiom. Extended Pareto means that whenever we partition a group of individuals into two subgroups, if both subgroups prefer one alternative over another (as indicated by their aggregated preferences), then the aggregated preference ordering of the union of the subgroups also prefers the first alternative over the second one.

I show that (1) the extended Pareto is equivalent to my weighted averaging axiom, and (2) I derive a generalization of Harsanyi's (1955) famous theorem on Utilitarianism. Harsanyi considers a single profile of individuals and a variant of Pareto to obtain Utilitarianism. However, in my approach, I partition a profile into smaller groups. Then, I aggregate the preference ordering of these smaller groups using the extended Pareto. Hence, I obtain Utilitarianism through this consistent form of aggregation. As a result, in my representation, the weight associated with each individual appears in all sub-profiles that contain her.

In another application, I find the class of extended Pareto social welfare functions. My result has a positive nature, compared to the claims by Kalai and Schmeidler (1977) and Hylland (1980) that the negative conclusion of Arrow's theorem holds even with vN-M preferences.

Finally, in Chapter 4, I derive a simple subjective conditional expectation theory of state-dependent preferences. In many applications such as models for buying health insurance, the standard assumption about the independence of the utility and the set of states is not a plausible one. Hence, I derive a model in which the main force behind the separation of beliefs and state-dependent utility comes from the extended Pareto condition. Moreover, I show that, as long as the model satisfies my strong minimal agreement condition, we can uniquely separate beliefs from the state-dependent utility.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter 1

A MATHEMATICAL THEORY OF CLASSIFIERS; REPRESENTATIONS AND APPLICATIONS

1.1 Introduction

In this chapter, we introduce a mathematical framework to study the properties of classifiers. Our paper defines the notion of classifiers, complexity of a classifier, and representation of a complex classifier through simpler classifiers.

Consider a data set in which each data points has different labels. Labels of data sets allow us to classify data points into different classes, each class is associated with the data points having some specific labels.

However, in many applications and data sets, we observe a classification rather than the underlying reasons that give us the observed classification. Each class may represent some data points with some properties. How can we understand how these classes have been generated? What are the possible simple criteria that classify the data set into the observed set of classes? If we have access to simple procedures that can divide data set into two classes (binary classifiers), then how many of them are needed to generate the same classification as the one observed in our data sets?

Our paper presents a framework to answer all these questions. Generally speaking, we try to understand the underlying structure of a complex classification procedure.

In the Machine Learning literature, a neural network is built from a set of simple linear classifiers. Given a data set and a set of labels, a researcher can add many linear classifiers to build a large neural network that can shatter data points to the correct classes. More generally, a researcher can combine many different functions to form a complex function with lots of parameters to shatter the set of data points into the correct classes. One of our goals is to do the same in our abstract setup without any notion of a norm, distance, or even a topology.

Our paper tries to formalize these ideas through some new characterization of closure operators. Our results have the same intuitions that are behind the duality of convex sets and support functionals in convex analysis.

The road starts with presenting some basic terms and primitives from abstract convex geometry literature. Then, we formally define our notion of classifiers using some motivational examples. In Section 1.4 we present our main characterizations of classifiers. Then, we discuss different notions of complexity associated with a classifier. We finish the section by providing a result on how to find the minimal decomposition of a classifier into simple classification procedures.

Finally, in a self-contained Section 1.5, we provide some applications of our results in the context of dynamic choice. Notably, our results generalize and shed some new lights on seminal papers by Kreps (1979), Dekel, Lipman, and Rustichini (2001), and Gul and Pesendorfer (2001) regarding the representation of decision-maker's preference ordering over a set of menus.

Our paper provides some new techniques and frameworks that have not been fully developed and used in the economics literature. We think there are many following questions, results, and consequences in decision theory.

1.2 Primitives and Axioms

The followings are standard definitions. Let X be a finite set of alternatives. A set $A \subseteq X$ is a ***menu*** of alternatives. We denote the set of all menus of X by 2^X. A binary relation $\gtrsim$ on X is a ***weak order*** on X, if it is transitive ($\forall x, y, z \in X$ if $x \gtrsim y$ and $y \gtrsim z$ then $x \gtrsim z$), and complete ($\forall x, y \in X$ whether $x \gtrsim y$ or $y \gtrsim x$); it is ***partial order*** on X, if it is reflexive ($\forall x \in X$ $x \gtrsim x$), transitive, and anti-symmetric ($\forall x, y \in X$ if $x \gtrsim y$ and $y \gtrsim x$ then $x = y$). A partial order which is complete is called ***total order*** or ***linear order***.

Let $(X, \gtrsim)$ be a partially ordered set and $Y \subseteq X$. We say that $(Y, \gtrsim)$ is a ***chain*** if it is a totally ordered set; it is an ***anti-chain*** if no two distinct element of Y are comparable. An element $x \in X$ is said to be an ***upper bound*** of Y if $x \gtrsim a$ for every $a \in Y$. We denote the ***lowest upper bound*** or ***join*** of Y (if it exists) by $\bigvee Y$. Similarly, we denote the ***greatest lower bound*** or ***meet*** of Y by $\bigwedge Y$.

A partially ordered set $(X, \gtrsim)$ is called a ***lattice*** if each $x, y \in X$ has a join and a meet, denoted by $x \vee y$ and $x \wedge y$ respectively. Let $(X, \gtrsim)$ be a lattice and $Y \subseteq X$. We say that Y is a ***sublattice*** of $(X, \gtrsim)$ if $x \vee y, x \wedge y \in Y$ for all $x, y \in Y$.

The following two definitions are the main definitions of our paper.

Definition 1. A ***closure operator*** on X is a map $f : 2^X \to 2^X$ that satisfies the following properties:

1. ***extensivity***: $A \subseteq f(A)$ and $f(\varnothing) = \varnothing$.
2. ***idempotence***: $f(f(A)) = f(A)$.
3. ***monotonicity***: $A \subseteq B$ implies $f(A) \subseteq f(B)$.

There are two special closure operators. We define the ***identity operator*** as the closure operator $I : 2^X \to 2^X$ such that $I(A) = A$ for every $A \in 2^X$. The ***trivial closure operator*** is defined as the closure operator $f : 2^X \to 2^X$ such that $f(A) = X$ for every nonempty $A \in 2^X$.

Definition 2. A set $A \subseteq X$ is ***closed*** with respect to a closure operator $f : 2^X \to 2^X$, if $f(A) = A$. We denote the set of all closed sets with respect to the closure operator f on X as $S(X, f)$.

There is a connection between the set of closed operators and the set of menus that are closed under intersection. We do not know who first noticed this observation [1]. However, it is a known fact in topology. We are using the following lemma throughout our paper.

Lemma 1. *Let $f : 2^X \to 2^X$ be a closure operator on X, then the set of closed sets $S(X, f)$ is closed under intersection and contains $\varnothing$ and X. Moreover, if S is any subset of 2^X which is closed under intersection and contains $\varnothing$ and X, there is a unique closure operator $f_S : 2^X \to 2^X$ such that $S(X, f_S) = S$.*

Proof. The proof is simple. However, to be complete we add it here. Let $A, B \in S(X, f)$ and $C = A \cap B$. We need to show that $C \in S(X, f)$. If $C = \varnothing$, by extensivity we are done. Otherwise, monotonicity, idempotence, and closedness of A, B imply that $f(C) \subset f(A) \cap f(B) = A \cap B = C$. Thus by extensivity we get that $f(C) = C$.

[1] In the context of convex geometry, the result has been noticed in Edelman and Jamison (1985).

For the other direction, assume that $S \subset 2^X$ is closed under intersection and contains $\varnothing$ and X. Define f for any $A \subseteq X$ as follows:

$$f(A) = \bigcap_{\{s | A \subseteq s, s \in S\}} s. \tag{1.1}$$

It is straightforward to check that f is a closure operator with $S(X, f) = S$.

$\square$

The above lemma shows that by understanding the set of closed sets, we can uniquely understand the closure operator associated with it.

Finally, given any closure operator $f : 2^X \to 2^X$, the set $S(X, f)$ endowed with the meet and join operators $A \wedge B = A \cap B$ and $A \vee B = f(A \cup B)$ becomes a lattice which is bounded by $\varnothing$ and X.

Remark 1. The basic concepts of abstract convex geometry and combinatorial convex hull operator are given in Edelman et al. (1985). Any closure operator $f : 2^X \to 2^X$ with ***anti-exchange*** property defines a ***combinatorial convex hull operator***. We say that f satisfies the anti-exchange property if given any closed set A and two unequal points $x, y \in X \backslash A$, then $x \in f(A \cup y)$ implies that $y \notin f(A \cup x)$.

A combinatorial convex hull operator is the abstraction of the regular convex hull operator in the Euclidean spaces. Any closed set of a combinatorial convex hull operator defines a convex set. An element $x \in X$ is said to be an extreme point of A if $f(A \backslash x) \neq f(A)$. The anti-exchange property captures the idea that if y is an extreme point of $f(A \cup y)$ and if $x \notin f(A)$ and $x \in f(A \cup y)$ then x cannot be an extreme point. In other words, any convex set can be generated using the set of extreme points inside the set. This is the combinatorial version of the Krein–Milman theorem [2].

The connection between the combinatorial convex hull operators and the path independent choice functions has been studied in Koshevoy (1999). He shows that for any combinatorial convex hull operator, the extreme point operator (the set of extreme points of a set with respect to the convex hull operator), satisfies the ***heritage*** and ***outcast*** properties. He shows that the reverse

[2] Let X be a locally convex Hausdorff topological vector space. Also, let A be a compact convex subset of X. Then, the theorem states that A is the closed convex hull of its extreme points.

is also possible, which means that given a choice function with heritage and outcast properties, we can derive the associated convex hull operator.

In a nutshell, the heritage property states that if $A \subseteq B$, then $ext(A) \supseteq ext(B) \cap A$. This property is just the α property of Sen (1977).

The outcast property states that if $ext(A) \subseteq B \subseteq A$, then $ext(A) = ext(B)$. This property is the weaker form of property β of Sen (1977).

These properties mentioned above have been well studied in choice theory.

The connection between convex hull operator and a choice function through extreme point operator allows for connecting propositions in the literature of both choice theory and abstract convex geometry.

In particular, a characterization of a combinatorial convex hull operator through a set of primitive orderings has been studied in Richter and Rubinstein (2018). Using their representation, they propose a notion of competitive equilibrium in an abstract environment.

A part of the goal of this paper is to generalize the work of Richter et al. (2018) for the class of closure operators. We show that the closure operator has a simple representation similar to the representation of the regular Euclidean convex hull operator through continuous linear operators. Moreover, we show how to find an additive representation of a decision-maker's preference ordering of a set of menus with some underlying revealed closure operator (we explain the notion in section 1.5). The representation sheds light on the connection between the seminal works of Kreps (1979) and Dekel, Lipman, and Rustichini (2001) regarding the representation of preferences over menus.

As mentioned before, the closure operator is a generalization of the convex hull operator. Before going to the details of our characterization, we present a simple environment that captures our interpretations of closure operators. We derive our results for our proposed environment.

1.3 Labeling Correspondences and Classifiers

Let X denote the set of data points. A finite set L represents a set of labels. A ***labeling correspondence*** on X using the set of labels L is a set value function $\Phi : X \rightrightarrows L$ that associates with each data point $x \in X$ a subset of labels $\Phi(x)$.

Given a labeling correspondence $\Phi : X \rightrightarrows L$, we define a ***classifier*** as a function $f : 2^X \to 2^X$ with $f(A) = \{x | x \in X, \bigcap_{y \in A} \Phi(y) \subseteq \Phi(x)\}$ for every $A \subseteq X$.

The interpretation of a labeling correspondence is straightforward. It attaches a set of labels to each data point. We interpret each label as a single feature or property attached to each data point. Hence, attaching two different labels $l_1, l_2 \in L$ to a data point $x \in X$, $\Phi(x) = \{l_1, l_2\}$, is interpreted as if the data point x has both of those properties.

To understand the definition of a classifier, assume that the classifier f, associated with a labeling correspondence Φ. Given a data point $x \subseteq X$, $\Phi(x)$ is the set of all labels associated with the point x. To find the set of data points that are in the same class (or category) as x, we need to consider all data points that have at least all the labels of the data points x. This is precisely the definition of $f(x)$.

More generally, for a given set of data points $A \subseteq X$, $f(A)$ is the set of all data points that at least have all the labels that are in common with all points in A. The idea is that if a decision-maker wants to find all data points that are in the same class as the observed data points in A (without any other information), she should consider all points $f(A)$.

Remark 2. Notice that any classifier related to a labeling function is a closure operator. The extensivity and monotonicity properties are simple to check. To check the idempotence, just notice that by monotonicity $f(A) \subseteq f(f(A))$. Hence, we only need to show that $f(f(A)) \subseteq f(A)$.

Assume that $x \in f(f(A))$. By the definition of f, we have $\bigcap_{y \in f(A)} \Phi(y) \subseteq \Phi(x)$. Again, by the definition of f, we know that for every $y \in f(A)$ we have $\bigcap_{z \in A} \Phi(z) \subseteq \Phi(y)$. Hence, $\bigcap_{z \in A} \Phi(z) \subseteq \bigcap_{y \in f(A)} \Phi(y) \subseteq \Phi(x)$. Thus, $x \in f(A)$. As a result, f satisfies the idempotence property.

As a result of the above remark, the set of closed sets, $S(X, f)$, represents all possible different ***classes*** of data points, and $f(A)$ is the smallest class that contains A as a subset. By considering Lemma 1, we know that the set of classes is closed under the intersection. Moreover, there is a lattice structure associated with the labeling correspondence.

To understand both concepts, we provide the following example.

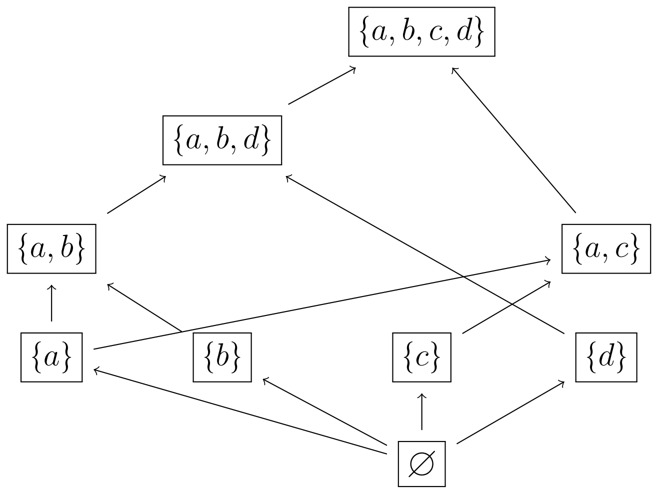

Figure 1.1: The lattice associated with the labeling correspondence Φ.

Example 1. Consider a set of four data points $X = \{a, b, c, d\}$. The set of labels is defined as $L = \{\text{human}, \text{cat}, \text{black}, \text{white}, \text{female}, \text{male}, \text{car}\}$. Assume that the labeling correspondence $\Phi : X \rightrightarrows L$ is as follows:

$$\begin{aligned}
\Phi(a) &= \{\text{human}, \text{black}, \text{female}\}\\
\Phi(b) &= \{\text{human}, \text{black}, \text{male}\}\\
\Phi(c) &= \{\text{cat}, \text{white}, \text{female}\}\\
\Phi(d) &= \{\text{car}, \text{black}\}
\end{aligned}$$

The classifier associated with the above labeling correspondence has eight classes. Class1={a} associated with the labels {human, black, female}, Class2={b} associated with the labels {human, black, male}, Class3={c} associated with the labels {cat, white, female}, Class4={d} associated with labels {car, black}, Class5={a,b} associated with the labels {human, black}, Class6={a,b,d} associated with the labels {black}, Class7={a,c} associated with the labels {female}, and the last class is Class8={a,b,c,d} associated with all data points in the set X.

Figure 1.1 depicts the structure of the classifier and the associated lattice structure.

It is important to notice that we can have another interpretation of a classifier through the set of classes. Consider the set of data points X, and a set of classes or categories $S \subseteq 2^X$. Let $x \in X$ be in two classes $A, B \in S$, which means that it has both properties of the classes A and B. Then, there should be another class $C \in S$, with both properties of the classes A and B, such that $x \in C$. The class, C, represents the data points with both properties of class A and class B. Thus, with the interpretation mentioned above, the set of classes should be closed under intersection. Moreover, we assume that there is a trivial class X, which represents the set of all data points with every possible property. Thus, through the eye of Lemma 1, there is a unique closure operator f with $S(X, f) = S$.

As a result of the above interpretation, we can define a labeling function $\Phi : X \rightrightarrows S$, which takes its labels from the set of classes S. For each data points $x \in X$, let $\Phi(x) = \{s|\ s \in S,\ x \in s\}$. Thus, for each data point $x \in X$, Φ attaches all the labels of each class that contains x. Interestingly, the classifier associated with the labeling correspondence Φ is exactly the unique closure operator f such that $S(X, f) = S$.

We summarize the above observations in the following proposition.

Proposition 1. *Let X be a set of data points. We have the followings:*

1. *Let L be a set of labels and $\Phi : X \rightrightarrows L$ be a labeling correspondence. Then, the unique classifier $f : 2^X \to 2^X$ associated with Φ (as defined above) is a closure operator. Moreover the set of classes are closed under intersection.*

2. *Let $f : X \to X$ be a closure operator on X. Then, there exists a set of labels L, and a labeling correspondence $\Phi : X \rightrightarrows L$ such that the classifier associated with Φ is f. Moreover, one choice of L and Φ is achieved by defining $L = S(X, f)$ and $\Phi : X \rightrightarrows L$ with $\Phi(x) = \{s|\ s \in S(X, f), x \in s\}$.*

Proof. We have already proved the first part in remark 2. To prove the second part, we define $L = S(X, f)$ and $\Phi : X \rightrightarrows L$ with $\Phi(x) = \{s|\ s \in S(X, f), x \in s\}$. Based on the proof of the first part, we know that the classifier $g : 2^X \to 2^X$ associated with Φ is a closure operator. Thus, we only need to prove that $g = f$.

First, we prove that $g(A) \subseteq f(A)$ for every $A \in 2^X$. Let $A \in 2^X$ and $x \in g(A)$. By the definition of g, $\bigcap_{y \in A} \Phi(y) \subseteq \Phi(x)$. Then, by definition of Φ, we have

$\{s \in S(X,f) | \ \forall y \in A, y \in s\} \subseteq \{s \in S(X,f) | \ x \in s\}$. Hence, if $s \in S(X,f)$ with $A \subseteq s$, then $x \in s$.

Now, consider the set $s = f(A)$. By monotonicity of the closure operator we have $A \subseteq s$. Since $s \in S(X,f)$ and $A \subseteq s$, then we obtain that $x \in s$. This means that $x \in f(A)$. Thus, $g(A) \subseteq f(A)$ for every $A \in 2^X$.

For the other side, we need to show that $f(A) \subseteq g(A)$ for every $A \in 2^X$. Let $A \in 2^X$ and $x \in f(A)$. By the definition of Φ and g, it remains to show that $\bigcap_{y \in A} \Phi(y) = \{s \in S(X,f) | A \subseteq s\} \subseteq \Phi(x) = \{s \in S(X,f) | \ x \in s\}$. Thus, it remains to show that if $s \in S(X,f)$ and $A \subseteq s$, then we would have $x \in s$. However, f is monotonic and s is a closed set respect to f. Thus, since $A \subseteq s$ then, $f(A) \subseteq f(s) = s$. Moreover, we assumed that $x \in f(A)$. As a result, we have $x \in s$. $\square$

In the second part of the above proposition, the labeling correspondence is not unique. However, using the second part of the proposition, we can get one choice of labeling correspondence.

Consider the classifier associated with Example 1. There are eight classes other than $\varnothing$. Let define $L = \{\text{Class1}, \ldots, \text{Class8}\}$. Using the result of Proposition 1, one choice of labeling correspondence is as follows:

$$\begin{aligned}
\Phi(a) &= \{\text{Class1}, \text{Class5}, \text{Class6}, \text{Class7}, \text{Class8}\} \\
\Phi(b) &= \{\text{Class2}, \text{Class5}, \text{Class6}, \text{Class8}\} \\
\Phi(c) &= \{\text{Class3}, \text{Class7}, \text{Class8}\} \\
\Phi(d) &= \{\text{Class4}, \text{Class6}, \text{Class8}\}
\end{aligned}$$

The above labeling correspondence is different that the original one in the example. However, they both have the same classifier with the same set of classes. In section 1.4.5, we show how many labels are needed to represent any given classifiers.

As mentioned in the introduction, in the Machine Learning literature, a neural network is built from a set of simple classifiers. Given a data set and a set of labels, a researcher can add many linear classifiers to build a large neural network that can shatter data points to the correct classes. More generally, a researcher can combine many different functions to form a complex function with lots of parameters to shatter the set of data points to the correct classes. One of our main goals is to do the same in this abstract setup without any notion of a norm, distance, or even a topology.

One of the primary results of the next section is to show that how a ***complex classifier*** (whether with lots of classes or with a complex structure of the underlying lattice) can be decomposed into simpler classifiers[3]. For example, how any classifier can be obtained by combining only ***binary classifiers***, which is the simplest non-trivial classifier. A binary classifier is a closure operator with only two different classes (other than the $\varnothing$). One class is the set of data points that satisfies a single property, and the other class is all data points. Any binary classifier is obtained by considering a labeling correspondence with only a single label. One class is the set of all data points that have the label, and the other class is the set of all data points. In this sense, a binary classifier can understand if a data point has that single label or not. In other words, it can understand if a single point is in a specific class or not. Combining these simple classifier can build any complex classifier.

In the next section, we show a general result on how any classifier can be decomposed into simpler classifiers (not only into the binary ones). Moreover, we find all possible decompositions. We also obtain the minimum number of simple classifiers that can generate a given classifier.

1.4 On Representation of Classifiers

In this section, we show how to decompose a classifier into many simpler classifiers. Our decomposition is closely related to the generation of standard convex hull operator using continuous linear functionals. Before going to the details of our results, let us remind our readers of the basic results of standard convex geometry and describe the relevant interpretation for our setup.

1.4.1 Linear Classifiers in Simple Euclidean Environment

Let $A \subseteq \mathbb{R}^n$. We denote by $Conv(A)$ the set of all convex combinations of vectors in A and $cl(A)$ as the closure of A. Let consider the linear functional $f : \mathbb{R}^n \to \mathbb{R}$. By the Fréchet–Riesz theorem any linear functional can be represented as an inner product with a vector. In other words, for any linear functional f, there exists a vector $v_f \in \mathbb{R}^n$ such that $f(\cdot) = \langle \cdot, v_f \rangle$, where $\langle \cdot, \cdot \rangle$

[3] We describe the notion of complexity in the next section. However, one might think of a lattice with lots of nodes and a large anti-chain as a complex structure.

represents the inner product in $\mathbb{R}^n$. Hence, we use the term linear functional f and the corresponding inner product representation $\langle \cdot, v_f \rangle$ interchangeably.

Let A be a bounded subset of $\mathbb{R}^n$. We define the ***support function*** of A in the direction $v \in \mathbb{R}^n$ by $h_A(v) = sup_{y \in A} \langle y, v \rangle$. The ***support hyper plane*** $\{y \in \mathbb{R}^n | \, \langle y, v \rangle = h_A(v)\}$ is the hyperplane tangent to A with the slope v. We denote the closed lower half-space that contains A and is tangent to A with the slope v by $H(v, h_A) = \{y \in \mathbb{R}^n | \, \langle y, v \rangle \leqslant h_A(v)\}$ and we call it the ***support half-space***.

By using the Separating Hyperplane Theorem (SHT), if A is a closed convex set and $x \in \mathbb{R}^n \backslash A$, then there exists a linear functional $f : \mathbb{R}^n \to \mathbb{R}$ that strictly separates A and x. In other words, there exist a positive number $c > 0$ and a vector $v \in \mathbb{R}^n$ such that $f(\cdot) =< \cdot, v >$ and $\forall y \in A \;\; f(y) < c < f(x)$. Hence, $x \notin H(v, h_A)$.

As a result of the previous discussion, if A is a closed convex set, we can write it in terms of the intersection of supporting half-spaces. More generally, for any bounded set $A \in \mathbb{R}^n$ we have the following dual property:

$$cl(Conv(A)) = \bigcap_{v \in \mathbb{R}^n} H(v, h_A). \tag{1.2}$$

There is a nice interpretation of Equation 1.2 related to the purpose of our paper. Assume that a decision-maker tries to classify different shapes in $\mathbb{R}^n$. If she has only access to the support functions, then she can distinguish two shapes that have different convex hulls. However, if two different shapes have the same convex hull, then from her perspective there is no difference between them.

In another more intuitive interpretation, a ***simple linear classifier*** is a closed half-space $H(v, a) = \{y \in \mathbb{R}^n | \langle y, v \rangle \leqslant a\}$ for some $v \in \mathbb{R}^n, a \in \mathbb{R}$. Given a data point x, a simple linear classifier can detect if the point is inside or not. More generally, given a set of point $A \in \mathbb{R}^n$, it can recognize if the whole set A is inside or not. A decision-maker that has access to all simple linear classifiers can detect differences between two shapes with the different convex hulls. However, any two different shapes with the same convex hull are similar through the eyes of any simple linear classifiers.

Notice that based on the above discussion, if a decision-maker wants to detect two different shapes with the same convex hull, she needs to have access to more complex classifiers (something that is not built using only linear classifiers in the above sense). However, before going to the details of our main results consider the following example.

Example 2. Let consider a decision-maker that tries to classify bounded shapes in $\mathbb{R}^2$. Assume that she has access to simple linear classifiers in the direction of $v_1 = (1, 0), v_2 = (0, 1)$. In other words, given any bounded set $A \in \mathbb{R}^2$ and any closed half-space $H(v, a) = \{y \in \mathbb{R}^n | \langle y, v\rangle \leqslant a\}$ with $v \in \{v_1, v_2\}$ and $a \in \mathbb{R}$, she can detect if A is inside $H(v, a)$ or not. Then from her perspective, there is no difference between the unit circle at the origin and a rhombus with congruent diagonals of length two which is centered at the origin. From her perspective, both are similar to a square with length two which is centered at the origin.

She can classify previous two shapes into two different classes, if she can access to another linear classifier in another direction, such as the direction of $v_3 = (1, 1)$.

However, using only v_1, v_2 she can still classify many different shapes, as long as they are different from the perspective of at least one of the linear classifiers of the form $H(v, a) = \{y \in \mathbb{R}^n | \langle y, v\rangle \leqslant a\}$ with $v \in \{v_1, v_2\}$ and $a \in \mathbb{R}$. Given two bounded data sets $A, B \subseteq \mathbb{R}^2$, she first finds the smallest rectangle f_A that contains A, and the smallest rectangle f_B that contains B. Then, she compares f_A with f_B. If both f_A and f_B are the same, then from her perspective A and B are the same. Otherwise, she classifies them as different categories.

Considering the above decision-maker, we define the function f from the set of bounded subsets of $\mathbb{R}^2$ to the set of rectangles in $\mathbb{R}^2$, which associates with any bounded $A \subseteq \mathbb{R}^2$ the smallest rectangle that contains A. It can be shown that f is a closure operator. The class of each bounded shape $A \in \mathbb{R}^2$ is just $f(A)$. Hence, the decision-maker classifies different shapes according to f.

Now, we consider the reverse problem in example 2. Assume that we observe that the decision-maker classifies different shapes using f. How can we decompose f to simple linear classifiers built using v_1, v_2? The more general question is that if a set of different classes revealed through the observation of the labels of a data set, how can we decompose the revealed classifier into its underlying simpler classifiers?

In the next section, our goal is to answer the previous questions for any general classifier in our abstract setup of section 1.3.

1.4.2 General Representation of Classifiers

In this section, we assume that X is a finite set of data points. Based on our discussion of section 1.3, we use the terms closure operator and classifier interchangeably. Our first proposition is rewriting the duality property of Equation 1.2 for the class of abstract classifiers (closure operators). To do that, we define the following notations.

Let $\succsim$ be a weak order on X and $A \subseteq X$. We denote the set of all weak orders on X by $\mathcal{R}$. The role of $\mathcal{R}$ for X is the same as the role of the dual of $\mathbb{R}^n$ (which is the set of all continuous linear functional, which in the case of finite dimensional vector spaces is the space itself) for $\mathbb{R}^n$. We define the ***support function*** of A, $h_A : \mathcal{R} \to 2^X$, by $h_A(\succsim) = \{x \in A |\ x \succsim y\ \forall y \in A\}$. Notice that the definition is similar to the Euclidean case. However, instead of the supremum of the inner product, we use all maximal elements of the set A with respect to the direction $\succsim$.

Similar to the Euclidean case, we define the ***support half-space*** of the set $A \neq \emptyset$ with respect to the order $\succsim$ by $H(\succsim, h_A) = \{x \in X |\ h_A(\succsim) \succsim x\}$. We define $H(\succsim, h_\emptyset) = \emptyset$.

Our first proposition is the general version of Equation 1.2 for our abstract environment. The result characterizes any closure operator as a combination of weak orders.

Proposition 2. *Let $f : 2^X \to 2^X$ be a closure operator (classifier). Then there exist some weak orders $\succsim_1, \ldots, \succsim_k$ on X, such that:*

$$f(A) = \bigcap_{i \in \{1,\ldots,k\}} H(\succsim_i, h_A). \tag{1.3}$$

Moreover, for any given set of week orders $\succsim_1, \ldots, \succsim_l$ on X, the operator $g : 2^X \to 2^X$ defined by $g(A) = \bigcap_{i \in \{1,\ldots,l\}} H(\succsim_i, h_A)$ is a closure operator.

Proof. First, we prove that if f is a closure operator, then there exists a set of weak orders $\{\succsim_1, \ldots, \succsim_k\}$ that generates it as in Equation 1.3. Consider the set of closed sets of f, which is $S(X, f)$. For any $C \in S(X, f)$, we associate a single weak order $\succsim_C$ as follows:

- for every $x, y \in C$, let $x \sim_C y$,
- for every $x, y \in X \backslash C$, let $x \sim_C y$,
- for every $x \in C$ and $y \in X \backslash C$, let $x \succ_C y$.

For any given $A \in 2^X$ and for any given $\succsim_C$ as as defined above, we have:

$$H(\succsim_C, h_A) = \begin{cases} \varnothing & \text{if } A = \varnothing, \\ C & \text{if } A \subseteq C, \\ X & \text{otherwise.} \end{cases} \tag{1.4}$$

We will show that $\{\succsim_C \mid C \in S(X, f)\}$ generates f, as in Equation 1.3.

Let $A \subseteq 2^X$ and $C \in S(X, f)$. Consider the following cases:

If $A \subseteq C$, then closedness of C and monotonicity of f together imply that $f(A) \subseteq f(C) = C$. Therefore, by Equation 1.4, $H(\succsim_C, h_A) = C$ and $f(A) \subseteq H(\succsim_C, h_A)$.

If $A \nsubseteq C$, then $H(\succsim_C, h_A) = X$. Therefore, $f(A) \subseteq H(\succsim_C, h_A)$.

As a result of the previous cases $f(A) \subseteq \bigcap_{C \in S(X,f)} H(\succsim_C, h_A)$. Since f is a closure operator, we have $f(A) \in S(X, f)$. Let $\succsim_{f(A)}$ represents the weak ordered associated with $f(A)$. Since $A \subseteq f(A)$, by Equation 1.4, $H(\succsim_{f(A)}, h_A) = f(A)$. Thus, $\bigcap_{C \in S(X,f)} H(\succsim_C, h_A) \subseteq H(\succsim_{f(A)}, h_A) = f(A)$. Therefore, $f(A) = \bigcap_{C \in S(X,f)} H(\succsim_C, h_A)$. The last result completes the first part of the proposition.

For the other direction, let $\{\succsim_1, \ldots, \succsim_k\}$ be a set of weak orders. We show that the operator f, defined as in Equation 1.3, is a closure operator.

We will check all the properties of the closure operator.

By the definition of support half-space, $H(\succsim_i, h_\varnothing) = \varnothing$ for every i. Hence, $f(\varnothing) = \varnothing$. Moreover, for every A and every i, we have $A \subseteq H(\succsim_i, h_A)$. Therefore, f is extensive.

Next, we show that f is monotonic. Let $A \subseteq B$ and $\succsim_i \in \{\succsim_1, \ldots, \succsim_k\}$. By the definition of support function, $h_B(\succsim_i) \succsim_i h_A(\succsim_i)$. Hence, $H(\succsim_i, h_A) \subseteq H(\succsim_i, h_B)$. Therefore, f is monotonic.

Finally, we show that f is idempotence. Since we have already shown that f is monotonic, we only need to show that $f(f(A)) \subseteq f(A)$ for every A.

Let $x \in f(f(A))$. By the definition of f, we have $x \in H(\succsim_i, h_{f(A)})$ for every $i \in \{1, \ldots, k\}$. As a result, we have $y \succsim_i x$ for all $y \in h_{f(A)}(\succsim_i)$.

Let $y \in h_{f(A)}(\succsim_i)$. Notice that $y \in h_{f(A)}(\succsim_i)$ implies that $y \in f(A) = \bigcap_{i \in \{1,\ldots,l\}} H(\succsim_i, h_A) \subseteq H(\succsim_i, h_A)$. Therefore, since $y \succsim_i x$ and $y \in H(\succsim_i, h_A)$, then we have $x \in H(\succsim_i, h_A)$. Therefore, $x \subseteq \bigcap_{i \in \{1,\ldots,l\}} H(\succsim_i, h_A) = f(A)$. The last result completes the proof. □

As a result of the above proposition, we say that the set of weak orders $\{\succsim_1, \ldots, \succsim_k\}$ ***generates*** the closure operator f, if $f(A) = \bigcap_{\succsim_i \in \{1,\ldots,k\}} H(\succsim_i, h_A)$.

There are many useful properties of the closure operators generated by a single weak order. To exploit them, consider $\succsim \in \mathcal{R}$. We denote the closure operator generated by $\succsim$ by $f_\succsim$. We call a closure operator (classifier) that generated by a single weak order a ***simple closure operator*** or a ***simple classifier***. The first observation is that $\forall A \in 2^X$, $f(A) = H(\succsim, h_A)$. Thus, the set of closed sets (classes) is $S(X, f_\succsim) = \{H(A, h_A) |\ A \subseteq X\}$. Moreover, $S(X, f_\succsim)$ is a single chain respect to the set inclusion. In other words, the lattice associated with $f_\succsim$ is a total order. Given any two classes of $f_\succsim$, one of them is the subset of the other one. The role of simple closure operators in our setup is similar to the role of continuous linear functionals in the previous section.

Interestingly, the reverse is also correct. If a lattice associated with a classifier f is a single chain (total order), then it is generated by a single unique weak order.

Lemma 2. *The lattice associated with a closure operator is a single chain if and only if a single weak order generates it. Moreover, the weak order generating the lattice is unique.*

Proof. If $\succsim$ generates f, then Proposition 2 guaranties that f is a closure operator. By the definition of f, the set of closed sets is $S(X, f_\succsim) = \{H(\succsim$

$, h_A)|\ A \subseteq X\}$. Using the definition of support half-space, $S(X, f_{\succsim})$ is a single chain respect to the set inclusion.

For the other side. If f is a closure operator such that $S(X, f)$ is a single chain, then define $a \succsim b$ if and only if $f(a) \supseteq f(b)$. Since $S(X, f)$ is a total order, $\succsim$ is a weak order.

To show that $\succsim$ generates f, we need to show that for every $A \in 2^X$ we have $f(A) = H(\succsim, h_A)$.

Notice that there should be some $x \in A$ with $f(x) = f(A)$. Otherwise, since $S(X, f_{\succsim})$ is a single chain, there should be a proper subset of $f(A)$ which contains all the closure of the singleton subsets of A, which is not correct.

Now, consider any $y \in f(A)$. Since $f(y) \subseteq f(x)$, then $x \succsim y$. Thus, $y \in H(\succsim, h_A)$. As a result, $f(A) \subseteq H(\succsim, h_A)$.

For the other side, since $f(x) = f(A)$, then x should be a maximal element in $H(\succsim, h_A)$. Hence, for all $y \in H(\succsim, h_A)$ we have $f(y) \subseteq f(x) = f(A)$. As a result $H(\succsim, h_A) \subseteq f(A)$. The last result completes the proof.

□

As a result of the previous lemma, there is no difference between the sets of weak orders, classifiers with a single chain, and support half-spaces generated by a weak order. All these three different sets are representing the class of simple classifiers.

Considering lemma 1, we can rewrite Proposition 2 in a different way:

Corollary 1. *Let $f : 2^X \to 2^X$ be a closure operator (classifier). Then there exist some simple closure operators (simple classifiers) $g_1, \ldots, g_k$ on X, such that:*

$$f(A) = \bigcap_{i \in \{1,\ldots,k\}} g_i(A). \tag{1.5}$$

Moreover, for any given set of simple closure operators $g_1, \ldots, g_k$ on X, the operator $g : 2^X \to 2^X$ defined by $g(A) = \bigcap_{i \in \{1,\ldots,l\}} g_i(A)$ is a closure operator.

Considering Proposition 2 and corollary 1, then the following observations are immediate:

Lemma 3. *Let $\succeq_1, \cdots, \succeq_k$ (or similarly the corresponding simple closure operators $g_1, \ldots, g_k$) generate the closure operator f and $S(X, f)$ is the closed sets of f. We have the following:*

1. *$\succeq_1, \ldots, \succeq_k$ (or similarly $g_1, \ldots, g_k$) are not unique to get f.*
2. *$\forall A \in 2^X$ $H(\succeq_i, h_A)$ (or $g_i(A)$) are all in $S(X, f)$.*
3. *If $\{\succeq_{i_1}, \cdots, \succeq_{i_l}\}$ is a subset of $\{\succeq_1, \cdots, \succeq_k\}$, by defining the closure operator h generated by $\succeq_{i_1}, \cdots, \succeq_{i_l}$ we have $S(X, h) \subseteq S(X, f)$*
4. *If $\succsim_n$ is a weak order and g is generated by $\{\succsim_n\} \cup \{\succeq_1, \ldots, \succeq_k\}$, then $g = f$ if and only if $S(X, f_{\succsim_n})) \subseteq S(X, f)$.*

Proof. We are going to prove a more general version of the result in Proposition 4. □

The above observations are crucial to build the smallest number of weak orders that can generate a closure operator f.

Before going into details of the minimal representation, we present a more general representation of the closure operators.

Equation 1.5 gives us a hint to generalize the representations of the closure operators. Consider a set of closure operators (classifiers) $\{g_1, \ldots, g_k\}$, where each g_i is not necessary a simple closure operator. We define an operator f as follows:

$$f(A) = \bigcap_{i \in \{1,\ldots,k\}} g_i(A). \tag{1.6}$$

The next proposition shows that f is a closure operator. As a result, we call f the classifier generated by the set of classifiers $\{g_1, \ldots, g_k\}$.

Proposition 3. *Consider any set of closure operators (classifiers) $\{g_1, \ldots, g_k\}$ on X. If f is the operator generated by them, as in Equation 1.6, then f is a closure operator.*

Proof. Extensivity and monotonicity of f come from extensivity and monotonicity of each of $g_1, \ldots, g_k$. For the idempotence property, we need to show that $f(f(A)) = f(A)$. By monotonicity of f, we only need to show that $f(f(A)) \subseteq f(A)$.

Let x be in $f(f(A))$. By the definition of f, for all $i \in \{1, \ldots, k\}$ $x \in g_i(f(A))$. Again by the definition of f, for all $i \in \{1, \ldots, k\}$ $x \in g_i(\bigcap_{j\in\{1,\ldots,k\}} g_j(A))$. Since $\bigcap_{j\in\{1,\ldots,k\}} g_j(A) \subseteq g_i(A)$ and g_i is monotonic, then $g_i(\bigcap_{j\in\{1,\ldots,k\}} g_j(A)) \subseteq g_i(g_i(A))$. But, since g_i is a closure operator, then $g_i(g_i(A)) = g_i(A)$. As a result, $g_i(\bigcap_{j\in\{1,\ldots,k\}} g_j(A)) \subseteq g_i(A)$. Thus, for all $i \in \{1, \ldots, k\}$ we have $x \in g_i(A)$, which means that $x \in \bigcap_{j\in\{1,\ldots,k\}} g_j(A)$. Again by the definition of f, the last result shows that $x \in f(A)$. Hence, $f(f(A)) \subseteq f(A)$, which completes the proof.

$\square$

Both Proposition 2 and 3 are about the representation of closure operators. However, they are silent on how to check if a set of closure operators $\{g_1, \ldots, g_k\}$ generates a closure operator f or not.

The next proposition is the answer to the previous question. The following result is the most general representation of the closure operators.

Proposition 4. *Let f and $g_1, \ldots, g_k$ be closure operators(not necessarily simple) on X. The closure operator f is generated by $g_1, \ldots, g_k$ if and only if:*

1. *$S(X, g_i) \subseteq S(X, f)$ for all $i \in \{1, \cdots, k\}$,*
2. *If $A \in S(X, f)$ and if $x \notin A$, then there exists a closure operator $g_i \in \{g_1, \ldots, g_k\}$ such that $x \notin g_i(A)$.*

Proof. First, we show that if f is generated by closure operators $g_1, \ldots, g_k$, then we have both conditions.

To prove the first condition, observe that by Proposition 3, f is a closure operator. Let $g_i \in \{g_1, \ldots, g_k\}$ and $A \in S(X, g_i)$. Since every $g_j \in \{g_1, \ldots, g_k\}$ is a closure operator, then g_j is monotonic. Hence, for every g_j, $A \subseteq g_j(A)$. Moreover, since $A \in S(X, g_i)$, then $g_i(A) = A$. As a result, $A = \bigcap_{j\in\{1,\ldots,k\}} g_j(A)$, which means that $A = f(A)$. Thus, $A \in S(X, f)$. Hence, for every $i \in \{1, \ldots, k\}$ we have $S(X, g_i) \subseteq S(X, f)$.

To prove the second condition, let $A \in S(X, f)$ and $x \notin A$. By the definition of f, $x \notin A = f(A) = \bigcap_{j\in\{1,\ldots,k\}} g_j(A)$. Therefore, there exists some $i \in \{1, \ldots, k\}$ such that $x \notin g_i(A)$. Thus, we complete the proof of the second condition.

For the other side of the proposition, we assume that both conditions are satisfied. We show that $f(A) = \bigcap_{j\in\{1,\dots,k\}} g_j(A)$.

By the first condition, for all $i \in \{1,\dots,k\}$ $S(X, g_i) \subseteq S(X, f)$. Hence, for every $A \in 2^X$ and for every $i \in \{1,\dots,k\}$ we have $A \subseteq g_i(A) \subseteq f(A)$. Therefore, $A \subseteq \bigcap_{j\in\{1,\dots,k\}} g_j(A) \subseteq f(A)$. Since $S(X, f)$ is closed under intersection and $g_i(A) \subseteq\ S(X, g_i) \subseteq S(X, f)$, we also have $\bigcap_{j\in\{1,\dots,k\}} g_j(A) \in S(X, f)$.

Since f is a closure operator, it is monotonic. Applying f to all terms of $A \subseteq \bigcap_{j\in\{1,\dots,k\}} g_j(A) \subseteq f(A)$, gives us $f(A) \subseteq f(\bigcap_{j\in\{1,\dots,k\}} g_j(A)) \subseteq f(f(A))$. By idempotence property of f, we have $f(A) \subseteq f(\bigcap_{j\in\{1,\dots,k\}} g_j(A)) \subseteq f(A)$, which results in $f(A) = f(\bigcap_{j\in\{1,\dots,k\}} g_j(A))$. But we have already shown that $\bigcap_{j\in\{1,\dots,k\}} g_j(A) \in S(X, f)$. Therefore, $f(A) = f(\bigcap_{j\in\{1,\dots,k\}} g_j(A)) = \bigcap_{j\in\{1,\dots,k\}} g_j(A)$. Thus, we show that $f(A) = \bigcap_{j\in\{1,\dots,k\}} g_j(A)$. The last result completes the proof. $\square$

There are some comments regarding the above proposition. Assume that the classifier f is generated by $g_1, \dots, g_k$. By the first condition of the proposition, the set of classes (or closed sets) of f is finer than any of the classifiers g_i. It means that f can separate data points into more classes than any of g_i. Moreover, there is no class that any of g_i can detect and f can not.

The second condition is similar to the Separating Hyper Plane Theorem in convex geometry. It means that if A is a class of f and $x \notin A$, then there should be a separating classifier $g_i \in \{g_1, \dots, g_k\}$ that detect that x is not in the class of A respect to the classifier g_i (which in the eyes of g_i, the class of A is $g_i(A)$, which is not necessarily $f(A)$).

The previous set of propositions let us fully understand different structures behind closure operators (classifiers). As a result, we have the following hierarchy of classifiers.

1.4.3 Hierarchy of Classifiers

A classifier $f : 2^X \to 2^X$ is more ***complex*** than a classifier $g : 2^X \to 2^X$ if $S(X, g) \subseteq S(X, f)$. In other words, in a virtue of Proposition 4, by combining some other classifiers with g we can generate f.

We have already defined a ***simple classifier*** f as a closure operator with a single chain structure underlying its associated lattice $S(X, f)$. More generally, we say that f is a ***simple classifier with length*** k, whenever the number of closed sets (classes) generated by f, $|S(X, f)|$, is $k+1$. Notice that $S(X, f)$ always contains $\varnothing$ and X. Hence, a simple classifier with length k has k nonempty different classes.

The ***trivial classifier*** is the closure operator with only two closed sets $\varnothing$ and X. The trivial classifier does not classify data points at all. It only understands that we have a data set X. It might represent a data set with no label or a data set with all data points having the same labels. As a result, there are no differences between our data points.

The simplest possible non-trivial classifier is a classifier with length two, which we call the ***binary classifier***. Any binary classifier can be built using a labeling correspondence with only one label. The binary classifier can detect if a data point has a single property or not. In other words, it can detect if a data point is in a single class associated with a single property or not.

Remark 3. Using the proof of Lemma 2, any simple classifier with length k is associated with a unique weak order with k different indifference classes. Let $\{A_1, \ldots, A_k\}$ be a partition of X into the indifference classes of a weak order $\gtrsim$ with $x \succ y$ whenever $y \in A_i, x \in A_j$ such that $i < j$. Then, $\{A_1, (A_1 \cup A_2), \ldots, (A_1 \cup, A_2 \cup \ldots \cup, A_k)\}$ is the set of closed sets of a simple classifier with length k. The reverse can be done in the same way.

Remark 4. Consider a binary classifier $f : 2^X \to 2^X$. Assume the that the corresponding set of closed sets is $S(X, f) = \{\varnothing, A, X\}$. Hence, the corresponding weak order $\gtrsim_f$ is $x \gtrsim y$ iff $x \in A, y \in X \backslash A$. From the perspective of the classification there is no difference between f and another binary classifier g with $S(X, g) = \{\varnothing, X \backslash A, X\}$. The weak order associated with g is the reverse order of $\gtrsim_f$. From The perspective of labeling correspondence, one can build the classifier f by attaching a single label l_1 to the points in A. The second one, g, can be built by attaching l_2 to $X \backslash A$, representing not having l_1.

Assume that we can infer a classifier f through a set of data points X. As discussed in the previous section, we want to find a representation of f using the simpler classifiers. For example, we may ask a decision-maker about different classes of goods in a supermarket. However, she might classify object through different labels or different criteria like different brands, sizes, colors, being closer to the entrance of a supermarket, and many more. Her complex classifier can be built based on simple classifiers that combine these single dimensional properties. Proposition 4 help us to identify all different possible representations.

Example 3. Let us revisit Proposition 2. We have already proved that any given classifier f can be generated by some weak orders $\succsim_1, \ldots, \succsim_k$. However, we want to use Proposition 4 to build some binary classifiers $\{g_1, \ldots, g_k\}$ to generate f. One simple way is to notice that for any binary classifier g_i, $S(X, g_i)$ should be a subset of $S(X, f)$. Moreover, for any class $\varnothing \neq A \in S(X, f)$ and $x \notin A$, there should be one of g_i to separate x and A. Hence, if we consider all binary classifiers g_A for every class A in $S(X, f)$, both requirements of Proposition 4 will be satisfied.

In the example above, the representation generated uses many binary classifiers. In fact, we used $|S(X, f)| - 1$ binary classifiers. In the next section, we show how we can get the minimum number of binary classifiers to generate a complex classifier. However, we can decrease the number of classifiers, if instead of using only binary classifiers we use simple classifiers with different lengths.

One way to do that (although it is not the best way) is to decompose the lattice $S(X, f)$ into chains. Consider sublattices of $S(X, f)$, each starting with $\varnothing$ and ends with X with only a single chain structure. If we consider some of the mentioned sublattices that cover the whole $S(X, f)$, the associated simple classifiers can generate f.

Consider the following example.

Example 4. Let $X = \{a, b, c\}$ be a set of data points and $f : 2^X \to 2^X$ be a closure operator with the set of closed sets $S(X, f) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, b, c\}\}$. We consider the following three chains:

$$
\begin{aligned}
S(X, g_1) &= \{\varnothing, \{a\}, \{a, b\}, \{a, b, c\}\} \\
S(X, g_2) &= \{\varnothing, \{b\}, \{a, b\}, \{a, b, c\}\} \\
S(X, g_3) &= \{\varnothing, \{c\}, \{a, b, c\}\}
\end{aligned}
$$

Notice that, since both conditions of Proposition 4 are satisfied, then $\{g_1, g_2, g_3\}$ generates f. Figure 1.2 illustrates the decomposition.

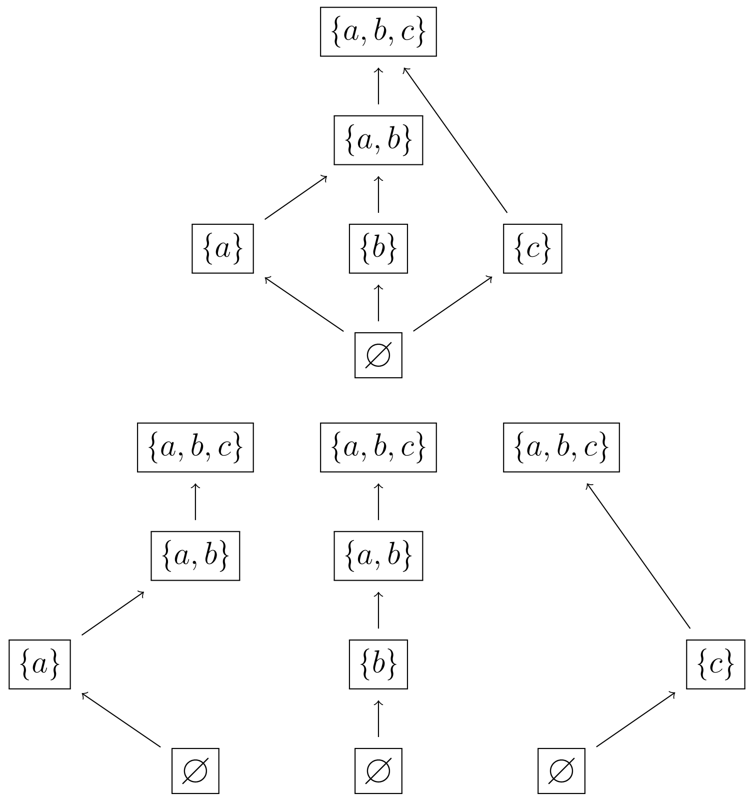

Figure 1.2: The lattice associated with the closure operator f is in the top. The associated decomposition into g_1, g_2, g_3 is in the bottom.

The above example illustrates one way to find the underlying simple classifiers that can generate a closure operator. In section 1.4.5, we discuss how many simple classifiers are needed in the minimal representation. Before that, we provide some comments on the notion of complexity that we defined.

1.4.4 Some Comments on the Notion of Complexity

In the previous section, in the virtue of Proposition 4, we define the notion of complexity over the set of classifiers which are defined over a given data set X. We say that, the classifier f is more complex than a classifier g if $S(X, g) \subseteq S(X, f)$. This notion of complexity induces a reflexive and transitive order over the set of classifiers defined over X. However, it is not a complete order. To be able to compare different classifier, there might be different ways to make it complete. We explain some.

First approach, for every classifier the associated lattice gives us some hints on how to define the order "more complex classifier". There are some constants associated with every lattice (or more generally with every partially ordered set).

- The first one is the number of non-empty closed sets (classes) that f can detect, which is $|S(X,f)|-1$. We might think that a classifier that can detect a larger number of classes is a more complex classifier than the other one which can detect less. Any classifier over the data set X can detect at most $2^{|X|}-1$ number of nonempty classes.

- The second one is the length of the largest chain, not including the empty set, in the lattice associated with a classifier, which we call it the ***depth of the classifier***. The depth of the classifier depends on the maximum number of subcategories which the classifier can detect. The largest possible depth of a classifier over the set X is $|X|$.

- The third one is the length of the largest antichain in the associated lattice, which we defined as the ***width of the classifier***. Through Dilworth's theorem (we will cover it in the next section), this number depends on the minimum number of chains that can cover the structure of the lattice. If we think of a single chain as a one-dimensional property, the width of a classifier is kind of the dimension of a classifier (as a decomposition into the minimum number of chains)

Depending on applications, one might use one or many of the numbers mentioned above to compare two different classifiers. For example, the sum of all three numbers is one possible choice. Notice that all three respect the notion of complexity that we defined at first. In other words, if f is more complex than g, in the sense of $S(X,g) \subseteq S(X,f)$, then the number of classes that f can detects, depth of f, and width of f is more than the corresponding numbers associated with g.

Second approach is through the lens of Proposition 2. As we have seen, any simple classifier on X is generated through one weak order over X. Hence, one notion of complexity might be defined as the minimum number of weak order ***(MNWO)*** that can generate a classifier. As we see in the next section, the minimum number of weak order that can generate a classifier is bounded by the width of the classifier, but not necessarily the same. This notion of complexity also respects our first notion of complexity. Again, if f is more complex than g, in the sense of $S(X,g) \subseteq S(X,f)$, then the minimum number of weak order needed to generate f is more than the minimum number of weak order to generate g.

The last approach is through the lens of Example 3. Any classifier can be generated with a finite number of binary classifiers, which are the simplest possible non-trivial classifiers. As a result, another notion of complexity might be defined as the minimum number of binary classifiers ***(MNBC)*** that can generate a classifier. Again, this notion of complexity also respects our first notion.

Remark 5. The last notion of complexity, through the decomposition into binary classifiers, is a trade-off between all other notions of complexity. For any given classifier f, it depends on the number of the classes of f, depth of f, width of f, and even the minimum number of simple classifiers that can generate f.

For example, it is not hard to show that any simple classifier with length k has depth k, width one, and the minimum number of binary classifiers needed to generate it is $k-1$. In the other direction, for any classifier with width k and depth two , the minimum number of binary classifiers needed to generate it is k.

Let us present some examples.

Example 5. Let $X = \{a, b, c, d\}$ be the set of data points. Consider the classifiers f_1, f_2 defined as follows:

$$S(X, f_1) = \{\varnothing, \{a\}, \{b\}, \{a, b, c, d\}\}$$
$$S(X, f_2) = \{\varnothing, \{a\}, \{a, b\}, \{a, b, c, d\}\}$$

Figure 1.3, illustrates the underlying lattice structures. The classifier f_1 has depth two and width two. The number of nonempty classes that it can detect is three. The minimum number of simple classifiers that can generate it is two. Finally, the minimum number of binary classifiers that can generate it is two.

The classifier f_2 has depth three and width one. The number of non-empty classes is three. The minimum number of simple classifiers that can generate

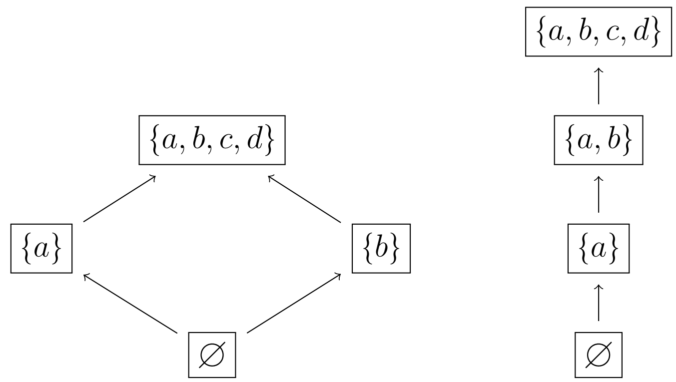

Figure 1.3: The lattices associated with the closure operator f_1 (the left one) and f_2.

it is one. The minimum number of binary classifiers that can generate it is two.

All numbers are easily seen through the lattices associated with f_1, f_2. We can observe that both have the same number of classes and the same MNBC (minimum number of binary classifiers that can generate it). Moreover, the sum of their width and depth is also the same.

However, in some cases, the numbers are not as easy to compute as f_1, f_2. Notably, in case of the minimum number of binary classifiers that are needed to decompose a classifier. The following classifiers have interesting associated values.

Consider the following classifiers f_3 and f_4:

$$S(X, f_3) = \{\varnothing, \{a\}, \{a,b\}, \{a,c\}, \{a,b,c,d\}\}$$
$$S(X, f_4) = \{\varnothing, \{a\}, \{a,b\}, \{a,b,c\}, \{a,b,c,d\}\}$$

Figure 1.4, illustrates the associated lattices. The width, depth, and the number of classes can easily be seen. The classifier f_3 has depth three, width two, and four nonempty classes. Similarly, the classifier f_4 has depth four, width one, and four nonempty classes.

Both have the same number of classes. Moreover, both have the same sum of width and depth. Two simple classifiers can generate the classifier f_3, and the classifier f_1 can be generated by one simple classifier.

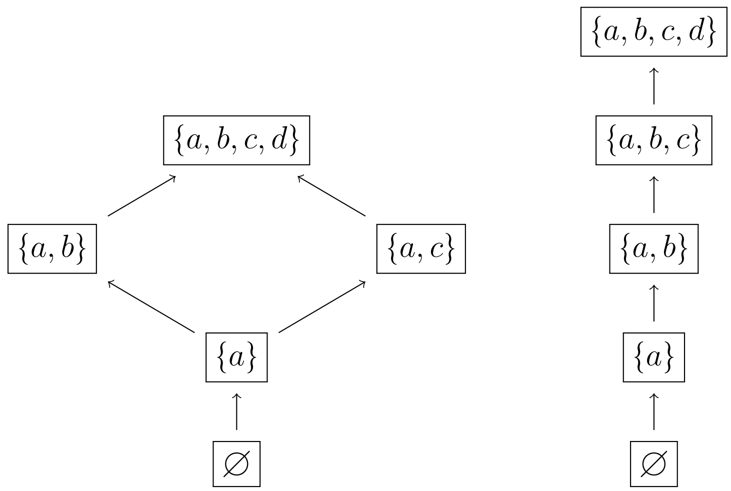

Figure 1.4: The lattices associated with the closure operator f_3 (the left one) and f_4.

However, when it comes to the decomposition into the binary classifiers, we need to be more careful.

The classifier f_4 can be generated with three binary classifiers. It is not hard to see that the following classifiers are the only minimal representations of f_4 through binary classifiers:

$$S(X, f_{4,1}) = \{\varnothing, \{a\}, \{a,b,c,d\}\}$$
$$S(X, f_{4,2}) = \{\varnothing, \{a,b\}, \{a,b,c,d\}\}$$
$$S(X, f_{4,3}) = \{\varnothing, \{a,b,c\}, \{a,b,c,d\}\}$$

Proposition 4 is the reason why we need all three.

Now consider the classifier f_3. At first sight, we might think we need at least three binary classifiers to generate it. However, surprisingly, in this case, we only need two binary classifiers. We define the classifiers $f_{3,1}, f_{3,2}$ as follows:

$$S(X, f_{3,1}) = \{\varnothing, \{a,b\}, \{a,b,c,d\}\}$$
$$S(X, f_{3,2}) = \{\varnothing, \{a,c\}, \{a,b,c,d\}\}$$

By the second condition in Proposition 4, since $\{a\} = \{a,b\} \cap \{a,c\}$, we can see that $f_{3,1}, f_{3,2}$ generate f_3.

Hence, if we think about MNBC as a notion of complexity, then interestingly f_4 is more complex than f_3.

In the next section, we discuss in more detail how many simple classifiers and binary classifiers are needed to generate a given classifier.

1.4.5 How Many Simple classifiers are Needed to Generate a Classifier?

Consider example 4 and the discussion right before that about how to decompose a classifier into simple classifiers that generate it. In this form of decomposition, the minimum number of the simple classifiers needed to generate a closure operator f can be bounded using the minimum number of chains that can cover the underlying lattice $S(X, f)$.

However, it is well known how many chains are needed to decompose a partially ordered set into chains that cover it. By using the ***Dilworth's Theorem***, any finite partially ordered set has a maximal antichain with the same size as the smallest chain decomposition. As a result of the mentioned theorem, the ***width*** of a partially ordered set defined as the length of its largest antichain or the size of the smallest chain decomposition.

As a result of the Dilworth's Theorem, given any closure operator f, we can decompose the underlying lattice to a chain covering with the size of the width of the underlying lattice structure. Then, adding the $\varnothing$ as the smallest element and X as the largest element to each of the chain in the smallest chain decomposition makes each chain a simple classifier. Therefore, the minimum number of simple classifiers needed to generate a classifier is bounded by the width of the underlying lattice structure. Hence, we have our first bound as follows.

Lemma 4. *Let f be a classifier over the set of data points X. The minimum number of simple classifiers needed to generate it is bounded by the width of the lattice $S(X, f)$.*

However, the minimum number of simple classifiers that can generate a given classifier f ***is not*** necessarily the width of the underlying lattice structure. Consider the following example.

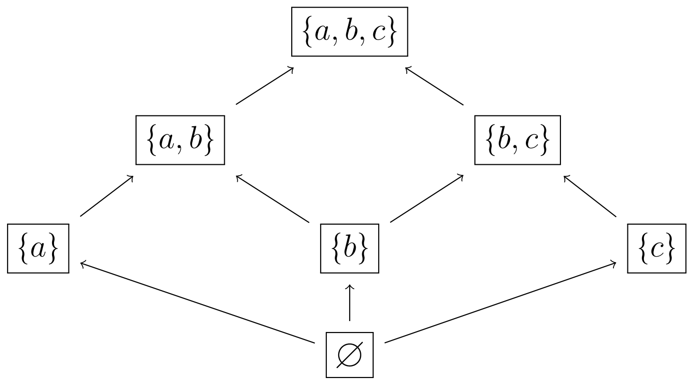

Figure 1.5: The lattice associated with the classifier f.

Example 6. Let $X = \{a, b, c\}$ be a set of data points. We define the classifier f as follows:

$$S(X, f) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, b, c\}\}$$

Figure 1.5 illustrates the lattice associated with f. The antichain with classes $\{a\}, \{b\}, \{c\}$ is the largest antichain in the lattice. Therefore, the width of the lattice $S(X, f)$ is three. As a result of the Dilworth's Theorem, one can decompose the lattice to three chains. For example, the three chains $C_1 = \{\varnothing, \{a\}, \{a, b\}, \{a, b, c\}\}$, $C_2 = \{\{c\}, \{b, c\}\}$, and $C_3 = \{\{b\}\}$. Attaching $\varnothing$ and $\{a, b, c\}$ to each of the three chains implies the following three simple classifiers:

$$\begin{aligned}
S(X, f_{C_1}) &= \{\varnothing, \{a\}, \{a, b\}, \{a, b, c\}\} \\
S(X, f_{C_2}) &= \{\varnothing, \{c\}, \{b, c\}, \{a, b, c\}\} \\
S(X, f_{C_3}) &= \{\varnothing, \{b\}, \{a, b, c\}\}
\end{aligned}$$

However, we can generate f by only two simple classifiers. Let the classifiers f_1, f_2 be as follow:

$$\begin{aligned}
S(X, f_1) &= \{\varnothing, \{a\}, \{a, b\}, \{a, b, c\}\} \\
S(X, f_2) &= \{\varnothing, \{c\}, \{b, c\}, \{a, b, c\}\}
\end{aligned}$$

By using the second condition of Proposition 4, since $\{b\} = \{a, b\} \cap \{b, c\}$ then f_1, f_2 can generate f. Therefore, the minimum number of weak order (MNWO) generating f is only two, which is less than its width which is three.

The above example shows that for some classifiers their MNWO is less than their widths. The reason lies behind the observation that, if any class is the same as the intersection of some other classes detected by some simple classifiers, then the class will be detected without adding it to any simple classifier. Formally, consider the following discussion.

Let f be a classifier and the set of classifiers $\{g_1, \ldots, g_k\}$ generates it. Using our Proposition 4, every class of f which is not the intersection of some other different classes of $g_1, \ldots, g_k$ should appear in at least one of the classifiers $g_1, \ldots, g_k$. However, any other class does not necessarily need to be a class in one of $g_1, \ldots, g_k$. As a result, we have the following procedure.

We remove all the classes that are the intersections of some other classes from the lattice $S(X, f)$. In other words, every class that has an out-degree of more than two, in the lattice $S(X, f)$, will be removed. The remaining set is a partially ordered set, with respect to the set inclusion. By the Dilworth's Theorem, the MNWO should be the width of the remaining partially ordered set. Therefore, we have the following proposition.

Proposition 5. *Let f be a classifier defined over the data set X. Let $S(X, f)$ denote the set of classes. The partial ordered $P(X, f)$ generated by removing the classes in $S(X, f)$ which are the intersections of some other classes. The minimum number of weak orders needed to generate f is equal to the width of $P(X, f)$.*

Example 7. Consider Example 6. If we use the procedure introduced in Proposition 5, then the partially ordered set $P(X, f)$ will be obtained be removing the classes $\varnothing$ and $\{b\}$ from $S(X, f)$. Figure 1.6 illustrates the structure of $P(X, f)$. The width of $P(X, f)$ is two, which is the same as the MNWO of f.

The same observation and technique work to find the minimum number of binary classifiers needed to generate a given closure operator. The following proposition explains the result.

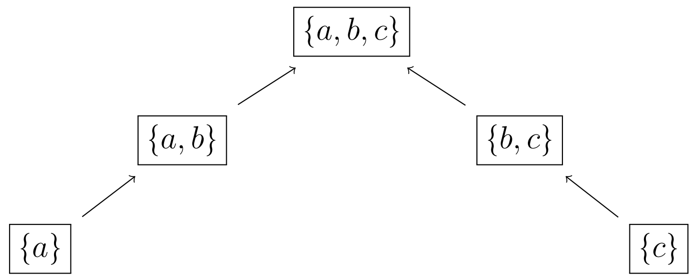

Figure 1.6: The lattice associated with $P(X, f)$.

Proposition 6. *Let f be a classifier defined over the data set X. Let $S(X, f)$ denote the set of classes. The partial ordered $B(X, f)$ generated by removing the classes in $S(X, f)$ which are the intersections of some other classes and also removing the classes $\varnothing$ and X. The minimum number of weak orders with two indifference classes that are needed to generate f is equal to $|B(X, f)|$.*

In the above proposition, since both $\varnothing$ and X are generated through any binary classifier, we can remove them from $S(X, f)$. Every other class in the remaining partially ordered set $B(X, f)$ should be contained in one of the binary classifiers.

Example 8. Consider Example 5. By using Proposition 6, the partially ordered set $B(X, f)$ is equal to $\{\{a, b\}, \{b, c\}\}$. The cardinality of $B(X, f)$ is two, which is the same as MNBC of the classifier f.

In the next section, we discuss some applications of our results in the context of preference ordering over menus. Moreover, we show an additive representation of our results in the same context.

1.5 Preference Ordering Over Menus

1.5.1 Overview

Kreps (1979) introduced the concept of preference ordering over menus to study the dynamic choice behavior of a decision-maker. Up to Kreps' work, in most standard models of decision making, a decision-maker has a preference ordering (weak order), $\succsim$, over a set of alternatives X. As a result, her preference ordering induces another preference ordering, $\succsim'$, over the set of menus, 2^X, as follows:

$$A \succsim' B \text{ iff } \forall b \in B, \exists a \in A \text{ such that } a \succsim b. \tag{1.7}$$

As Kreps mentioned, the preference ordering $\succsim'$ satisfies the following property:

$$A \succsim' B \text{ implies } A \sim' A \cup B \text{ for every menus } A \text{ and } B. \tag{1.8}$$

In other words, the decision-maker does not care to have an option to have a larger menu $A \cup B$ instead of the smaller menu A.

However, in many applications, a decision-maker might have the ***desire to have the flexibility of choices***. For example, consider a decision-maker that has an option first to decide which menu she wants to select for tomorrow's lunch. Then in the next day, she can decide which meal she wants to eat from the selected menu. In this scenario, the decision-maker might like to have the option of the larger menu, since she is not sure about her tomorrow's mood or taste.

As a result, Kreps relaxed the assumption in Equation 1.8. He replaced it by the following two axioms.

1. Desire for flexibility: $B \subseteq A$ implies $A \succsim' B$,

2. Ordinal submodularity: $A \sim' A \cup B$ implies that for all C, $A \cup C \sim' A \cup B \cup C$.

Kreps' main result is as follows:

Theorem 1. *(Kreps, 1979)*

Let X is a finite set of alternatives. A binary relation $\succsim'$ is a weak order satisfies the desire for flexibility and ordinal submodularity if and only if there exist a finite set S and a function $U : X \times S \to \mathbb{R}$ such that

$$v(A) = \sum_{s \in S} [\max_{a \in A} U(a, s)] \tag{1.9}$$

represents $\succsim'$.

The representation captures the idea that the decision-maker has a subjective state space S, and in each state $s \in S$, she has a preference ordering represented by $U(., s)$ over the set of alternatives. Therefore, whenever she compares two menus, she compares the expected values of the menus with respect to her state-dependant utility.

Kreps also has another result to prove his main theorem, which is an interesting result by itself.

Theorem 2. *(Kreps, 1979)*

Let X is a finite set of alternatives. A binary relation $\succsim'$ is a weak order satisfies the desire for flexibility and ordinal submodularity if and only if there exist a finite set S, a function $U : X \times S \to \mathbb{R}$, and a strictly increasing function $u : \mathbb{R}^S \to \mathbb{R}$ such that

$$u([\max_{a \in A} U(a, s)]_{s \in S}) \tag{1.10}$$

represents $\succsim'$.

Following Kreps (1979), the seminal papers by Dekel, Lipman, and Rustichini (2001)(DLR) and Gul et al. (2001) extend the domain of the set of alternatives to the set of simple lotteries, $\Delta(X)$. In their models, the preference ordering is over the set of menus of lotteries. Their setup not only considers the desire for flexibility, but also the ***desire for commitment***. Consider the following example as a motivation behind the desire for commitment.

Again, consider a decision-maker that wants to select a menu for tomorrow's lunch. Assume she knows that if there is a heavy meal in a menu, she will be tempted to select it. As a result, she might select a smaller menu to control herself from her future temptation in the following day.

The paper by DLR is demanding more than Kreps' axioms. As oppose to Kreps, their representation also heavily depends on tools from convex and functional analysis. One of their core axioms is the ***indifference to randomization (IR)***: for every menu $A \subseteq \Delta(X)$, $A \sim' Conv(A)$.

By assuming (IR), any given menu has the same ordering as its convex hull. One of their main results is as follows:

Theorem 3. *(Dekel, Lipman, and Rustichini, 2001) A binary relation $\succsim'$ is a weak order, continuous, non-trivial, and satisfies (IR) if and only if there exist* *a set S and a utility function $U : \Delta(X) \times S \to \mathbb{R}$ and a function $u : \mathbb{R}^S \to \mathbb{R}$ such that*

$$u([\max_{a\in A} U(a,s)]_{s\in S}) \tag{1.11}$$

represents $\succsim'$, continuous, each $U(,s)$ is an expected-utility, and two other conditions which are not crucial for our setup.

In the representation above, they show that adding the Kreps' desire for flexibility axiom makes the aggregator function u strictly increasing.

In other words, in their setup, they replaced the submodularity and desire for flexibility with continuity, non-triviality, IR, and monotonicity (kreps' desire for flexibility) to get a similar representation.

They also show that by strengthening IR to independence axiom, then the aggregator, u, attains an additive form. In other words, there exists a finitely additive measure μ (not necessarily positive) over S such that:

$$u([\sup_{a\in A} U(a,s)]_{s\in S}) = \int_S \sup_{a\in A} U(a,s)\ \mu(ds). \tag{1.12}$$

In the above representation, adding the monotonicity axiom (Kreps' desire for flexibility) makes μ to be a positive measure, which makes the representation similar to the kreps' representation in Equation 1.9.

However, in the more general representation of Equation 1.12, one can associate the set of states which are positive, respects to the measure μ, to the desire for flexibility. Similarly, the negative states are the ones associated with the desire for commitment. Finally, as mentioned by DLR, ***the size of the minimal state space can be a measure of the decision-maker's uncertainty about her future's taste***. Unlike the Kreps' model, they can construct the minimal state space uniquely.

One of our goals is to explain the finite setting of Kreps through a concept similar to the DLR's IR. In the next section, we show that our notion of ***a preference ordering respecting a classifier*** gives us some interesting results, even more general than both DLR and Kreps (in finite setting). Moreover, we have a bound on how many states needed to rationalize a preference ordering over the set of menus. Finally, we prove an additive representation similar to the DLR's version for a finite setting of Kreps. We should mention that, as a corollary of our results, any preference ordering over the set of menus can be rationalized through the additive representation.

1.5.2 A Preference Ordering Respecting a Classifier

In this section, let X be a finite set of alternatives. The preference ordering $\succsim'$ is a weak order over the set of menus of X. We assume that a classifier $f : 2^X \to 2^X$ is revealed through some observations. Moreover, we assume that the preference ordering $\succsim'$ respects f in the following sense.

Definition 3. Let $f : 2^X \to 2^X$ be a classifier and $\succsim'$ be a transitive preference ordering over the set of menus of X. We say that ***the preference ordering*** $\succsim'$ ***respects the classifier*** f , if for every menu $A \in 2^X$, we have $A \sim' f(A)$.

The idea is that we learn some parts of the indifference classes of a preference ordering through some observations. The part we have learned can be represented by the classifier f.

Notice that every preference ordering always respects the identity classifier, which is defined as $I(A) = A$ for every $A \in 2^X$. In other words, if there is no information available, we can only learn the identity classifier. In this case, certainly every preference ordering respects the identity classifier.

Remark 6. In the DLR's setting they put the IR axiom as their core axiom. The closure operator is a generalization of the convex hull operator in our abstract setup. However, we do not put it as an axiom. We assume that there is always one choice of the classifier which is revealed through some observations. If there is no such revelation, we consider the identity classifier as a choice which every preference ordering respects it. Hence, our results are as general as possible.

There are two applications of our definition in this paper. In the next section, we show how to find an additive representation of a preference ordering that has a revealed classifier f, with at most $2(|S(X, f)| - 1)$ number of subjective states.

In the rest of this section, we show the other application of our definition. Consider the two main axioms in the Kreps' setting. Assume that a preference ordering $\succsim'$ satisfies both the desire for flexibility and ordinal submodularity. We define the operator $f : 2^X \to 2^X$ for every $A \in 2^X$ as follows.

$$f(A) = \bigcup_{B \in 2^X,\ A \sim' A \cup B} B. \tag{1.13}$$

It follows easily from the definitions of the two axioms, or lemma 1 in Kreps (1979), that:

1. f is a closure operator,

2. $\succsim'$ respects f,

3. $A \sim' A \cup B$ if and only if $f(B) \subseteq f(A)$,

4. $f(B) \subset f(A)$, then $A \succ' B$.

The whole idea is that, if $A \sim' A \cup B$, then by the desire for flexibility axiom it will be revealed that $A \succsim' B$.

Now, consider our results in section 1.4.2 and consider the following procedure.

Let the set of simple classifiers $G = \{g_1, \ldots, g_k\}$ generates f. Let the set $\{\succsim_1, \ldots, \succsim_k\}$ be the corresponding weak orders associated with the set G. Finally, let a function $U_i : X \to \mathbb{R}$ be a representation of $\succsim_i$ for every $i \in \{1, \ldots, k\}$. Then, using our representations, it is not difficult to see that $f(B) \subset f(A)$ if and only if each coordinates of the vector $(max_{b \in B} U_i(b))_{i \in \{1,\ldots,k\}}$ is less than or equal to $(max_{a \in A} U_i(a))_{i \in \{1,\ldots,k\}}$ and the two vectors are not the same.

As a result, we can define the strictly increasing function $u : \mathbb{R}^k \to \mathbb{R}$ such that $u((max_{a \in A} U_i(a))_{i \in \{1,\ldots,k\}})$ represents $\succsim'$.

It is not hard to see that every representation as in Equation 1.10 can be derived using the procedure mentioned above. Using our Proposition 5, the minimum number of states needed to represent $\succsim'$ is precisely the width of $P(X, f)$, which is the MNWO of f.

We summarize the above discussions in the following proposition.

Proposition 7. *Let a binary relation $\succsim'$ be a weak order and satisfies both the desire for flexibility and ordinal submodularity, then $\succsim'$ can be represented as in Equation 1.10. Moreover, any representation in this form can be generated using the procedure defined above. Finally, the minimum number of states (cardinality of S) needed for the representation is precisely the MNWO of the associated classifiers f defined as in Equation 1.13, which is $|P(X, f)|$.*

1.5.3 Additive Representation

In this section, we wrap up our discussion by presenting our last result on the additive representation of a preference ordering that has a revealed classifier f.

Let X be a set of alternatives. The classifier f has been revealed. For every transitive preference ordering $\succsim'$ that respects f we have the following proposition.

Proposition 8. *If the classifier f has been revealed, then for every transitive preference ordering $\succsim'$ that respects f, there exist a state space S, where $S = S^+ \cup S^-$ with $S^+ \cap S^- = \varnothing$ and the cardinality of at most $2(|S(X, f)| - 1)$, and a state-dependent utility $U : X \times S \to \mathbb{R}$, such that:*

$$U(A) = \sum_{s \in S^+} \underset{a \in A}{max}\, U(a, s) - \sum_{s \in S^-} \underset{a \in A}{max}\, U(a, s) \tag{1.14}$$

represents $\succsim'$.

Proof. We use the Möbius inversion formula to prove the result. Appendix A.1 explains the technique in more detail.

Consider the lattice $S(X, f)$. Define the partial order $\gtrsim$ over $S(X, f)$ by reversing the partial order induced by the set inclusion. In other words, $A \gtrsim B$ if and only if $A \subseteq B$. We can check that the meet and join of the lattice $S(X, f)$ will be swapped.

Since $\succsim'$ is a transitive binary relation, we can make it a weak order over the finite set 2^X. Then there should be a representation by real-valued functions. Consider any utility function $U : 2^X \to \mathbb{R}$ that represents $\succsim'$.

We define the Möbius operator $\Phi : (S(X, f))^{\mathbb{R}} \to (S(X, f))^{\mathbb{R}}$ as follows:

$$\Phi(f)(A) = \sum_{\substack{A \gtrsim B \\ B \in S(X,f)}} U(B). \tag{1.15}$$

Möbius inversion formula guarantees that the Möbius operator is bijective and the inverse is $\Phi^{-1}(g)(A) = \sum_{A \gtrsim B} \mu(B, A) g(B)$, where μ is the Möbius function.

As a result, if we define the function $h : S(X,f) \to \mathbb{R}$, for every $A \in S(X,f)$ as:

$$h(A) = \sum_{\substack{A \succsim B \\ B \in S(X,f)}} \mu(B,A)U(B), \tag{1.16}$$

Then for every $A \in S(X,f)$, U can be retrieved as follows:

$$U(A) = \sum_{\substack{A \succsim B \\ B \in S(X,f)}} h(B). \tag{1.17}$$

However, $A \succsim B$ if and only if $A \subseteq B$. Therefore, we have

$$U(A) = \sum_{\substack{A \subseteq B \\ B \in S(X,f)}} h(B). \tag{1.18}$$

Notice that the above equation is only correct for $A \in S(X,f)$. However, since $\succsim'$ respects f and U represents $\succsim'$, then for every $A \in 2^X$ we have $U(A) = U(f(A))$. Therefore, for every $A \in 2^X$, since $f(A) \in S(X,f)$, we have

$$U(A) = U(f(A)) = \sum_{\substack{f(A) \subseteq B \\ B \in S(X,f)}} h(B). \tag{1.19}$$

Note that, since f is a closure operator, then $A \subseteq B$ if and only if $f(A) \subseteq B$ for every $A \in 2^X$ and every $B \in S(X,f)$. Therefore, for every $A \in 2^X$ we have

$$U(A) = \sum_{\substack{A \subseteq B \\ B \in S(X,f)}} h(B). \tag{1.20}$$

We define $h^+(B) = max(0, h(B))$ and $h^-(B) = max(0, -h(B))$. Since $h = h^+ - h^-$, then we have

$$U(A) = \sum_{\substack{A \subseteq B \\ B \in S(X,f)}} h^+(B) - \sum_{\substack{A \subseteq B \\ B \in S(X,f)}} h^-(B). \tag{1.21}$$

By comparing the above equation and Equation 1.14, we only need to make some changes to make them equal. The trick is as follows.

We define functions $U^+, U^- : (X \times S(X,f) \backslash \varnothing) \to \mathbb{R}$ as follows:

$$U^+(x,B) = \begin{cases} -h^+(B) & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$$
$$U^-(x,B) = \begin{cases} -h^-(B) & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$$

Now, consider any $A \in 2^X$ and $B \in S(X,f)\backslash\varnothing$. By our definition of U^+, U^-, we have

$$\max_{a\in A} U^+(a,B) = \begin{cases} -h^+(B) & \text{if } A \subseteq B \\ 0 & \text{otherwise} \end{cases}$$
$$\max_{a\in A} U^-(a,B) = \begin{cases} -h^-(B) & \text{if } A \subseteq B \\ 0 & \text{otherwise} \end{cases}$$

As a result of the above observation, we get the following result:

$$\begin{aligned} U(A) &= \sum_{\substack{A\subseteq B \\ B\in S(X,f)}} h^+(B) - \sum_{\substack{A\subseteq B \\ B\in S(X,f)}} h^-(B) \\ &= \Big(\sum_{\substack{A\subseteq B \\ B\in S(X,f)\backslash\varnothing}} -\max_{a\in A} U^+(a,B)\Big) - \Big(\sum_{\substack{A\subseteq B \\ B\in S(X,f)\backslash\varnothing}} -\max_{a\in A} U^-(a,B)\Big) \\ &= \Big(\sum_{B\in S(X,f)\backslash\varnothing} -\max_{a\in A} U^+(a,B)\Big) - \Big(\sum_{B\in S(X,f)\backslash\varnothing} -\max_{a\in A} U^-(a,B)\Big) \\ &= -\Big(\sum_{B\in S(X,f)\backslash\varnothing} \max_{a\in A} U^+(a,B)\Big) + \Big(\sum_{B\in S(X,f)\backslash\varnothing} \max_{a\in A} U^-(a,B)\Big) \\ &= \Big(\sum_{B\in S(X,f)\backslash\varnothing} \max_{a\in A} U^-(a,B)\Big) - \Big(\sum_{B\in S(X,f)\backslash\varnothing} \max_{a\in A} U^+(a,B)\Big) \end{aligned} \tag{1.22}$$

Equation 1.22 and 1.14 are similar except their indexes. To make them the same, we consider any two disjoint sets $S^+, S^- \subseteq \mathbb{N}$ [4], with both have $|S(X,f)| - 1$ elements. We consider any two bijection $index_1 : S^+ \to S(X,f)\backslash\varnothing$, $index_2 : S^- \to S(X,f)\backslash\varnothing$. Let $S = S^+ \cup S^-$. We define the function $U : X \times S \to \mathbb{R}$ as follows:

[4]The choice of $\mathbb{N}$ is arbitrary. As long as S^+, S^- are disjoint and each has $|S(X,f)| - 1$ elements, our argument follows.

$$U(x,s) = \begin{cases} U^-(x, index_1(s)) & \text{if } s \in S^+ \\ U^+(x, index_2(s)) & \text{if } s \in S^- \end{cases}$$

Then, using Equation 1.22 and our definition of function U, we have the following result:

$$\begin{aligned} U(A) &= \Big(\sum_{B \in S(X,f)\setminus\varnothing} \max_{a\in A} U^-(a,B)\Big) - \Big(\sum_{B \in S(X,f)\setminus\varnothing} \max_{a\in A} U^+(a,B)\Big) \\ &= \sum_{s\in S^+} \max_{a\in A} U(a,s) - \sum_{s\in S^-} \max_{a\in A} U(a,s) \end{aligned} \tag{1.23}$$

Equation 1.23 finishes our proof. □

Our previous proposition works for any choice of classifiers that respect the preference ordering $\succsim'$. However, the identity operator respects any given preference ordering and $|S(X,I)| = 2^{|X|}$. Therefore, as a corollary to the above proposition, we have the following result.

Corollary 2. *For every preference ordering $\succsim'$ over the set of menus, there exists a representation as in Equation 1.14 with at most $2 \times (2^{|X|} - 1)$ states.*

Remark 7. Notice that the above representation is not a minimal additive representation. For example, let a one-to-one utility function $U : X \to \mathbb{R}$ induces a preference ordering $\succsim'$ over the set of menus. It is not hard to see that the minimal classifier f that respects $\succsim'$ has at least $|X|$ nonempty classes. Thus, our construction in Proposition 8 generates a representation with at least $2(|X| - 1)$ subjective states. However the original construction only needs one state, which can be represented by U itself [5].

1.6 Literature Review

The basic concepts of abstract convex geometry and combinatorial convex hull operator are developed in Edelman et al. (1985). In decision theory, Koshevoy (1999) studies the connection between the combinatorial convex hull operators and the path independent choice functions. The closest paper is by Richter

[5] We are working on another paper which might provide an answer on how to build the minimal additive representation.

et al. (2018). They provide a characterization of a combinatorial convex hull operator through a set of primitive orderings. Using their representation, they propose a notion of competitive equilibrium in an abstract environment.

In the context of dynamic choice behaviors, following Kreps (1979), Dekel, Lipman, and Rustichini (2001)(DLR), and Gul et al. (2001) many papers study different aspects of choice behaviors. However, the main related papers are by Kopylov (2009), Kopylov (2018), and Gorno (2016). Kopylov (2009) determines the number of positive and negative states in DLR's setting. Gorno (2016) shows that any preference ordering in Kreps' setting has a representation as a DLR's representation. Finally, Kopylov (2018) proposes a combinatorial model of subjective states. By relaxing axioms of Kreps' setting, he presents a weaker model of coherent aggregation.

Chapter 2

A THEORY OF RECURSIVE AGGREGATION WITH APPLICATIONS

2.1 Introduction

2.1.1 Overview

This paper presents a general model for studying aggregation in different economic applications. The outcome of our model depends on aggregation of a group of features. We focus on a recursive form of aggregation, where the aggregate outcome for larger collections of features results from aggregating the outcomes of smaller subsets. Specifically, the aggregate outcome of the union of two disjoint collections of features is a weighted average of the outcome of each collection of features separately. We show that this form of recursive aggregation is a common structure that lies behind many seemingly unrelated results in economic theory.

In our model, there exist a set of features and a set of conditional outcomes. We associate each conditional outcome with a subset of features, representing the outcome of the model conditional on aggregation of that subset of features. Our central axiom, the *weighted averaging axiom*, is a simple formalization of the recursivity. It imposes a structure on how the outcome of the union of two disjoint subsets of features relates to the outcome of each of the subsets separately. The axiom states that the outcome of a set of features can be recursively computed by first partitioning the set of features into two disjoint subsets. Then, the aggregated outcome is a weighted average of the outcome of each of the two smaller subsets.

This paper makes two contributions. The first contribution is to find all aggregation procedures that satisfy weighted averaging. The result unifies aggregation procedures across many different economic environments, showing that all of them rely on the same basic result. The second contribution is to show extensions of the result in different domains of economics. Notably, we show applications and extensions in the context of Belief Formation, Choice Theory, and Welfare Economics.

In this paper, we model the dependency of the outcome on the set of aggregated features through what we call an *aggregation rule*. Formally, we define an aggregation rule as a function on the set of subsets of features that maps each subset of features to an outcome of the model.

Our main result finds all aggregation rules that satisfy recursivity in the form of our weighted average axiom. We show that for any two disjoint subsets of features as long as the outcome of their union is a weighted average (with non-negative weights) of the outcome of each subset, then the aggregation rule has a simple form (with a technical richness condition):

There exist a strictly positive weight function and a weak order (a transitive and complete order) over the set of features, with the outcome of any subset of features being the weighted average of the outcomes of each of the highest-ordered features of the subset separately.

The importance of the result is that the weight of each feature is independent of the group of features which the model tries to aggregate. The role of the weak order in the main representation is to partition the set of features to different equivalence classes and rank them from the highest class to the lowest class. If all features of a subset of features are in the same class, then the outcome is the weighted average of the outcomes of each member of the set. However, if some features have a higher ranking than others, then the aggregation rule will ignore lower-ordered features.

Following the main result, we elaborate on two special cases of our main result. In the first case, we propose the *strict weighted averaging* axiom to represent the case where the outcome of the union of two disjoint subsets of features is inside the "relative" interior of the outcomes of each subset separately. The strict weighted averaging axiom captures the idea that the model cannot ignore the role of one of the subsets in finding the outcome of their union.

We show that the strict weighted averaging axiom is the necessary and sufficient condition that the weak order, in the main representation, has only one equivalence class. Hence, the outcome of a subset of features is just the weighted average (with strictly positive weights) of the outcomes of each feature separately.

In the second case, we look into applications where there is a similarity between some features. Specifically, we model the space of features as a subset of a vector space. By considering the distance between vectors, we capture the notion of similarity or closeness of features. In this context, an appealing property is that by replacing a feature in a subset of features with another closely similar feature, the outcome of this new subset stays close to the outcome of the previous one. Under this property, which we define as the *continuity* axiom, all similar enough features attain the same ranking with respects to the weak order. Moreover, the weight function is a continuous function over the set of features. In other words, the weight between two close (or similar) features should be close. In a special case, where the space of features is a convex set, we show that all features attain the same ranking. In this case, there is no difference between the weighted averaging and the strict weighted averaging axiom.

Depending on the application, each feature and the aggregation rule may have different interpretations. Each feature may represent a signal or an event containing some information about the true state of nature. In this case, the role of an aggregation rule is to form a belief about the true state of nature. In the context of choice theory, features may represent choice objects, where an aggregation rule behaves as a decision maker that selects a lottery or a random choice out of a group of choice objects. Another interpretation is in the context of welfare economics, where each feature represents a preference of an individual over some alternatives. In this case, an aggregation rule represents a social welfare function that associates with each preference profile, a single preference ordering over the set of alternatives.

The most immediate interpretation of the result is in the case of modeling an agent who makes a prediction about the true state of nature, conditional on observing a set of events. In this context, each feature represents an event, and the outcome of the model conditional on observing a set of events is the belief about the true states of nature. The main result is the necessary and sufficient condition that the belief formation process behaves as a *Bayesian Updater*. Under the main axiom, there exists a conditional probability system associated with the set of events, and the belief formation process conditional on observing a set of events behaves as a conditional probability. The weak order of the main result is capturing the idea that, conditional on observing even a zero probability event, the belief formation still behaves as a Bayesian updater.

To motivate the framework and results, sections 2.5, 2.6, 3.2, and 4.2 show applications and extensions of our main representations. We show that the weighted averaging axiom is closely related to many known axioms in different topics, from the Pareto axiom in Social Choice Theory to the path independence axiom in Choice Theory.

2.1.2 Applications

2.1.2.1 Belief Formation

In Section 2.5, we interpret the set of features as signals. Each signal contains some information about the distribution of states of nature. The role of an aggregation rule is an agent who makes a prediction about the true state of nature, based on observing some signals. In this context, the range of an aggregation rule is the probability distributions over the states of nature. Following Billot, Gilboa, Samet, and Schmeidler (2005), an aggregation rule is a *belief formation process* that associates with each finite set of signals, a *belief* over the states of nature.

The representation of the belief formation process under the weighted averaging axiom is a straightforward application of the main results. Using our representation, on the one hand, we propose an extension, where the timing of signals may be important. We consider the case where an agent can receive signals in different time zones in the past. The agent tries to form a prediction at the present time, and it may perceive signals closer to the time of the prediction as more credible. To capture the representation, we introduce the *stationarity* axiom, in which a belief induced by a set of received signals and their timing is the same as the belief induced by shifting the timings of all signals by a constant number to the past.

Under stationarity, any belief formation process satisfying the strict weighted averaging axiom has a weight function over the set of signals and an exponential discount factor over each time zone. The belief associated with a set of received signals is the discounted weighted average of the beliefs associated with each signal. In this case, the weight function captures the time-independent value of each signal.

On the other hand, we interpret the set of signals as the information structure of an agent who wants to predict the true state. We interpret each subset of signals as an event in her information structure. We show that as long as the information structure has a finite cardinality, the strict weighted averaging axiom is the necessary and sufficient condition for a rich belief formation process to appear as a Bayesian updater. This result answers the question in Shmaya and Yariv (2007) regarding finding a necessary and sufficient condition for a belief formation process to act as a *Bayesian updating rule*.

2.1.2.2 Average Choice Functions

In Section 2.6, the set of features is a subset of $\mathbb{R}^n$. We interpret each feature as a choice object. The interpretation of the aggregation rule is a decision maker that selects a choice randomly from a menu of choice objects. We model the decision maker as an *average choice function* that associates with any menu of choice objects, an average choice (mean of the distribution of choices) in the convex combination of choice objects. Average choice is easier to report and obtain rather than the entire distribution[1]. However, except for the case where elements of a menu are affinely independent, average choice does not uniquely reveal the underlying distribution of choices.

First, using our main representation, we show that it is possible to uniquely extract the underlying distribution of choices as long as the average choice function satisfies the weighted averaging axiom.

Then, we illustrate two applications of the result. In one application, we consider the class of average choice functions that can be rationalized by a *Luce rule*, *i.e.*, a stochastic choice function that satisfies the *independence of irrelevant alternatives* axiom (IIA) proposed by Luce (1959). We show that the average choice functions satisfying the strict weighted averaging axiom are exactly the ones that can be rationalized by a Luce rule. More generally, we show that the class of average choice functions satisfying the weighted averaging axiom is the same as the class of average choice functions rationalizable by a *two-stage Luce* model proposed by Echenique and Saito (2018).

In the second application, we consider continuous average choice functions. First, we show that any continuous average choice function under the weighted averaging axiom is rationalizable by a Luce rule. This means that there is no continuous average choice function that is rationalizable by a two-stage Luce rule but not with a Luce rule.

[1]Check Ahn, Echenique, and Saito (2018) for the complete discussion on merits of average choice.

Then, we illustrate a connection of our result with the one by Kalai and Megiddo (1980), regarding the impossibility of an average choice function to satisfy both the *path independence* axiom and continuity.

2.2 Primitives and Axioms

Let X be a nonempty set of features. Observe that we make no assumptions about the cardinality or topology of X. We denote the set of all real numbers by $\mathbb{R}$, and the set of all positive real numbers by $\mathbb{R}_{++}$. We represent n dimensional real vectors by $\mathbb{R}^n$. For any $A \subseteq \mathbb{R}^n$, we denote by $Conv(A)$ the set of all convex combinations of vectors in A. Respectively, we denote the relative interior and boundary of $\text{Conv}(A)$ by $Conv^{\circ}(A)$ and $\partial^r(\text{Conv}(A))$.

Definition 4. A set $A \subseteq X$ is a ***coalition*** of features of X if it is a nonempty finite subset of X. We denote the set of all coalitions of X by X^*.

Definition 5. A nonempty set $A \subseteq X$ is a ***sub-coalition*** of a set $B \subseteq X$ if $A \subseteq B$.

Definition 6. A binary relation $\succcurlyeq$ on X is a ***weak order*** on X, if it is reflexive, transitive, and complete.

We denote the symmetric and asymmetric parts of a weak order $\succcurlyeq$ by $\sim$ and $\succ$. Given a weak order $\succcurlyeq$ on X for any nonempty coalition $A \in X^*$, we define the set $\boldsymbol{M(A,} \succcurlyeq)$ as:

$$M(A, \succcurlyeq) := \{x \in A | x \succcurlyeq y \ \forall y \in A\} \tag{2.1}$$

The weak order $\succcurlyeq$ captures a member ranking of the set X. Given any coalition A, $M(A, \succcurlyeq)$ contains the highest-ranked features of A. Since A has a finite number of features, $M(A, \succcurlyeq)$ is a nonempty subset of A.

Definition 7. An ***aggregation rule*** on X is a function $f : X^* \to \mathbb{R}^n$, that associates with every coalition $A \in X^*$ a vector $f(A) \in \mathbb{R}^n$.[2]

An aggregation rule captures how outcomes depend on each coalition of features. In this paper, we focus on recursive aggregation rules. The idea is as follows: the value of the aggregation rule over a coalition of features depends

[2]All discussions of this paper continue to hold if $\mathbb{R}^n$ is replaced by any general (possibly infinite dimensional) normed vector space.

on the aggregated values of any two disjoint sub-coalitions. Specifically, the aggregated outcome is a weighted average (with non-negative weights) of the value returned by the aggregation rule for each of the two sub-coalitions.

The idea of recursive aggregation is embodied in the following axiom, which is central to our analysis:

Axiom 1. ***(Weighted averaging)*** If $A, B \in X^*$ and $A \cap B = \varnothing$, then

$$f(A \cup B) \in \mathrm{C}onv(f(A), f(B))$$

or equivalently:

$$\exists\ \lambda \in [0,1] \text{ s.}t.\ f(A \cup B) = \lambda f(A) + (1-\lambda) f(B).$$

In the axiom, there is no restriction on how the weights are connected to each sub-coalition. The weight λ may depend on A, B, and $A \cup B$. However, the main result implies that there is a simple connection between the weights (which are, furthermore, unique). But first, we will go over some special cases of the weighted averaging axiom.

The weight λ in the axiom can be any number between zero and one. There are two special cases: strict and extreme. In the case of the strict version, λ cannot be zero or one. Therefore, the value of an aggregation rule over any two disjoint coalitions must be strictly between the values of each separate coalition. In the extreme case, λ is either zero or one. In this case, the value of an aggregation rule of the union of two disjoint coalitions becomes *equal* to the value over one of the coalitions.

To capture the difference between the two special cases, we introduce the following axioms:

Axiom 2. ***(Strict weighted averaging)*** If $A, B \in X^*$ and $A \cap B = \varnothing$, then

$$f(A \cup B) \in \mathrm{C}onv^{\circ}(f(A), f(B))$$

or equivalently:

$$\exists\ \lambda \in (0,1) \text{ s.}t.\ f(A \cup B) = \lambda f(A) + (1-\lambda) f(B).$$

Axiom 3. ***(Extreme case)*** If $A, B \in X^*$ and $A \cap B = \varnothing$, then

$$f(A \cup B) \in \{f(A), f(B)\}$$

or equivalently:

$$\exists\ \lambda \in \{0,1\} \text{ s.}t.\ f(A \cup B) = \lambda f(A) + (1-\lambda) f(B).$$

We provide two examples of rules satisfying the weighted averaging axiom. Our examples clarify the strict and extreme cases. The first example is the standard notion of rational choice, and illustrates the extreme version of the axiom.

Example 9. Consider a complete strict order $\succ$ on the set of features X. Given any feature $x \in X$, let $f(x) \in \mathbb{R}^n$ be the outcome of the model based on observing the single feature x. Given any coalition $A \in X^*$, there is a single highest-ordered element $M(A, \succ)$ in the coalition. We define an aggregation rule $f : X^* \to X$ over any coalition $A \in X^*$ to be the outcome of the highest-ordered elements of the coalition A. Formally, we have:

$$f(A) = f(M(A, \succ)).$$

There are two observations. First, the aggregation rule satisfies the weighted averaging. Second, it is an example of the extreme case. Due to the completeness of the order $\succ$, for any two disjoint coalitions $A, B \in X^*$, we have:

$$f(A \cup B) = \begin{cases} f(A) & \text{if } M(A, \succ) \succ M(B, \succ), \\ f(B) & \text{otherwise.} \end{cases} \tag{2.2}$$

The second example captures the case of strict weighted averaging. The example has the same flavour as the standard Luce model of individual choice behavior (section 2.6 covers average choice functions in more detail).

Example 10. Consider a weight function $w : X \to \mathbb{R}_{++}$ on a set of features. Given any feature $x \in X$, let $f(x) \in \mathbb{R}^n$ be the outcome of the model based on observing the single feature x. We define an aggregation rule $f : X^* \to \mathbb{R}^n$ over any coalition $A \in X^*$ as follows:

$$f(A) = \sum_{x \in A} \left(\frac{w(x)}{\sum\limits_{y \in A} w(y)} f(x) \right).$$

This form of aggregation also satisfies the weighted averaging axiom. The reason is that given any two disjoint coalitions $A, B \in X^*$, we have the following property:

$$f(A \cup B) = \left(\frac{\sum\limits_{x \in A} w(x)}{\sum\limits_{y \in A \cup B} w(y)} \right) f(A) + \left(\frac{\sum\limits_{x \in B} w(x)}{\sum\limits_{y \in A \cup B} w(y)} \right) f(B).$$

In this case, since the weights are strictly positive, the aggregation rule satisfies the strict weighted averaging.

These examples are two separate and simple forms of aggregation rules. Both satisfy the weighted averaging axiom. The main result of the next section shows that the combination of these two special cases is the whole class of aggregation rules under the weighted averaging axiom.

We should mention that both the *weighted averaging* axiom and *strict weighted averaging* axiom are closely related to the *concatenation axiom* of Billot et al. (2005) in the context of case-based prediction, *path independence or Plott's condition* of Plott (1973), *partial path independence* of Ahn et al. (2018) in the context of stochastic choice functions, *extended Pareto axiom* of Dhillon (1998) and Baucells and Shapley (2008), and also the *coherence axiom* of Skiadas (1997a) in the context of preference aggregation rules. Sections 2.5, 2.6, 3.2, and 4.2 describe the connections of these axioms.

But first, in Section 2.3, we provide the representation of aggregation rules under the weighted averaging axiom and special cases. Then, in Section 2.4, by providing a norm over the set of features, X, we define continuous aggregation rules, and get our last representation for the case of continuous aggregation rules.

2.3 Main Representation

This section presents our main representation. Examples 9 and 10 in Section 2.2, show two simple forms of aggregation satisfying the weighted averaging axiom. Our main result, under a technical richness condition, shows that the class of aggregation rules under the weighted averaging axiom consists of the combination of these two simple forms of aggregation rules.

Before stating our main result, we present the special case of the representation under the strict weighted averaging axiom. We show that under a technical richness condition, all aggregation rules satisfying the strict version have the simple form of Example 10.

In this case, the richness condition determines that the range of the aggregation rule should be more than a single line. Formally, we have:

Definition 8. An aggregation rule $f : X^* \to \mathbb{R}^n$ is ***rich*** if the range of f is not a subset of a line.

The above condition means that we can always find three disjoint features such that under the aggregation rule they form a triangle in $\mathbb{R}^n$.

The special case of our main result is as follows:

Theorem 4. *Let an aggregation rule $f : X^* \to \mathbb{R}^n$ satisfy the richness condition. The following are equivalent:*

1. *The aggregation rule f satisfies the strict weighted averaging axiom.*
2. *There exists a weight function $w : X \to R_{++}$ such that for every $A \in X^*$:*

$$f(A) = \frac{\sum\limits_{x \in A} w(x) f(x)}{\sum\limits_{x \in A} w(x)}. \tag{2.3}$$

Moreover, the function w is unique up to multiplication by a positive number.

The strict weighted averaging axiom states that for any $A, B \in X^*$, there exists a weight $\lambda_{A,B,A\cup B} \in (0,1)$, possibly depending on A, B, and $A \cup B$, such that $f(A \cup B) = \lambda_{A,B,A\cup B} f(A) + (1 - \lambda_{A,B,A\cup B}) f(B)$. However, this result shows that there should be a strictly positive weight function w over the set of features and $\lambda_{A,B,A\cup B} = \frac{\sum_{x \in A} w(x)}{\sum_{y \in A \cup B} w(y)}$. The complete proof of the result is in the appendix.

The previous representation is the building block of our main representation. However, to get the representation, we need a stronger richness condition.

Definition 9. An aggregation rule $f : X^* \to \mathbb{R}^n$ is ***strongly rich*** if for any $x \in X$ there exist two other different features $y, z \in X$ such that:

1. $f(\{x,y\}) \notin \{f(x), f(y)\}$ and $f(\{x,z\}) \notin \{f(x), f(z)\}$[3].

2. $f(\{x\}), f(\{y\})$, and $f(\{z\})$ are not on the same line.

The condition states that for any feature, we can find another two features that (1) their aggregated coalitions satisfy the strict weighted averaging axiom, (2) they are non-collinear in the range of f.

The idea behind the strong richness condition and the richness condition follows from the main representation. We provide the discussion after the main result.

Our main result shows that we can identify the class of aggregation rules under the weighted averaging axiom as long as they satisfy the strong richness condition. The representation is a combination of the rules in Example 9 and 10. The main representation is as follows:

Theorem 5. *Let an aggregation rule $f : X^* \to \mathbb{R}^n$ be strongly rich. The following are equivalent:*

1. *The aggregation rule f satisfies the weighted averaging axiom.*

2. *There exist a unique weak order $\succcurlyeq$ on X and a weight function $w : X \to R_{++}$ such that for every $A \in X^*$:*

$$f(A) = \sum_{x \in M(A,\succcurlyeq)} \left(\frac{w(x)}{\sum\limits_{y \in M(A,\succcurlyeq)} w(y)} \right) f(x). \tag{2.4}$$

Moreover in this case, the function w is unique up to multiplication by a positive number in each of the equivalence classes of the weak order $\succcurlyeq$.

There are two forces behind the representation: one is captured by the weak order $\succcurlyeq$; the other is the weight function w.

The weak order plays the role it has in example 9. It partitions the set of features into equivalence classes and ranks them from top to bottom. If all features of a coalition have the same ranking, then the outcome is the weighted

[3] In the proof of our main result, we show that as long as $f(\{x\}), f(\{y\})$, and $f(\{z\})$ are not on the same line, then $f(\{x,y\}) \notin \{f(x), f(y)\}$ and $f(\{x,z\}) \notin \{f(x), f(z)\} \Rightarrow f(\{y,z\}) \notin \{f(y), f(z)\}$.

average of the outcomes of each member of the coalition. However, if some features have a higher ranking than others, then the aggregation rule will ignore the lower-ordered features.

Hence, the assessment of the aggregation rule has two steps. First, it only considers the highest-ordered elements. Then, it uses the weight function and finds the weighted average among the highest-ordered features.

The importance of the representation is that the weighted averaging axiom states that for any $A, B \in X^*$, there exists a weight $\lambda_{A,B,A\cup B} \in [0,1]$, possibly depending on A, B, and $A \cup B$, such that $f(A \cup B) = \lambda_{A,B,A\cup B} f(A) + (1 - \lambda_{A,B,A\cup B}) f(B)$. However, Theorem 5 shows that there should be a weak order $\succcurlyeq$ and a strictly positive weight function w over the set of features and for any $A, B \in X^*$, we have:

$$\lambda_{A,B,A\cup B} = \frac{\sum\limits_{x \in M(A,\succcurlyeq) \cap M(A\cup B,\succcurlyeq)} w(x)}{\sum\limits_{y \in M(A\cup B,\succcurlyeq)} w(y)}.$$

The idea behind the proof is as follows. First, we define an order $\succcurlyeq$ over binary coalitions. Given $x, y \in X$, if $f(x) \neq f(y)$ we define $x \succcurlyeq y$ if $f(\{x,y\}) \neq f(y)$. Using the strong richness condition, we prove that this order is transitive. Moreover, we can make it into a complete order. In the next step we show that for any coalition A, $f(A) = f(M(A, \succcurlyeq))$. This part of the proof is the most difficult part. Finally, the strong richness condition states that in each equivalence class of the weak order $\succcurlyeq$, there should be three non-collinear points. Hence, we use the result of Theorem 4 and in each equivalence class. In other words, we can find the weight function that represents the rule in each equivalence class. The conditions that (1) $f(A) = f(M(A, \succcurlyeq))$ and (2) all points in $M(A, \succcurlyeq)$ are in a same equivalence class, will complete the proof.

In both of the representations above, the richness condition is crucial. The following example shows that without the richness, there are aggregation rules that satisfy the strict weighted averaging axiom, but do not have a weighted average representation.

Example 11. Let $X = \{x, y, z\}$ with $f(\{x\}) = 0, f(\{y\}) = 1/2, f(\{z\}) = 1, f(\{x,y\}) = 1/4, f(\{y,z\}) = 3/4, f(\{x,z\}) = 3/8$, and $f(\{x,y,z\}) = 7/16$.

Assume that there exists a positive weight function on X, and the aggregation rule over any coalition of X has a representation as a weighted average of its elements.

Assume that $w : X \to \mathbb{R}_{++}$ is the corresponding weight function. In order to have such a representation, we should have $f(\{x, y\}) = \frac{w(x)f(x)+w(y)f(y)}{w(x)+w(y)}$. By considering the value of $f(\{x, y\}), f(\{x\})$, and $f(\{y\})$, we get $\frac{w(x)}{w(y)} = 1$. Similarly, by considering the coalition $\{y, z\}$ we get $\frac{w(y)}{w(z)} = 1$. By combining these two observations, we get $\frac{w(x)}{w(z)} = 1$. However, considering the coalition $\{x, z\}$, and the representation $f(\{x, z\}) = \frac{w(x)f(x)+w(z)f(z)}{w(x)+w(z)}$, we get $\frac{w(x)}{w(z)} = 5/3$, which is a contradiction. Hence, the representation does not work in this case.

The problem with the example is that the range of the aggregation rule is a subset of a single dimensional vector space. Therefore, it does not satisfy the richness condition.

The example also shows the reason behind the strong richness condition in Theorem 5. To be able to define a weight function in each equivalence class of the weak order $\succcurlyeq$, the aggregation rule should satisfy the richness in each equivalence class. This is precisely the idea behind the definition of the strong richness condition: the first condition states that the three points are in a same equivalence class. The second condition states that they are non-collinear.

Remark 8. In the case of aggregation rules under the strict weighted averaging axiom, the richness condition is equivalent to the strong richness condition. Since, if an aggregation $f : X^* \to \mathbb{R}^n$ satisfies the richness condition, there exist three non-collinear vectors $f(\{x\})$, $f(\{y\})$, and $f(\{z\})$ in the range on f. Under the strict weighted averaging axiom, the vectors $f(\{x, y\}) \in \text{Conv}^\circ(\{f(\{x\}), f(\{y\})\})$ and $f(\{x, z\}) \in \text{Conv}^\circ(\{f(\{x\}), f(\{z\})\})$. Therefore, f satisfies the strong weighted averaging axiom. This means that Theorem 4 is a corollary of Theorem 5.

Remark 9. Note that, by the richness condition, the dimension of the range of the aggregation rule should be at least 2. As a result, the dimension of $\mathbb{R}^n$ in the definition of aggregation rules should be strictly more than 1.

Remark 10. As a consequence of Theorem 4, for any two members of the space X, the ratio of their weights in any coalition is constant. This observation guarantees that by only considering the coalition of any two features, we can get the ratio of their weights in any coalition containing them. Instead of the strict weighted averaging axiom, by fixing the ratio of weights of two features in any coalition including them, we can get a similar result as Theorem 4. This observation is closely related to the *independent of irrelevant of alternative* axiom of Luce (1959), which is discussed in detail in Section 2.6.2.

The next section shows that for continuous aggregation rules over a convex domain, the weighted averaging axiom is the same as the strict weighted averaging axiom. Therefore, the weak order $\succcurlyeq$ in the main representation has only one equivalence class.

2.4 Continuous Aggregation Rules

In this section, in order to discuss the closeness of members of coalitions, we assume that X is a subset of a normed vector space. Two features of X are close to each other if their distance, with respect to the norm on X, is close to each other.

One appealing property of an aggregation rule would be that by replacing a member of a coalition with another feature close to that member, the value of the aggregation rule over this new coalition stays close to the value of the aggregation rule over the previous coalition. To capture this idea, we define a continuous aggregation rule as follows:

Axiom 4. ***(Continuity)*** An aggregation rule $f : X^* \to \mathbb{R}^n$ is continuous if, for any coalition $A \in X^* \cup \{\varnothing\}$, and any feature $x \in X \backslash A$, if a sequence $(x_n)_{n=1}^{\infty} \in X$ and $x_n \to x$, then:

$$f(A \cup \{x_n\}) \to f(A \cup \{x\}).$$

Remark 11. In the definition of the previous axiom, convergence in X is with respect to the norm on X, and the convergence in the range of the aggregation rule is with respect to the Euclidean norm of $\mathbb{R}^n$.

By considering a continuous aggregation rule $f : X^* \to \mathbb{R}^n$ and a single-element coalition $\{x\} \in X^*$, if $x_n \in X$ and $x_n \to x$, then by the definition of the continuity, we should have: $f(\{x_n\}) \to f(\{x\})$. In other words, the value of an aggregation rule over a feature close to a particular feature is close to the value of the aggregation rule over that particular feature.

By adding the weighted averaging axiom, we can get the representation in Theorem 5. The next theorem shows that for continuous aggregation rules under the weighted averaging axiom, the weight function in the representation is also a continuous function. In other words, not only the values of the aggregation rule of two close features are close, but also their weights are close. Moreover, if two features are close enough, they are in the same equivalence class of the weak order $\succcurlyeq$. In other words, they have the same ranking. Formally, we have:

Theorem 6. *Let X be a subset of a normed vector space and an aggregation rule $f : X^* \to \mathbb{R}^n$ be strongly rich and satisfy weighted averaging axiom. By Theorem 5 there exist a unique weak order $\succcurlyeq$ on X and a weight function $w : X \to R_{++}$ such that for every $A \in X^*$:*

$$f(A) = \frac{\sum\limits_{x \in M(A,\succcurlyeq)} w(x) f(x)}{\sum\limits_{x \in M(A,\succcurlyeq)} w(x)}.$$

Moreover, if f is continuous, then:

1. *The weight function w is continuous, i.e., for any $x \in X$ and a sequence $(x_n)_{n=1}^{\infty} \in X$ such that $x_n \to x$, then $w(x_n) \to w(x)$.*
2. *For any $x \in X$, there exists a neighborhood $\mathcal{N}_x$ of x such that for all $y \in \mathcal{N}_x \cap X$: $y \sim x$. In other words, $\forall\ x \in X\ \exists\ \epsilon > 0$ such that $\forall\ y \in B_\epsilon(x) :\ y \sim x$, where $B_\epsilon(x) = \{z | z \in X, |z - x| < \epsilon\}$.*

Next, we assume that X is a convex subset of a normed vector space. Our next theorem shows that under this assumption, any continuous aggregation rule on X under weighted averaging axiom can only have a single equivalence class. To clarify, for any $x \in X$, not only any feature close enough to x has the same order as x, but also all other members of X has the same order as x.

The theorem states that under a convex domain, for continuous aggregation rules, there is no difference between the weighted averaging axiom and strict weighted averaging axiom.

Theorem 7. *Let X be a convex subset of a normed vector space, and $f : X^* \to \mathbb{R}^n$ a rich aggregation rule that satisfies weighted averaging and continuity axioms. Then, there exists a continuous weight function $w : X \to R_{++}$ such that for every $A \in X^*$:*

$$f(A) = \frac{\sum\limits_{x \in A} w(x) f(x)}{\sum\limits_{x \in A} w(x)}.$$

The proof is simple. Therefore, we include it here.

Proof. We prove Theorem 7 by contradiction. Assume that there exist $x, y \in X$ such that $x \nsim y$. Without loss of generality, we assume that $f(x) \neq f(y)$. Since if $f(x) = f(y)$, by the strong richness condition there exists another point $z \in X$ in which $z \sim x$ and $f(z) \neq f(x)$, and we can use z instead of x.

Hence, assume that $x \nsim y$ and $f(x) \neq f(y)$. By convexity of X, the whole segment $[x, y] = \{\alpha x + (1-\alpha)y \mid \alpha \in [0,1]\}$ should be in X. Define a point $\alpha_1 \in [0,1]$ as $\inf\{\alpha \mid \alpha \in [0,1],\ \alpha x + (1-\alpha)y \nsim x\}$. The "inf" is well defined, since it is defined on a bounded subset of $[0,1]$ and additionally $x \nsim y,\ f(x) \neq f(y)$.

Consider the point $z = \alpha_1 x + (1-\alpha_1)y$. First, we prove that the point z is different than x. By the result of Theorem 6, the point x has a neighborhood $B_\epsilon(x)$ with all points in this neighborhood having the same order as x. By considering $B_\epsilon(x) \cap [x, y]$, z is different than x. Similarly, z is different than y.

By Theorem 6, there exists a neighborhood $B_{\epsilon'}(z)$ of z such that all points in the neighborhood have the same order as z. By considering the set $B_{\epsilon'}(z) \cap [x, y]$ and the fact that $z \neq x$ and $z \neq y$, α_1 cannot be $\inf\{\alpha \mid \alpha \in [0,1],\ \alpha x + (1-\alpha)y \nsim x\}$, which is a contradiction.

The contradiction proves that all points in X have the same order with respect to the weak order $\succcurlyeq$, which completes the proof.

□

We can interpret the representation as an impossibility result. It means that it is not possible to have the extreme case and continuity. This interpretation is especially useful in the context of choice theory. We show that this is a more general version of the impossibility result by Kalai and Megiddo (1980). Section 2.6.3 covers the interpretation.

2.5 Belief Formation

In this section, we discuss an application of our main results in the context of Belief Formation. The role of an aggregation rule is an agent who makes a prediction about the true state of nature, based on observing some signals containing information about the true state. In this context, the range of an aggregation rule is the probability distributions over the states of nature.

Section 2.5.1 provides the main definitions and application of our main result in this context. In Section 2.5.2, we consider an extension where the timing of signals changes their credibility. Finally, in Section 2.5.3, by interpreting the set of signals as the agent's information structure, we find the necessary and sufficient condition that a belief formation is a Bayes updater.

2.5.1 Belief Formation Processes

Let $\Omega = \{1, 2, \ldots, n\}$ be a set of states of nature. Depending on the application, there might be different interpretations of the elements of X as disjoint cases, events, or experts such that each has some information about the true state. To be consistent, we interpret the elements of the set X as disjoint signals.

The role of an aggregation rule over a finite subset of signals is to predict the true state of nature by assigning probabilities to each state. Therefore, following Billot et al. (2005), the aggregation rules can be interpreted as a *belief formation process*, in which by observing a finite subset of signals it assigns a *belief* to the set of states of nature. More precisely, let $\Delta(\Omega)$ be the set of all probability distributions over the set of states of nature.

Definition 10. A ***belief formation process*** is a function $f : X^* \to \Delta(\Omega)$, that associates with every finite set of signals $A \in X^*$, a ***belief*** $f(A) \in \Delta(\Omega)$ on the states of nature.

Theorem 5 shows that if the belief induced by the union of two disjoint finite sets of signals is on the line segment connecting the beliefs induced by each set of signals separately, then, under the strong richness condition, there exists a strictly positive weight function and a weak order over the set of signals such that the belief over any finite subset of signals is a weighted average of the beliefs induced by each of the highest-ordered signals of that subset.

By enforcing the belief formation process to use both of the induced beliefs, *i.e.*, the belief induced by the union of two disjoint finite sets of signals is on the "interior" of the line segment connecting the induced belief of each set of signals separately, we can use Theorem 4 to find the representation. Formally, we have:

Corollary 3. *Let a belief formation process $f : X^* \to \Delta(\Omega)$ be strongly rich and satisfies weighted averaging axiom. Then, there exist a unique weak order $\succcurlyeq$ on X and a weight function $w : X \to R_{++}$ such that for every $A \in X^*$:*

$$f(A) = \sum_{x \in M(A,\succcurlyeq)} \left(\frac{w(x)}{\sum\limits_{x \in M(A,\succcurlyeq)} w(x)} \right) f(x).$$

Moreover, if the function f satisfies the strict weighted averaging axiom, then the weak order $\succcurlyeq$ has only one equivalence class and for every $A \in X^*$:

$$f(A) = \sum_{x \in A} \left(\frac{w(x)}{\sum\limits_{x \in A} w(x)} \right) f(x).$$

The special case of the representation above, under the strict weighted averaging axiom, is similar to the one in Billot et al. (2005). However, their belief formation process is defined over "sequences" of signals, in which each sequence can have multiple copies of the same signal. In contrast, we define the belief formation process over "sets" of signals, and there can be only one copy of a signal in each set. Billot et al.'s main axiom, concatenation axiom, is defined over any two sequence of signals, and the role is to count the number of each signals in each sequence. However, our strict weighted averaging axiom does not allow the same signal to be in sub-coalitions, which is crucial for our representation (otherwise, since we have the "union", the representation does not hold).

As a result, besides the conceptual differences, there is a mathematical difference between the concatenation axiom and our strict weighted averaging axiom.

2.5.2 Role of Timing

In this section, we explore the role of the timing of signals. We assume that an agent may receive signals in different time zones in the past. Among a set of received signals, the agent may perceive a signal closer to the time of the prediction as more credible compared to the same signal if it was received further in the past. Therefore, the agent may add more weight to the belief induced by that specific signal if it was received closer to the time of the prediction.

To capture the idea, in the rest of this section, we assume that a belief formation process receives signals in possibly different time zones in the past and it tries to form a belief at the present time.

We show that if a belief induced by a set of received signals and their timing is the same as the belief induced by a constant shift of timings of the same received signals, then the belief formation process, under the strict weighted averaging axiom, has a simple form:

There exist a weight function over each signal and an exponential discount factor over the timing. The belief induced by a set of received signals is just the discounted weighted average of the beliefs induced by each of the received signals separately.

Formally, let X be the set of signals. The present time denoted by 0, and time $t \in \mathbb{N}$ represents t units of time before the present time. For a given finite subset of signals $A \in X^*$, let a function $T_A : A \to \mathbb{N}$, represent the timing of each signal in the set A, *i.e.*, for any signal $x \in A$, $T_A(x)$ is the time of receiving the signal x. Given a $c \in \mathbb{N}$, $T_A + c$ represents a *time shift* of size c over the timing T_A of a set of received signals A. Finally, the set $X^T = \{(A, T_A) \mid A \in X^*,\ T_A : A \to \mathbb{N}\}$ represents all possible realizations of the received signals. In this context, a belief formation process is a function $f : X^T \to \Delta(\Omega)$.

Our main axiom, in addition to the strict weighted averaging axiom, is the *stationarity* axiom. A belief formation process is stationary if a belief induced by a set of received signals and their timing is the same as the belief induced by a constant shift of timings of the same received signals. More precisely:

Axiom 5. ***(Stationary)*** If $A \in X^*$, $T_A : A \to \mathbb{N}$, then for any $c \in \mathbb{N}$:

$$f((A, T_A + c)) = f(A, T_A).$$

The stationarity axiom specifies that shift of the timing of signals, with the number for all the signals, does not change the belief.

The next proposition characterizes the belief formation processes under both the stationarity and strict weighted averaging axioms.

Proposition 9. *Let a rich belief formation process $f : X^T \to \Delta(\Omega)$ satisfy the strict weighted averaging and stationarity axioms. Then, there exist a unique* *discount factor $q \in (0, \infty)$ and a unique (up to multiplication by a positive number) weight function $w : X \to \mathbb{R}_{++}$, such that for all $(A, T_A) \in X^T$:*

$$f(A, T_A) = \frac{\sum\limits_{x \in A} q^{T_A(x)} w(x) f(x)}{\sum\limits_{x \in A} q^{T_A(x)} w(x)}. \tag{2.5}$$

The idea behind the proof is to use the uniqueness of the weight w in Theorem 4. Next, by using the stationarity axiom, we show that there is a constant q, in which by moving the timing of any signal in a coalition $A \in X^*$ to one unit in the past and keeping the timing of others the same, its relative weight compared to other elements scales by q.

As a consequence of the representation, under the assumption of the proposition, the weight over a received signal $x \in A$ can be separated into two separate entities. One is the intrinsic value of the signal, captured by $w(x)$. The other one is the role of timing, captured by $q^{T_A(x)}$. Moreover, the only discounting that captures the role of the timing is the exponential form.

If $q = 1$, the timing is not important. Hence, the belief formation process only considers the intrinsic value of each signal. However, when $q \neq 1$, the belief formation process places relatively more weight, in the case of $q \in (0, 1)$, and less weight, in the case of $q \in (1, \infty)$, on a signal received closer to the time of the prediction.

2.5.3 Bayesian Updating

In this section, the set of signals represents the information structure of an agent who wants to predict the true state of nature. We interpret each subset of signals as an event in her information structure. We show that as long as the information structure has a finite cardinality, the strict weighted averaging axiom is the necessary and sufficient condition for a rich belief formation process to appear as a Bayesian updater. In other words, the belief formation (1) attaches a probability measure to each event; (2) by observing a set of disjoint events, she updates her belief through the Bayes rule.

Formally, let $(X, X^* \cup \{\varnothing\})$ be the measure space of events, where X has a finite number of disjoint events. The space of events captures the information structure of the belief formation process. Similarly, by considering the set $\Omega = \{1, \ldots, n\}$, we denote $(\Omega, 2^\Omega)$ as the measure space of states of nature, where

2^Ω is the set of subsets of the set Ω. For any probability distribution $d \in \Delta(\Omega)$ and any subset of the state of nature $B \in \Omega$, let $d(B)$ denote the probability of B which is induced by the distribution d. Hence, $d(B) = \sum_{\omega \in B} d(\omega)$.

Definition 11. A belief formation process $f : X^* \to \Delta(\Omega)$ is ***Bayesian***, if there exists a probability measure P on the space $(\Omega \times X, 2^{\Omega \times X})$, such that for every $A \in X^*$ and $B \in 2^\Omega$ we have:

$$\big(f(A)\big)(B) = \frac{P(B \times A)}{P_X(A)}, \tag{2.6}$$

where, P_X is the marginal probability distribution of P over X.

The right hand side of the previous equation is the conditional probability of B given A. Therefore, a Bayesian belief formation process f behaves as a Bayesian updater: by observing an event A in her information structure X^*, her prediction about the probability of the true state being in a subset $B \in \Omega$ comes from the Bayes rule. To put it differently, $\big(f(A)\big)(B)$ is equal to the conditional probability $P(B|A)$.

Our next proposition shows that our strict weighted averaging axiom is the necessary and sufficient condition for a rich belief formation process to be Bayesian. Since the result is an immediate consequence of our Theorem 4, we put the proof right after the proposition.

Proposition 10. *A rich belief formation process is Bayesian if and only if it satisfies the strict weighted averaging axiom.*

Proof. First, we prove the necessary part. Let the probability measure P rationalize a belief formation process $f : X^* \to \Delta(\Omega)$. Consider two disjoint events $A_1, A_2 \in X^*$. Since f is rationalized by P, we have $\big(f(A_1 \cup A_2)\big)(B) = \frac{P(B \times (A_1 \cup A_2))}{P_X(A_1 \cup A_2)}$ for all $B \in 2^\Omega$. Since $A_1 \cap A_2 = \emptyset$, we have $P(B \times (A_1 \cup A_2)) = P(B \times A_1) + P(B \times A_2)$ and $P_X(A_1 \cup A_2) = P_X(A_1) + P_X(A_2)$. Hence, $\big(f(A_1 \cup A_2)\big)(B) = \frac{P(B \times (A_1 \cup A_2))}{P_X(A_1 \cup A_2)} = \frac{P_X(A_1)}{P_X(A_1)+P_X(A_2)} \frac{P(B \times A_1)}{P_X(A_1)} + \frac{P_X(A_2)}{P_X(A_1)+P_X(A_2)} \frac{P(B \times A_2)}{P_X(A_2)}$.

By defining the $\lambda := \frac{P_X(A_1)}{P_X(A_1)+P_X(A_2)}$, for all $B \in 2^\Omega$, we have $\big(f(A_1 \cup A_2)\big)(B) = \lambda\big(f(A_1)\big)(B) + (1-\lambda)\big(f(A_2)\big)(B)$. Therefore, $f(A_1 \cup A_2) = \lambda f(A_1) + (1-\lambda) f(A_2)$. Hence, f satisfies the strict weighted averaging axiom.

Since f is rationalizable by P, in equation 2.6, in order for f to be well defined, the denominator should be positive, *i.e.*, for all $A \in X^*, P_X(A) > 0$.

Therefore, $\lambda > 0$. As a result, f satisfies the strict weighted averaging axiom.

For the sufficient part, if a rich function $f : X^* \to \Delta(\Omega)$ satisfies the strict weighted averaging axiom, then by applying Theorem 5, there exists a positive weight function $w : X \to \Delta(\Omega)$, such that for all $A \in X^*$:

$$f(A) = \frac{\sum_{x\in A} w(x)f(x)}{\sum_{x\in A} w(x)}.$$

Therefore for all $A \in X$ and $B \in 2^{\Omega}$ we have:

$$\big(f(A)\big)(B) = \frac{\sum_{x\in A} w(x)\Big(\big(f(x)\big)(B)\Big)}{\sum_{x\in A} w(x)}. \tag{2.7}$$

Since the space X is finite, $\sum_{x\in X} w(x) < \infty$. Hence, for every $A \in X^*$ and $B \in 2^{\Omega}$, the probability measure P on the measurable space $(\Omega \times X, 2^{\Omega} \times X^*)$ can be defined as $P(B \times A) = \frac{\sum_{x\in A} w(x)\Big(\big(f(x)\big)(B)\Big)}{\sum_{x\in X} w(x)}$. Since each $f(x)$ is a probability measure, the marginal distribution of P over X is just $P_X(A) = \sum_{x\in A} w(x)$. Therefore, by equation 2.7, for all $A \in X^*$ and $B \in 2^{\Omega}$, we have:

$$\big(f(A)\big)(B) = \frac{P(B \times A)}{P_X(A)}.$$

This representation completes the proof. $\square$

Note that the richness condition is crucial. Otherwise, as shown in Example 11, there are cases where a belief formation process satisfies the strict weighted averaging axiom, but it is not a Bayesian updater.

Remark 12. In our proof, it is necessary that the set of disjoint events, X, be a finite set. Otherwise, with the same technique, we can show that there exists a finite additive measure P that rationalizes the belief formation process under the strict weighted averaging axiom. However, the measure is not a probability measure.

Remark 13. Since the range of a belief formation process is in the simplex $\Delta^{n-1} = \{p|\ p \in \mathbb{R}^n,\ \sum_{i=1}^n p_i = 1\}$, the dimension of the range of f is at most the dimension of Δ^{n-1}, which is $n-1$. Therefore, in order for a belief formation process to satisfy the richness condition, the dimension of the space Ω should be at least 3.

We can show the more general form of the result by adding the strong richness condition and weakening the strict weighted averaging axiom to the weighted averaging axiom. In the more general form, it is possible to have zero probability events. The belief formation process behaves as a Bayesian updater even conditional on observing a zero probability event. To capture the idea, we need the following definition.

Definition 12. A class of functions $\{P_A|\ P_A : 2^{\Omega} \times X^* \to [0,1],\ A \in X^*\}$ is a ***conditional probability system*** if it satisfies the following properties:

1. For every $\varnothing \neq A \in X^*$, P_A is a probability measure on $\Omega \times X$ with $P_A(\Omega \times A) = 1$.
2. For every disjoint events $A_1, A_2 \in X^*$ and for every $C \in \Omega \times X$, we have:

$$P_{A_1 \cup A_2}(C) = P_{A_1 \cup A_2}(\Omega \times A_1)P_{A_1}(C) + P_{A_1 \cup A_2}(\Omega \times A_2)P_{A_2}(C).$$

In the definition above, the probability measure P_{Ω} represents the agent's *ex-ante* prediction. Suppose the agent received information in the form of an event $A \in X^*$. The probability measure P_A represents the agent's *ex-post* prediction, which is a conditional probability measure given the event A. It means that for any set $B \in \Omega$, $P_A(B \times A)$ is the conditional probability of B given A. Moreover, for any two events $A_2 \subset A_1$ in X^*, $P_{A_1}(\Omega \times A_2)$ is the conditional probability of event A_2 given A_1.

In the definition, the first property states that the agent, conditional on realizing an event A in her information structure, does not put any probability on the complement event A^c.

The second property states that regardless of the *ex-ante* probability of an event $A_1 \cup A_2$, conditional on observing $A_1 \cup A_2$, the Bayes updating rule should be satisfied. It is important to mention that the *ex-ante* probability of the event $A_1 \cup A_2$, which is $P_{\Omega}(\Omega \times A_1 \cup A_2)$, might be zero. However, conditional on observing the event $A_1 \cup A_2$, the *ex-post* prediction measure $P_{A_1 \cup A_2}$ is the updated prediction.

A belief formation that is rationalizable by a conditional probability system behaves as a Bayes updater regardless of what event it encounters (even if the *ex-ante* probability of an event is zero). By adding the strong richness condition, the next theorem shows that the weighted averaging axiom is the necessary and sufficient condition for rationalizing a belief formation process by a conditional probability system.

Proposition 11. *A strongly rich belief formation process is rationalizable by a conditional probability system if and only if it satisfies the weighted averaging axiom.*

This proof is similar to the proof of proposition 10. The only difference is that we use the representation under the weighted averaging axiom to derive the result. First, we get a weak order and a weight function using Theorem 5. Then, we can define the probability distribution over each equivalence class of the weak order similar to the proof of Proposition 10 and put zero probability over all lower-ordered events. Therefore, we can define the conditional probability system associated with the belief formation process.

Finally, we should mention that Shmaya et al. (2007) considers the problem of characterizing the *updating rules* (in our context the belief formation processes) that appear to be Bayesian. By providing an example, they show that their *soundness condition*, our strict weighted averaging axiom, is not a sufficient condition for an updating rule to behave as a Bayesian updater. However, we show that the strict weighted averaging axiom is the necessary and sufficient condition as long as the belief formation process satisfies our richness condition.

2.6 Average Choice Functions

2.6.1 Primitives

In this section, an aggregation rule represents an agent choosing randomly from a given menu. The outcome of the aggregation rule is the average choice (mean of the distribution of choices) rather than the entire distribution of choices[4]. Average choice is easier to report and obtain rather than the entire distribution. However, an average choice does not uniquely reveal the underlying distribution of choices.

Our goal is to show that (1) it is possible to uniquely extract the underlying distribution of choices as long as the average choice satisfies the weighted averaging axiom, and (2) there is a connection between our weighted averaging axiom and the Path Independent, Luce, and two-stage Luce choice models.

[4]Check Ahn et al. (2018) for the complete discussion of the merits of average choice.

In this section, X is a nonempty subset of $\mathbb{R}^n$, which is not a subset of a line. In other words, the dimension of X, defined as the dimension of the smallest linear variety that contains X, is greater than one. An aggregation rule in this context is a *average choice function*, in which over any given menu of choice objects, chooses an average choice in the convex combination of members of the menu. Formally we have the following definition:

Definition 13. An aggregation rule $f : X^* \to \mathbb{R}^n$ is called an ***average choice function***, if for any menu $A \in X^*$, $f(A) \in \text{Conv}(A)$.

Based on this definition, the choice over any single-element menu is the element itself. In a larger menu, the value of an average choice function is a vector in the convex combination of the elements of the menu, which can be viewed as the mean of a probability distribution over the elements.

Since the choice from any single-element menu is the element itself, given two different elements in X, the choice from each of them is different. This property of the average choice functions is more restrictive than the previous section where there was a possibility that the values of an aggregation rule over two different single-element menu are the same.

Since average choice functions are a special form of aggregation rules; by adding the strict weighted averaging axiom, the weighted averaging axiom, or the continuity, we can get the representations of the previous section. In this context, the weighted averaging axiom specifies that the choice from the union of two disjoint menus is on the line segment joining the choices from each of the menus.

Under the assumptions of Theorem 5, for any average choice function under the weighted averaging axiom, with a minor richness condition, there exist a weight function and a weak order over elements such that the choice from any menu is just the weighted average (with respect to the weight function) of the highest order (with respect to the weak order relation) elements of the menu.

Adding continuity requires the elements to have the same order with respect to the weak order relation. Under the representation of Theorem 7, the average choice from any given menu is just the weighted average of the menu. To be more precise, and as a starting point for our discussion, we have the following corollary:

Corollary 4. *Let an average choice function $f : X^* \to Conv(X)$ be strongly rich. The following statements are equivalent:*

1. *The average choice function f satisfies the weighted averaging axiom.*

2. *There exists a unique weak order $\succcurlyeq$ on X and a unique weight function $w : X \to R_{++}$, up to multiplication over equivalence classes of the weak order such that for every $A \in X^*$:*

$$f(A) = \frac{\sum\limits_{x \in M(A,\succcurlyeq)} w(x)x}{\sum\limits_{x \in M(A,\succcurlyeq)} w(x)} = \sum_{x \in M(A,\succcurlyeq)} \left(\frac{w(x)}{\sum\limits_{x \in M(A,\succcurlyeq)} w(x)} \right) x. \tag{2.8}$$

Moreover, if the average choice function f satisfies continuity and the weighted averaging axiom, the weight function w is continuous and the weak order $\succcurlyeq$ is the equivalence order. In this case, for every $A \in X^*$:

$$f(A) = \frac{\sum\limits_{x \in A} w(x)x}{\sum\limits_{x \in A} w(x)} = \sum_{x \in A} \left(\frac{w(x)}{\sum\limits_{x \in A} w(x)} \right) x. \tag{2.9}$$

Using the results of this corollary, in the following two subsections, we look into two forms of choice behavior captured by the weighted averaging axiom. In Section 2.6.2, we give a necessary and sufficient condition for an average choice function to be rationalizable by a *Luce* Model. Then, we generalize the result and provide the class of all average choice functions rationalizable by a *two-stage Luce* Model. In Section 2.6.3, we provide a stronger version of the impossibility result by Kalai and Megiddo (1980), regarding the impossibility of a choice function satisfying both the *path independence* axiom , proposed by Plott (1973), and continuity.

2.6.2 Luce Rationalizable Average Choice Functions

In this section, we show that any average choice function under the weighted averaging axiom can be rationalized by a two-stage Luce rule. As discussed in Section 2.6.1, except for the case that elements of a menu are affinely independent, average choice does not uniquely reveal the underlying distribution of choices. However, we show that under the strong richness condition the two-stage Luce model is the only choice model that satisfies the weighted averaging axiom.

In the rest of this section, first we consider a special case. We show that the Luce model is the only choice model satisfying the strict weighted averaging axiom. Therefore, the strict weighted averaging axiom is conceptually equivalent to the *Luce's independence of irrelevant alternatives* axiom, which states that the probability of selecting one element over another in all menus containing both of them should be the same. Following the first result, we show the more general case, in which the two-stage Luce model is the only choice model satisfying the weaker weighted averaging axiom.

The following definitions are standard definitions in the context of individual decision making.

Definition 14. A ***stochastic choice*** is a function $\rho : X^* \to \Delta(X)$, that for any $A \in X^*$, $\rho(A) \in \Delta(A)$.

For an average choice function $f : X^* \to Conv(X)$ and a menu $A \in X^*$, $f(A) \in \text{Conv}(A)$. Therefore, there exists a stochastic choice $\rho : X^* \to \Delta(X)$ (which may not be unique) that rationalizes the average choice function f, *i.e.*, $f(A) = \sum_{x \in A} \rho(x, A)x$, where $\rho(x, A)$ is the probability of selecting the element x from the menu A.

One appealing form of a stochastic choice function is the one that satisfies *Luce's IIA*, *i.e.*, the probability of selecting an element over another element is independent of any other element. Luce (1959) shows that stochastic choices that satisfy the IIA axiom are in the form of Luce rules.

Definition 15. A stochastic choice $\rho : X^* \to \Delta(X)$ is a ***Luce rule*** if there is a function $w : X \to R_{++}$, such that:

$$\rho(x, A) = \frac{w(x)}{\sum_{y \in A} w(y)}.$$

Furthermore, if w is continuous, then ρ is a continuous Luce rule.

Definition 16. An average choice function f is rationalizable by a stochastic choice ρ, if for all $A \in X^*$:

$$f(A) = \sum_{x \in A} \rho(x, A)x.$$

Furthermore, if there exists a Luce rule that rationalizes the average choice function f, then f is ***Luce rationalizable***.

By considering our Theorem 4 and corollary 4, a choice f has a Luce form representation, *i.e*, $f(A) = \sum_{x\in A}(\frac{w(x)}{\sum_{x\in A} w(x)})x$ if and only if it satisfies the strict weighted averaging axiom. As a result:

Corollary 5. *An average choice function is Luce rationalizable if and only if it satisfies the strict weighted averaging axiom. Moreover, the Luce rule that rationalizes the average choice function is unique.*
Furthermore, an average choice function is continuous Luce rationalizable if and only if it satisfies the strict weighted averaging axiom and continuity.

In the Luce model, the decision maker selects each element of a given menu with a strictly positive probability. However, this is not a plausible assumption in many situation. The decision maker may always select a better choice between two alternatives. We model this behavior by a two-stage Luce model. Echenique et al. (2018) introduce the two-stage Luce model. In this model, there exist a ranking order and a weight function over elements. A decision maker facing a menu only selects the highest-ordered elements from the menu. The probability of the selection of each highest-ordered element is related to the weight associated with the element. Formally:

Definition 17. A stochastic choice $\rho : X^* \to \Delta(X)$ is a ***two-stage Luce rule*** if there are a function $w : X \to R_{++}$ and a weak order $\succcurlyeq$ over elements of X, such that:

$$\rho(x, A) = \begin{cases} \frac{w(x)}{\sum_{y\in M(A,\succcurlyeq)} w(y)} & \text{if } x \in M(A, \succcurlyeq), \\ 0 & \text{otherwise.} \end{cases} \tag{2.10}$$

Given a menu A, the decision maker only selects the elements in $M(A, \succcurlyeq)$, that are the highest-ordered elements of A. She chooses each element of $M(A, \succcurlyeq)$ with a probability associated with its weight.

By considering our Theorem 5, any average choice function under the weighted averaging axiom is rationalizable by a two-stage Luce rule.

Corollary 6. *A strongly rich average choice function is two-stage Luce rationalizable if and only if it satisfies the weighted averaging axiom. Moreover, the two-stage Luce rule that rationalizes the average choice function is unique.*

Remark 14. Note that by adding the continuity axiom using our Theorem 7, both the two-stage Luce model and Luce model are equivalent. The next section discusses this observation in more detail.

2.6.3 Continuous Average Choice Functions

In this section, we consider the class of average choice functions satisfying both the weighted averaging axiom and continuity. First, we reinterpret our corollary 4 as an impossibility result. This means that no continuous average choice function is rationalizable by a two-stage Luce model but not by a Luce model. Then, we show the connection with the impossibility result by Kalai and Megiddo (1980), regarding the impossibility of a choice function satisfying both the path independence and continuity.

Plott (1973) extensively studies choice functions under the path independence axiom. Plott's notion of path independence requires a choice from the union of two disjoint menu $A \cup B$, to be the choice between the choice from A and the choice from B. Using his axiom, the choice from any menu can be recursively obtained by partitioning the elements of the menu into disjoint sub-menus. Then, the choice from the whole menu would be the choice from the choices of each sub-menu. In our setup, for an average choice function f, we have:

Axiom 6. ***(Path Independence)*** If $A, B \in X^*$ and $A \cap B = \varnothing$, then

$$f(A \cup B) = f(\{f(A), f(B)\}).$$

The path independence axiom is stronger than our weighted averaging axiom. In other words, any average choice function under Plott's notion of path independence satisfies the weighted averaging axiom. More precisely, given a choice function $f : X^* \to Conv(X)$ and two disjoint menus $A, B \in X^*$, under the path independence axiom, $f(A \cup B) = f(f(A), f(B))$. By the definition of average choice functions, $f(\{f(A), f(B)\}) \in Conv(f(A), f(B))$, which shows that the choice function f satisfies the weighted averaging axiom.

As discussed in Section 2.4, continuity is an appealing property of an average choice function. It specifies that by replacing an element of a menu with another element close to it, with respect to the norm of X, the average choice of the new menu is close to the average choice of the previous menu.

Kalai and Megiddo (1980) show that there is no average choice function that satisfies both path independence axiom and continuity. Here, we reinterpret the result of Corollary 4 to show a more general result for average choice functions.

Corollary 4 states that for an average choice function $f : X^* \to Conv(X)$ under continuity and the weighted averaging axiom, as long as it satisfies the strong richness condition, there exists a unique weight function $w : X \to \mathbb{R}_{++}$ such that for any $A \in X^*$:

$$f(A) = \sum_{x \in A} \left(\frac{w(x)}{\sum_{x \in A} w(x)}\right) x.$$

There are two important observations regarding the representations above.

First, through discussions in Section 2.6.2, the representation shows that any continuous average choice function that is rationalizable by a two-stage Luce model is also rationalizable by a Luce model. Second, since the function w is strictly positive, the average choice of any menu should be in the relative interior of the convex hull of members of the menu.

As a result, our impossibility result specifies that for an average choice function that satisfies the weighted averaging axiom, it is impossible to satisfy the continuity axiom and also to have a choice from a menu that is on the relative boundary of the elements of the menu. We summarize the observation in the following corollary.

Corollary 7. *If X is a nonempty convex subset of a vector space that contains at least three non-collinear points, then an average choice function $f : X^* \to X$ that satisfies the weighted averaging axiom cannot be both continuous and contains a menu $A \in X^*$, with $f(A) \in \partial^r(Conv(A))$.*

To see the connection between our corollary 7 and the result in Kalai and Megiddo (1980), it is enough to consider a menu with three non-collinear members. Theorem 1 in Kalai and Megiddo (1980) shows that the average choice of a path independent average choice function from any menu is the average choice of the average choice function from a sub-menu of two members of the menu. This shows that the average choice from a menu with three non-collinear members is on the line segment connecting two of the member of the menu. As a result, the choice should be on the relative boundary of the menu. That is why it cannot satisfy continuity.

2.7 Related Literature

Our methods are applicable to different areas of economic theory, and generalize existing ideas in those areas. In particular, instances of our weighted averaging axiom appears in several different papers.

The theory of Case-Based Prediction is developed by the seminal works of Gilboa and Schmeidler (1995) and Gilboa and Schmeidler (2012) and Billot et al. (2005). In this context, the *concatenation* axiom proposed by Billot et al. (2005), is closely related to the strict case of our axiom. However, there are differences between the two axioms. As discussed in more detail in Section 2.5.1, their belief formation process is defined over "sequences" of cases, in which each sequence can have multiple copies of the same case. The role of the concatenation axiom is to count the number of each case. However, in our framework, we define our axiom over "sets" of signals, in which in each set there is only one copy of each signal. Moreover, our axiom is defined over disjoint sets. By weakening our definition for any two general sets, our result does not hold anymore.

In the paper by Shmaya et al. (2007), they provide an example, on binary state space, to show that their *soundness* condition is not a sufficient condition for an updating rule to behave as a Bayesian rule. However, we show that under our richness assumption, the strict weighted averaging axiom (which is the same as their soundness condition) is the necessary and sufficient condition for an updating rule to behave as a Bayesian. We also generalize our result for the class of updating rules that can be rationalized by a conditional probability system.

In the context of choice theory, Ahn et al. (2018) introduce a model of continuous average choice over convex domains. In this application, we generalized their result in many ways. First, their result holds for the strict case of our axiom. Moreover, continuity and convexity are the two important forces behind their result. However, we show that the strictness of an average choice function, continuity, or convexity of the domain are not the main forces behind extracting the underlying distribution of choices. The main force is our weighted averaging axiom. Moreover, we show that it is possible to rationalize an average choice function by a two-stage Luce model, as long as it satisfies our weighted averaging axiom.

The path independence choice functions are extensively studied by Plott (1973).

Our representation of average choice functions under the weighted averaging axiom and continuity, generalizes the results by Kalai and Megiddo (1980) and Machina and Parks (1981), regarding the impossibility of a choice function under both the path independence and the continuity.

Chapter 3

EXTENDED PARETO AGGREGATION

3.1 Introduction

In this chapter, we consider the problem of aggregation of preference ordering of a group of individuals to form a social preference ordering. The goal is to find an aggregation procedure with some appealing properties. In our setup each preference ordering satisfies the axiom of Von-Neumann and Morgenstern (1944). The role of an aggregation rule is to associate with each coalition of individuals another vN-M preference ordering over the set of alternatives.

An appealing property of an aggregation rule, in this context, is to satisfy the *extended Pareto* axiom. Shapley and Shubik (1982) introduced the extended Pareto. It specifies that, if two disjoint coalitions of individuals, each prefers an outcome over another outcome, then the union of the coalitions also should prefer the same outcome over the other one. Moreover, if one of them strictly prefers one outcome over the other one, then the union of the coalitions should also strictly prefer the same outcome over the other one.

First, we show that under a normalization of cardinal utilities of individuals and a minor richness condition, aggregation rules under the strict weighted averaging (weighted averaging) axiom are exactly aggregation rules under the *extended Pareto* (*extended weak Pareto*) axiom.

Following the equivalence, we use our main representation result on weighted averaging aggregation rules as a technical tool to pin down the representation of the extended Pareto aggregation rules. We show that the only possible extended Pareto aggregation is to have a positive weight over each individual in the society. Then, the aggregated preference ordering of a given group of individual is the weighted sum of their preference ordering.

The representation can be considered as a generalization of the theorem by Harsanyi (1955) on utilitarianism. Harsanyi considers a single profile of individuals and a variant of Pareto to get the Utilitarianism. However, in our approach, we partition a profile to smaller groups. Then, we aggregate the preference ordering of these smaller groups using the extended Pareto. Hence, we get the Utilitarianism through this consistent form of aggregation. As a result, in our representation, the weight associated with each individual appears in all sub-profiles that contain her.[1]

We provide two more applications of this representation. One is a result with a positive nature on social choice functions, which we discussed in Section 3.2.3. The other application is to provide a subjective conditional expected theory of state-dependent preferences, which we discuss in Section 4.2.

In Section 3.2.3, we extend our result on extended Pareto aggregation rules to the class of *generalized social welfare function*. Unlike our previous model, each individual may have any different preference ordering. Therefore, the domain of the generalized social welfare function is a set of all different groups (with all possible sizes) of individuals with each individual having all different possible preference orderings. Our definition of generalized social welfare function extends the standard definition used by Arrow (1963), in which the domain is a set of fixed-length profiles of individuals.

For a technical reason, we restrict the set of vN-M preferences to those which all of them strictly prefer one fixed lottery to another fixed one. We show that the only possible extended Pareto generalized social welfare functions are the ones that associate a positive number to each individual's preferences (unlike the previous section, in which each weight depends on both the individual and the whole profile) and it associates each coalition with the weighted sum of their cardinal utility using the weight associated to their preferences.

The important observation is that, *each positive weight in the representation is independent of the other individuals in any profiles.* The weight only depends on each individual and her own preference ordering.

Our representation above has a positive nature, compare to the claims by Kalai and N. Schmeidler (1977) and Hylland (1980) that the negative conclusion of Arrow's theorem holds even with vN-M preferences. Moreover, the representation provides an answer to the main concern of Borgers and Choo (2017a) and Borgers and Choo (2017b) regarding the correctness of the main theorem of Dhillon (1998).

---

[1] Similar to the discussion of Weymark (1991) regarding the debate of Sen-Harsanyi, our result is better to be interpreted as a representation rather than a justification of the utilitarianism.

Dhillon (1998) by considering a set of axioms, other than the ones by Arrow, provides one of the first axiomatizations of relative utilitarianism as a possibility result. However, Borgers et al. (2017a) shows a counterexample to their representation. Our representation fixes the error using our variant of the extended Pareto axiom and our restricted domain of the generalized social welfare function.

Finally, adding the anonymity and the weak IIA axiom of Dhillon (1998) gives us the relative utilitarianism as one possible choice of the weight function. However, the primary concern of our paper is to show that *the weighted averaging of preferences is the only generalized social welfare function that respects extended Pareto*. The possible choices of weights are not our focus in this paper.

3.2 Extended Pareto Aggregation Rules

In this section, each individual has a preference ordering of a set of alternatives. The role of an aggregation rule is to associate with each group of individuals another preference ordering of alternatives, representing the group's aggregated preference.

We consider the class of aggregation rules satisfying the *extended Pareto* axiom. This means that whenever we partition a group of individual into two subgroups, if both aggregated preferences of subgroups prefer one alternative over another, the aggregated preference of the union of the subgroups also prefers the first alternative over the second.

Section 3.2.1 provides the main definitions. In section 3.2.2, we show that (1) the extended Pareto is equivalent to the strict weighted averaging axiom, and (2) Theorem 5 will imply a general version of Harsanyi (1955) famous theorem on Utilitarianism. Harsanyi considers a single profile of individuals and a variant of Pareto to obtain Utilitarianism. However, in our approach, we partition a profile into smaller groups. Then, we aggregate the preference ordering of these smaller groups using the extended Pareto. Hence, we model Utilitarianism through this consistent form of aggregation. As a result, in our representation, the weight associated with each individual appears in all sub-profiles that contain her. Finally, section 3.2.3 generalizes the result of section 3.2.2 for generalized social welfare functions by extending the domain of aggregation rules to the set of all different groups of individuals with each individual having all possible preference orderings. We discuss the connection to relative Utilitarianism at the end of the section.

3.2.1 Primitives and axioms

Let the set $M = \{0, 1, \ldots, m\}$ represent $m + 1$ pure prospects. The simplex $L = \{(p_1, \ldots, p_m) | \sum_{i=1}^{m} p_i \leqslant 1, p_i \geqslant 0\}$ represents the set of lotteries over the set M. A lottery $p \in L$ associates the probability p_i to the prospect $i \in M\backslash\{0\}$ and $1 - \sum_{i=1}^{m} p_i$ to the prospect 0.

A *vN-M preference* over the set L is a preference relation that satisfies the axioms of Von-Neumann et al. (1944), *i.e.*, it is a weak order that satisfies both the continuity and independence axiom. Hence, a vN-M preference has an expected utility representation. We denote the set of all vN-M preferences over the set of lotteries L, by $\mathcal{R}$. We denote the strict part of the preference $R \in \mathcal{R}$, by $\overline{R}$.

Remark 15. Let $R \in \mathcal{R}$, be a vN-M preference over the set L. Using the vN-M Theorem, there exists an affine representation of the preference R. For notational convenience, we normalize all affine representations to have the value 0 over the prospect 0. Therefore, there exists a utility $u \in \mathbb{R}^m$ that represents R, *i.e.*, for any two lotteries $x, y \in L$, xRy if and only if $u \cdot x \geqslant u \cdot y$, where "$\cdot$" represents the inner product in $\mathbb{R}^m$. Moreover, the "unique" *ray* $U = \{\alpha u | \ \alpha > 0\}$ contains all normalized affine utilities that represent the vN-M preference R.

In this section, we interpret each feature as an individual. The set of features $X = \{1, \ldots, n\}$ represents the set of all agents. The set X^* represents the set of all coalitions of agents. We denote the X-Fold Cartesian product of $\mathcal{R}$, by $\mathcal{R}^X$. Every $R^X \in \mathcal{R}^X$ defines a *preference profile* of the set of agents over the set of lotteries.

The main representation of this section is about a specific form of aggregation rules.

Definition 18. A ***group aggregation rule*** on X is a function $f : X^* \to \mathcal{R}$, that associates with every coalition of agent $A \in X^*$ a vN-M preference $f(A) \in \mathcal{R}$.

Let $f : X^* \to \mathcal{R}$ be a group aggregation rule and $i \in X$. We interpret $f(i)$ as the preference of agent i. For every coalition $A \in X^*$, we interpret the preference $f(A)$ as the aggregated vN-M preference of the members of coalition A. By considering remark 15, for every coalition $A \in X^*$, we associate the unique ray U_A that represents the vN-M preference $f(A)$. Every cardinal utility $u \in U_A$ represents the vN-M preference $f(A)$.

An appealing property of group aggregation rules is that whenever two disjoint coalitions, *e.g.* $A, B \in X^*$, both prefer a lottery x to another lottery y, then their union, $A \cup B$, also prefers the lottery x to the lottery y. Formally, we have:

Axiom 7. A group aggregation rule $f : X^* \to \mathcal{R}$ satisfies the ***extended Pareto axiom*** if for all disjoint coalitions of agents $A, B \in X^*$, and for all lotteries $x, y \in L$,

$$x\ f(A)\ y,\ x\ f(B)\ y \Rightarrow x\ f(A \cup B)\ y \tag{3.1}$$

$$x\ \overline{f(A)}\ y,\ x\ f(B)\ y \Rightarrow x\ \overline{f(A \cup B)}\ y. \tag{3.2}$$

The first condition, which we refer to as the *extended weak Pareto axiom*, states that whenever two disjoint coalitions $A, B \in X^*$ prefer lottery x to lottery y, then the coalition formed as the union of A and B also prefers the lottery x to the lottery y. The second condition states that if one of the coalitions strictly prefers lottery x over lottery y, then the union of the coalition also strictly prefers lottery x over lottery y.

Our last condition requires the existence of two lotteries in the set of lotteries, in which all agents strictly prefers one over the other.

Definition 19. A group aggregation rule $f : X^* \to \mathcal{R}$ satisfies the ***minimal agreement condition*** if there exist two lotteries $\overline{x}, \underline{x} \in L$ such that for every agent $i \in X$, $\overline{x}\ \overline{f(i)}\underline{x}$.

Remark 16. Let a group aggregation rule $f : X^* \to \mathcal{R}$ satisfy both the minimal agreement and extended Pareto axiom. Given two agents $i, j \in X$, by applying the strict part of the definition of the extended Pareto axiom, we have $\overline{x}\ \overline{f(\{i,j\})}\ \underline{x}$. Similarly, for every coalition of agents $A \in X^*$, recursively using the strict part of the extended Pareto axiom, we can show that $\overline{x}\ \overline{f(A)}\ \underline{x}$.

Remark 17. There is another useful equivalent definition for the minimal agreement condition. Let the vector $v \in \mathbb{R}^m$ be $\overline{x} - \underline{x}$, where $\overline{x}, \underline{x}$ are the two lotteries in the definition of the minimal agreement condition. By considering remark

15, let a cardinal utility $u_i \in \mathbb{R}^m$ represent the vN-M preference $f(i)$. Hence, $\overline{x}\ \overline{f(i)}\ \underline{x}$ if and only if $u_i \cdot v > 0$. Therefore, the definition of the minimal agreement condition is equivalent to the existence of a direction $v \in \mathbb{R}^m$ such that for all $i \in X$, $u_i \cdot v > 0$.

In the next section, first, we show that any group aggregation rule (under the minimal agreement condition) satisfies the extended Pareto if and only if it satisfies our strict weighted averaging axiom. Then, using the result of Theorem 4, we characterize the representation of extended Pareto group aggregation rules.

3.2.2 The Representation of Extended Pareto Group Aggregation Rules

In this section, we assume that the group aggregation rule $f : X^* \to \mathcal{R}$ satisfies the minimal agreement condition. In particular, we assume that all agents strictly prefer the lottery $\overline{x} \in L$ over the lottery $\underline{x} \in L$. By considering remark 17, we define $v = \overline{x} - \underline{x}$ as the direction that every agent agrees on. Following the comments right after the definition 18, for a coalition of agents $A \in X^*$, the ray U_A represents the vN-M preference $f(A)$.

We define the space H by $H = \{u \in \mathbb{R}^m |\ u \cdot v = 1\}$. The space H represents the normalization of utilities in which the difference of the value of utility of the lottery $\overline{x}$ and the lottery $\underline{x}$ is exactly 1. For every coalitions of agents $A \in X^*$ and for every cardinal utility $u_A \in U_A$, by the minimal agreement condition and remark 16, we have $u_A \cdot v > 0$. Therefore, for every coalition of agents $A \in X^*$, there is a unique cardinal utility $\hat{u}_A \in U_A$, such that $\hat{u}_A$ is in H.

For the rest of the section, for every coalition $A \in X^*$, we consider the unique cardinal utility $\hat{u}_A \in H$ to represent the vN-M preference $f(A)$. Using this representation, we can represent the group aggregation rule $f : X^* \to \mathcal{R}$, by a ***normalized group aggregation rule*** $f_H : X^* \to \mathbb{R}^m$, where $f_H(A) = \hat{u}_A$.

The role of f_H is to "normalize" the cardinal utilities to be in H. In other words, we represent each vN-M preference by a unique cardinal utility that the difference of the utility between the lottery $\overline{x}$ and $\underline{x}$ is one.

Remark 18. Without loss of generality, we can assume that the lottery $\underline{x}$ in the definition of the minimal agreement condition is just the lottery 0. In that case, the space H is vN-M preferences with the value 0 for the lottery 0 and the value 1 for the lottery $\overline{x}$. The role of f_H is to normalize the cardinal utilities to get the value 0 for the lottery 0 and the value 1 for the lottery $\overline{x}$.

The next proposition shows that under the representation of the vN-M preference $f(A)$ by the $\hat{u}_A$, the extended Pareto axiom is equivalent to the strict weighted averaging axiom. Formally, we have:

Theorem 8. *Let a group aggregation rule $f : X^* \to \mathcal{R}$ satisfy the minimal agreement condition with $v \in \mathbb{R}^m$ as the direction on which all agents agree. Then, the following are equivalent:*

1. *f satisfies the extended Pareto axiom.*
2. *f_H satisfies the strict weighted averaging axiom.*

The idea behind the proof is as follows. Assume that $A, B \in X^*$ are two disjoint coalitions of agents. We select two utilities $u_A \in U_A$ and $u_B \in U_B$. We assume that $u_{A\cup B}$ represents the preference of the union of them, *i.e.*, $u_{A\cup B} \in U_{A\cup B}$. First, an immediate application of Farkas' Lemma shows that $u_{A\cup B}$ should be in the relative interior of the cone generated by the vectors u_A and u_B. Then, by intersecting the cone generated by u_A, u_B and the hyperplane H, we show that $\hat{u}_{A\cup B} \in \mathrm{Conv}^{\circ}(\{\hat{u}_A, \hat{u}_B\})$, which completes the proof.

Finally, using the result of Theorem 4, we immediately attain the representation of the extended Pareto group aggregation rules.

Corollary 8. *Let a rich group aggregation rule $f : X^* \to \mathcal{R}$ satisfy both the extended Pareto axiom and minimal agreement condition. Then, there exists a weight function $w : X \to \mathbb{R}_{++}$ such that for every coalition of agents $A \in X^*$,*

$$f_H(A) = \sum_{i\in A} \left(\frac{w(i)}{\sum\limits_{j\in A} w(j)} \right) f_H(i). \tag{3.3}$$

Moreover, the weight function is unique up to multiplication by a positive number.

As shown in Example 11, the richness condition is crucial. The richness here is equivalent to the existence of three non-collinear "normalized" cardinal utilities in the space H. Therefore, the richness in this case is equivalent to the existence of three linearly independent cardinal utilities in the range of the aggregation rule.[2]

The representation above finds the unique cardinal utility associated with the $f(A)$ in the space H. Hence, we can interpret the theorem as a generalization of the main theorem of Harsanyi (1955) on Utilitarianism. However, our result shows the connection between weights of individuals in different sub-coalitions of the main profile.

To see the connection with Harsanyi's result, we rewrite the theorem in an additive form: let the group aggregation rule $f : X^* \to \mathcal{R}$ satisfy both the extended Pareto axiom and minimal agreement condition. Then, there exists a weight function $w : X \to \mathbb{R}_{++}$ such that for every coalition of agents $A \in X^*$, $f(A)$ has a representation as follows:

$$\sum_{i \in A} w(i) f_H(i). \tag{3.4}$$

Moreover, it is important that the representation is unique up to multiplication.

By defining $u(i) := w(i) f_H(i)$ for $i \in X$, we can rewrite equation 3.4 in a more familiar additive form of:

$$\sum_{i \in A} u(i). \tag{3.5}$$

Moreover, if we consider only the representations with the value 0 for the lottery 0, this representation is "unique" up to multiplication by a positive number.

There are two immediate applications of corollary 8. First, Section 3.2.3 shows an application in characterizing the class of extended Pareto generalized social welfare functions and relative utilitarianism. Then, Section 4.2 covers the other application, or interpretation, by providing a simple theory of subjective conditional expected utility of state-dependant preferences.

3.2.3 The Representation of Extended Pareto Generalized Social Welfare Functions

In this section, by using corollary 8, we provide the characterization of the class of extended Pareto *generalized social welfare* functions. The represen-

[2] This is similar to the argument of remark 13.

tation is a generalization of corollary 8, in which, unlike the previous model, each individual can get any vN-M preference relation. The standard domain of a social welfare function is a set of fixed-length profiles of individuals, where each individual has all possible preference ordering. However, the domain of our generalized social welfare function is the set of all different groups (with all different lengths) of individuals, where each individual has all possible preference orderings.

The setup is the same as the one in the previous section. However, in order to define the domain of generalized social welfare functions, we need some definitions. Without loss of generality, we assume that the lottery $\underline{x} \in L$, in the definition of the minimal agreement, is the vector 0. Let $\overline{x} \in L$ be any lottery other than 0. Define $\mathcal{R}_{\overline{x}} \subset \mathcal{R}$ as the set of all vN-M preferences that strictly prefer $\overline{x}$ to 0. Let $\mathcal{R}_{\overline{x}}^{X}$ be the X-fold Cartesian product of $\mathcal{R}_{\overline{x}}$. Every $R \in \mathcal{R}_{\overline{x}}^{X}$ defines a preference profile of the set of individuals. For any coalition $A \in X^*$ and for any preference profile $R \in \mathcal{R}_{\overline{x}}^{X}$, let $R_A \in \mathcal{R}_{\overline{x}}^{A}$ denote the restriction of the profile R to the coalition A.

Based on remark 15, we can represent each preference $R \in \mathcal{R}$ by a unique ray $U_R = \{\alpha u | \ \alpha > 0\}$, where $u \in \mathbb{R}^m$ is a cardinal utility representing R. Moreover, for any preference $R \in \mathcal{R}_{\overline{x}}$, there should be a unique cardinal utility $u_R \in U_R$ with $u_R \cdot \overline{x} = 1$. Denote $H = \{u \in \mathbb{R}^m | \ u.\overline{x} = 1\}$ as the space of all cardinal utilities attaining value 0 at the lottery 0 and the value 1 at the lottery $\overline{x}$. Let the function $u_H : \mathcal{R}_{\overline{x}} \to H$ associate each preference $R \in \mathcal{R}_{\overline{x}}$ with the unique cardinal utility $u_H(R) \in H$ that represents it. This function is a bijection associating each preference to the unique cardinal utility attaining value 0 at the lottery 0 and value 1 at the lottery $\overline{x}$.

We denote $\mathcal{R}_X \subset \mathcal{R}_{\overline{x}}^{X}$ as the set of all profiles where the representation of individuals' cardinal utilities in the space H is not a subset of a single line. Formally, we define $\mathcal{R}_X = \{R \in \mathcal{R}_{\overline{x}}^{X} | \ d(\{u_H(R_i) | \ i \in X\}) > 1\}$, where $d(\{u_H(R_i) | \ i \in X\})$ is the dimension of the smallest linear variety containing all $u_H(R_i), \ i \in X$.[3]

Finally, we denote $\mathcal{R}_X^* = \{R \in \mathcal{R}_{\overline{x}}^{A} | \ A \subseteq X, R \in \mathcal{R}_X\}$ as all the profiles in $\mathcal{R}_X$ and all sub-coalitions of those profiles. $\mathcal{R}_X^*$ is the domain of our generalized social welfare functions. Formally, we have:

[3] There should be at least four alternatives; otherwise, $\mathcal{R}_X$ is the empty set.

Definition 20. A ***generalized social welfare function*** on $\mathcal{R}_X$ is a function $f : \mathcal{R}_X^* \to \mathcal{R}$, that associates with any coalition $A \in X$ and any profile $R \in \mathcal{R}_{\overline{x}}^X$ a preference $f(R_A) \in \mathcal{R}$. Moreover, we assume that for any individual $i \in X$, and any profile $R \in \mathcal{R}_{\overline{x}}^X$, $f(R_i) = R_i$.

In our setup, the domain of generalized social welfare functions is a rich set of all sizes of profiles. Moreover, it satisfies the ***Individualism*** axiom, which means that it associates any individual preference to the same preference.

The connection between profiles of different sizes is the ***extended Pareto axiom***. The extended Pareto states that if the associated preference ordering of two disjoint coalitions of individuals, A and B, each prefer a lottery x to y, then the associated preference ordering of the union of the coalition with the same preference as before should also prefer x to y.

Definition 21. A generalized social welfare function $f : \mathcal{R}_X^* \to \mathcal{R}$ satisfies the ***extended Pareto*** axiom if for every preference profile $R \in \mathcal{R}_X$ and for any two disjoint coalitions $A, B \in X^*$, and for all lotteries $x, y \in L$,

$$x\ f(R_A)\ y,\ x\ f(R_B)\ y \Rightarrow x\ f(R_{A \cup B})\ y \tag{3.6}$$

$$x\ \overline{f(R_A)}\ y,\ x\ f(R_B)\ y \Rightarrow x\ \overline{f(R_{A \cup B})}\ y. \tag{3.7}$$

Our main result of this section characterizes the class of extended Pareto generalized social welfare functions.

Theorem 9. *Let X be a set of individuals with $|X| \geqslant 4$. The generalized social welfare function $f : \mathcal{R}_X^* \to \mathcal{R}$ satisfies the extended Pareto axiom if and only if there exists a weight function $w : X \times \mathcal{R}_{\overline{x}} \to \mathbb{R}_{++}$, such that for any coalition $A \subseteq X$ and any preference profile $R \in \mathcal{R}_X$, $f(R_A)$ has the following representation:*

$$u_H(f(R_A)) = \sum_{i \in A} \left(\frac{w(i, R_i)}{\sum\limits_{j \in A} w(j, R_j)} \right) u_H(R_i). \tag{3.8}$$

Moreover, the weight function is unique up to multiplication by a positive number.

Note that the representation above finds the unique cardinal utility in the space H, which is associated with the aggregated preference relation $f(R_A)$. In other words, it normalizes the aggregated utility representation of $f(R_A)$ to have the value 0 at the lottery 0 and value 1 at the lottery $\overline{x}$.

Remark 19. We can rewrite the theorem to specify that the generalized social welfare function $f : \mathcal{R}_X^* \to \mathcal{R}$ satisfies the extended Pareto axiom if and only if there exists a weight function $w : X \times \mathcal{R}_{\overline{x}} \to \mathbb{R}_{++}$, such that for any coalition $A \subseteq X$ and any preference profile $R \in \mathcal{R}_X$, $f(R_A)$ has the following representation:

$$\sum_{i \in A} w(i, R_i) u_H(R_i). \tag{3.9}$$

It is important that each weight depends only on the associated individual's preferences and not on the other individuals.

Note that, based on corollary 8, for any fixed profile $R \in \mathcal{R}_X$, there exists a weight function (depending on the whole profile R and each individual index) satisfying equation 3.4. However, in equation 3.9, the weight function depends only on each individual's own preference and not the whole profile. The reason the theorem works is the structure we put on the domain, $\mathcal{R}_X^*$, in the definition of generalized social welfare functions. The proof of the result is in the Appendix.

The weight function in the representation depends on each individual's index. However, adding the classical ***Anonymity*** axiom makes the weight function independent of individual's indexes.

Axiom 8. An extended Pareto aggregation rule $f : \mathcal{R}_X^* \to \mathcal{R}$ satisfies the ***Anonymity*** axiom, if any permutation of the indexes of individuals does not change the generalized social welfare function.

The anonymity axiom makes any extended Pareto generalized social welfare functions independent of the individual's indexes. Hence, the uniqueness of the weight function in Theorem 9 makes the weight function, associated with an anonymous extended Pareto aggregation rule, independent of the indexes. Therefore, we have:

Corollary 9. *Let X be a set of individuals with $|X| \geqslant 5$. The extended Pareto generalized social welfare function $f : \mathcal{R}^*_X \to \mathcal{R}$ satisfies the Anonymity axiom if and only if there exists a weight function $w : \mathcal{R}_{\overline{x}} \to \mathbb{R}_{++}$, such that for any coalition $A \subseteq X$ and any preference profile $R \in \mathcal{R}_X$, $f(R_A)$ has the following representation:*

$$u_H(f(R_A)) = \sum_{i \in A} \left(\frac{w(R_i)}{\sum_{j \in A} w(R_j)} \right) u_H(R_i). \tag{3.10}$$

Or in a more familiar way, $f(R_A)$ has a representation as follows:

$$\sum_{i \in A} w(R_i) u_H(R_i). \tag{3.11}$$

Moreover, the weight function is unique up to multiplication by a positive number.

The representation above is the main result of this section. It is important to point out the positive nature of our result, given the claims by Kalai and N. Schmeidler (1977) and Hylland (1980) that the negative conclusion of the impossibility theorem by Arrow (1963) holds even with vN-M preferences. However, other than the differences between our model and theirs, we only consider the restricted domain where all preferences prefer the lottery $\overline{x}$ over the lottery $\underline{x}$. As discussed before, the definition of our restricted domain is crucial in corollary 9.

Remark 20. Adding different other axioms can restrict choices of the weight function. For example, relative utilitarianism can be obtained by adding the weak IIA axiom of the paper by Dhillon (1998). By adding the weak IIA, the weight function normalizes each preference such that the difference between the cardinal utility of the best alternative and the worst alternative becomes 1. In other words, for any preference $R \in \mathcal{R}_{\overline{x}}$, $w(R) = \frac{1}{\underset{j}{max}\ (u_H(R))_j - \underset{j}{min}\ (u_H(R))_j}$.

3.3 Related Literature

In the context of social choice, Dhillon (1998) and Baucells et al. (2008) study variants of extended Pareto rules.

Baucells et al. (2008) study the extended Pareto rule over vN-M preferences by relaxing the completeness axiom. Besides the technical and conceptual differences between the two approaches and results, their model depends on their *non-degeneracy* condition. The condition is only satisfied when there is a spanning tree over the preferences, and every three consecutive preferences in the spanning tree are linearly independent. However, the richness condition of our theorem only requires three linearly independent vectors among the whole set of preferences. Moreover, our result can be applied even for the class of extended weak Pareto aggregation rules, under our strong richness condition. Note that our primary goal in this paper is to show that extended Pareto and extended weak Pareto are special cases of our weighted averaging axiom (under the minimal agreement condition).

The papers by Dhillon (1998), Dhillon and Mertens (1999), and Borgers et al. (2017b) each by considering different sets of axioms, other than Arrow's, provide an axiomatization of relative utilitarianism as a positive result. The paper by Dhillon (1998) is the closest one to ours. Dhillon considers a variant of extended Pareto to get a weighted averaging structure. However, Borgers et al. (2017a) show a counterexample to the representation. We restrict the domain and use our definition of extended Pareto to get the weighted averaging structure as a consequence of our main theorem on the representation of weighted averaging aggregation rules. Finally, the technique we developed can also be used to provide a representation of the extended weak Pareto social welfare functions.

Chapter 4

A SUBJECTIVE CONDITIONAL EXPECTED UTILITY THEORY OF STATE-DEPENDENT PREFERENCES

4.1 Introduction

The choice-theoretic foundation of subjective expected utility was developed by the seminal works of Ramsey (1931), Savage (1954), and Anscombe and Aumann (1963). In the standard model, the decision maker has a ranking over acts (state-contingent outcomes). The representation of this ranking consists of a subjective probability over the set of states, capturing the decision maker's beliefs, and a cardinal utility representing the decision maker's tastes over the set of outcomes, independent of the realization of the true state. However, in many applications, such as models for buying health insurance, the independence of the utility and the set of states is not a plausible assumption[1].

In Section 4.2, we provide a simple theory of subjective expected utility of state-dependent utility by reinterpreting our representation of extended Pareto aggregation rules. We build our model using the framework of Anscombe et al. (1963). In our model, the decision maker has a preference ordering over the set of conditional constant acts. This means that given any fixed event, the decision maker has a hypothetical conditional preference ordering over the set of lotteries, representing her conditional preference condition on learning that only that event is happening[2]. Each of these hypothetical conditional preferences satisfies the axioms of Von-Neumann et al. (1944), which means each has an affine representation. We show that as long as the class of hypothetical conditional preferences satisfies the extended Pareto axiom, there is a subjective probability measure over the set of states and a state-dependent utility over the set of alternatives. The class of hypothetical conditional preferences has a representation in the form of conditional expectation with respect to the subjective probability and the state-dependent utility.

The result shows that the extended Pareto is the main force behind the separation of the belief and the state-dependent utility. However, the representa-

[1]Check Arrow (1974), Cook and Graham (1997), and Karni (1985) for more discussions.

[2]In Section 4.2, we illustrate another interpretation of hypothetical conditional preferences by providing a preference ordering over the set of conditional constant acts.

tion is not unique. Hence, the challenge is to provide meaning to a decision maker's prior beliefs, when utility is state-dependent. We get the uniqueness by adding a stronger version of our minimal agreement condition. The strong minimal agreement condition specifies that there exist two lotteries where one is strictly preferred to the other, regardless of states. Moreover, the decision maker's conditional preference for each of them is independent of the set of states.

We show that under the strong minimal agreement, the belief is unique. Moreover, the state-dependent utility is unique up to affine transformation.

4.2 Main Model and Results

4.2.1 Primitives

In this section, we develop a simple theory of subjective expected utility of state-dependant utility by reinterpreting the results of sections 3.2.2 and 2.5.3. We show that the extended Pareto is the main force behind the separation of a subjective belief and a state-dependant utility. However, adding a stronger version of our minimal agreement condition allows us to uniquely separate the belief and the state-dependant utility.

Our model is built using the framework of Anscombe et al. (1963). Let $\Omega = \{1, 2, \ldots, n\}$ be a finite set of states of nature. The finite set M represents outcomes. The simplex $L = \Delta(M)$ represents the set of lotteries over the set M. A lottery $l \in L$ associates the probability l_i to the outcome $i \in M$.

In this model, the decision maker faces two levels of uncertainty. First, the decision maker does not know which states will obtain. Second, after realization of the state, she faces the distribution induced by the lottery over outcomes.

There is a single outcome O other than the outcomes of the set M. The outcome O represents that the decision maker can *ex-ante* refuse any lottery.

In our setup, the objects of choice are ***conditional constant acts***. For any lottery $l \in L$, and any event $A \in 2^{\Omega} \backslash \varnothing$, the function $f : \Omega \to L \cup \{O\}$ such that $f(w) = l$ for $\omega \in A$ and $f(w) = O$ for $\omega \in A^c$ is termed a conditional constant act and denoted by $f = (l, A, O, A^c)$. The interpretation is that if event A is realized, the decision maker faces the lottery l. Otherwise, O will be realized. We assume that the decision maker has a preference relation $\succcurlyeq$, and not necessarily a complete relation, over the set of conditional constant acts.

Let $F = \{(l, A, O, A^c)|\ \varnothing \neq A \in 2^\Omega, l \in L\}$ represent the set of conditional constant acts. For any event $\varnothing \neq A \in 2^\Omega$, let $F_A = \{(l, A, O, A^c)|l \in L\}$ be the set of all conditional constant acts attaining a lottery on the event A and staying out on the event A^c. We represent the conditional preference ordering of the decision maker over F_A by $\succcurlyeq_A$. For any two lotteries l_1, l_2, we write $l_1 \succcurlyeq_A l_2$ as a shorthand of $(l_1, A, O, A^c) \succcurlyeq (l_2, A, O, A^c)$.

Our interpretation of conditional preference ordering is related to the models developed by Luce and Krantz (1971), Fishburn (1973), Skiadas (1997a), and Karni and D. Schmeidler (2007). However, there is another interpretation of the conditional preference similar to the conditional decision model of Ghirardato (2002). In this interpretation, we assume that the decision maker may receive some information that only $\omega \in A$ can be realized. In this case, $\succcurlyeq_A$ represents the decision maker's *ex-post* preference over the set of lotteries. Similarly, $\succcurlyeq_\Omega$ represents her *ex-ante* preference over exactly the same set of lotteries.

Regardless of the interpretation, the goal is to provide a theory that connects the class of conditional preferences through the Bayes updating. Formally, our goal is to find the condition that there exists a state-dependent utility function $u : \Omega \times M \to \mathbb{R}$ and a subjective probability measure $P : \Omega \to \mathbb{R}_{++}$, such that for every two lotteries $x, y \in L$, and any event A the following holds:

$$x \succcurlyeq_A y \Leftrightarrow \sum_{\omega \in A} P(\omega|A) E^x[u(\omega, \cdot)] \geqslant \sum_{\omega \in A} P(\omega|A) E^y[u(\omega, \cdot)]. \tag{4.1}$$

In the equation above, $E^x[u(w, \cdot)]$ represents the expected utility of the state-dependent utility u in the state ω and with respect to the lottery x. The right hand side of the equation is comparing the conditional expectation utility of the lottery x and y, with respect to the subjective probability measure P and the state-dependent utility u. The importance of the result is that the probability measure P depends on the event A through the Bayes rule.

4.2.2 Main Axioms and Representation

In order to get equation 4.1, we define the following axioms.

Axiom 4.2.1. ***(Weak Order)*** For any event A, the conditional preference $\succcurlyeq_A$ is complete and transitive.

Axiom 4.2.2. ***(vN-M Continuity)*** For any event A and for every $x, y, z \in L$, if $x \succcurlyeq_A y \succcurlyeq_A z$, there exist $\alpha, \beta \in (0,1)$ such that

$$\alpha x + (1-\alpha)z \succcurlyeq_A y \succcurlyeq_A \beta x + (1-\beta)z.$$

Axiom 4.2.3. ***(Independence)*** For any event A, every $x, y, z \in L$, and every $\alpha \in (0,1)$,

$$x \succcurlyeq_A y \Rightarrow \alpha x + (1-\alpha)z \succcurlyeq_A \alpha y + (1-\alpha)z.$$

The first three axioms are the standard vN-M axioms. Together, they imply that there is a representation of each conditional preference $\succcurlyeq_A$ in the form of an affine utility function. However, the affine utility function in the representation depends on the event A. The next axiom is the main connection of the class of conditional preferences.

Axiom 4.2.4. ***(extended Pareto)*** For any two disjoint events A, B, and for every $x, y \in L$,

$$x \succcurlyeq_A y, \; x \succcurlyeq_B y \Rightarrow x \succcurlyeq_{A \cup B} y \tag{4.2}$$

$$x \succ_A y, \; x \succcurlyeq_B y \Rightarrow x \succ_{A \cup B} y \tag{4.3}$$

This is the main axiom of the previous section. The role is the same as before, to get the Bayesian form of aggregation. Theorem 8 connects the extended Pareto (extended weak Pareto) to the strict weighted averaging axiom (weighted averaging axiom).

We add two more axioms, similar to the richness and minimal agreement conditions of Section 3.2.

Axiom 4.2.5. ***(Minimal Agreement)*** There exist two lotteries, $\overline{x}, \underline{x} \in L$, such that for every $\omega \in \Omega$, $\overline{x} \succ_\omega \underline{x}$.

In the axiom, the two lotteries $\overline{x}, \underline{x}$ are not necessarily the best and worst lotteries. The decision maker, regardless of the realization of the states, always prefers the lottery $\overline{x}$ to the lottery $\underline{\text{x}}$.

The last axiom is the richness condition of Section 2.3.

Axiom 4.2.6. ***(Richness)*** There exist three states $\omega_1, \omega_2, \omega_3 \in \Omega$ such that for any $\omega \in \{\omega_1, \omega_2, \omega_3\}$, there exist two lotteries, $x, y \in L$, where $x \succ_\omega y$ and $y \succcurlyeq_{\omega'} x$ for $\omega' \in \{\omega_1, \omega_2, \omega_3\} \backslash \{\omega\}$.

The role of the axiom is similar to the role of the richness condition in Section 2.3. The axiom implies that there exists a representation of the class of conditional preferences with affine functions having a dimension of greater than two. It is another way of forcing the conditional preferences to satisfy the definition of the $\mathcal{R}_X$ in Section 3.2.3.

By considering these six axioms, we can rationalize the behavior of the decision maker as a subjective expected utility maximizer with a state-dependent utility.

Theorem 10. *Suppose that the decision maker's conditional preferences satisfy axioms 6.1-6.6, then there exist a function $u : \Omega \times M \to \mathbb{R}$ and a probability measure $P : \Omega \to \mathbb{R}_{++}$, such that for every two lotteries $x, y \in L$, and any event A, the following holds:*

$$x \succcurlyeq_A y \Leftrightarrow \sum_{\omega \in A} P(\omega|A) E^x[u(\omega, \cdot)] \geqslant \sum_{\omega \in A} P(\omega|A) E^y[u(\omega, \cdot)]. \tag{4.4}$$

The proof is similar to the proof of corollary 8.

It is important to note that the probability measure P is not unique. Let $Q : \Omega \to \mathbb{R}_{++}$ be any probability measure on Ω; by defining a state-dependent utility $w(\omega, x) = \frac{u(\omega,x)}{Q(\omega)}$, equation 4.4 continues to hold with Q and w. However, if we change the minimal agreement axiom to a stronger version, we attain the uniqueness. In the stronger version of the minimal agreement, we assume that the decision maker's preferences over the lotteries $\overline{x}, \underline{x}$ is indifferent to the realization of the states. Formally:

Axiom 4.2.7. ***(Strong Minimal Agreement)*** There exist two lotteries $\overline{x}, \underline{x} \in L$ such that for every $\omega \in \Omega$, $\overline{x} \succ_\omega \underline{x}$. Moreover, $(\overline{x}, \{\omega_1\}, O, \Omega\backslash\{\omega_1\}) \sim (\overline{x}, \{\omega_2\}, O, \Omega\backslash\{\omega_2\})$ and $(\underline{x}, \{\omega_1\}, O, \Omega\backslash\{\omega_1\}) \sim (\underline{x}, \{\omega_2\}, O, \Omega\backslash\{\omega_2\})$ for all $\omega_1, \omega_2 \in \Omega$.

Conceptually, this axiom is closely related to A.0 axiom by Karni and D. Schmeidler (2007). However, unlike Karni's axiom, we do not need these two lotteries to be the best and worst lotteries in the set of lotteries. Our model only needs two lotteries with one strictly preferred to the other, regardless of states. Moreover, the decision maker's conditional preference for each of them is independent of the set of states.

By replacing the minimal agreement axiom with the strong minimal agreement axiom, we can "uniquely" separate the belief from the state-dependent preference.

Theorem 11. *Suppose that the decision maker's conditional preferences satisfy axioms 6.1-6.7, then there exist a function $u : \Omega \times M \to \mathbb{R}$ and a probability measure $P : \Omega \to \mathbb{R}_{++}$, such that for every two lotteries $x, y \in L$, and every event A, the following holds:*

$$x \succcurlyeq_A y \Leftrightarrow \sum_{\omega \in A} P(\omega|A) E^x[u(\omega, \cdot)] \geqslant \sum_{\omega \in A} P(\omega|A) E^y[u(\omega, \cdot)]. \tag{4.5}$$

Moreover, the probability measure P is unique and the function u is unique up to affine transformations.

Proof. Based on Theorem 10, there exists a pair (P, u) satisfying equation 4.5.

To prove the uniqueness, we assume that (P_1, u_1) and (P_2, u_2) both represent the same class of conditional preferences. By considering the conditional preference $\succcurlyeq_\omega$ and the vN-M Theorem, we know that $u_2(\omega, .) = \alpha_\omega u_1(\omega, .) + \beta_\omega$. By using the strong minimal agreement axiom, we have $u_1(\omega_1, \overline{x}) = u_1(\omega_2, \overline{x})$, $u_2(\omega_1, \overline{x}) = u_2(\omega_2, \overline{x})$, $u_1(\omega_1, \underline{x}) = u_1(\omega_2, \underline{x})$, and $u_2(\omega_1, \underline{x}) = u_2(\omega_2, \underline{x})$ for any two states $\omega_1, \omega_2 \in \Omega$. Therefore, $\alpha_{\omega_1} = \alpha_{\omega_2}$ and $\beta_{\omega_1} = \beta_{\omega_2}$ for all $\omega_1, \omega_2 \in \Omega$. Hence, $u_2(\omega, .) = \alpha u_1(\omega, .) + \beta$ for all $\omega \in \Omega$.

We consider an event A. Both (P_1, u_1) and (P_2, u_2) represent the conditional preference $\succcurlyeq_A$. Considering the pair (P_2, u_2), $\succcurlyeq_A$ has the representation $\sum_{\omega \in A} P_2(\omega|A) E^{(.)}[u_2(\omega, .)] = \sum_{\omega \in A} P_2(\omega|A) E^{(.)}[\alpha u_1(\omega, .) + \beta] = \alpha \sum_{\omega \in A} P_2(\omega|A) E^{(.)}[u_1(\omega, .)] + \beta$.

Since, α is strictly positive, the last representation is the same as $\sum_{\omega \in A} P_2(\omega|A) E^{(.)}[u_1(\omega, .)]$. However, using the other pair, (P_1, u_1), we get the representation $\sum_{\omega \in A} P_1(\omega|A) E^{(.)}[u_1(\omega, .)]$.

Therefore, for any event A, $\sum_{\omega \in A} P_2(\omega|A)E^{(.)}[u_1(\omega,.)]$ and $\sum_{\omega \in A} P_1(\omega|A)E^{(.)}[u_1(\omega,.)]$ both represent the conditional preference $\succcurlyeq_A$.

Using the richness axiom, strong minimal agreement condition, and uniqueness of corollary 8, we have $P_1 = P_2$. This completes the proof.

$\square$

The result shows that adding the strong minimal agreement allows us to uniquely define the belief and the state-dependent utility.

4.3 Related Literature

There are many papers and different approaches to address the shortcomings of subjective expected utility theory. Note that our goal, in this context, is to explain the basic underlying structure that let us separate beliefs and state-dependent utilities.

Karni, D. Schmeidler, and Vind (1983) and Karni and D. Schmeidler (2016) use hypothetical preferences on hypothetical lotteries to obtain the identification of the beliefs and state-dependent preferences. Dereze (1987), Dereze and Rustichini (1999), and Karni (2006) present different theories to identify state-dependent preferences in situations where moral hazard is present.

Luce and Krantz (1971) and Fishburn (1973) use preferences on enlarged choice space of all conditional acts to model subjective expected utility of state-dependent preferences. However, our paper only considers the hypothetical conditional preferences on the set of conditional "constant" acts. We find the necessary and sufficient condition that our conditional preferences are related to each other through a subjective probability and a state-dependent utility.

The papers by Skiadas (1997a), Ghirardato (2002), and Karni and D. Schmeidler (2007) are closely (conceptually) related to our main result of Section 4.2. However, there are many differences between each result. Moreover, our goal is to build the model that only extended Pareto derives the separation of beliefs and state-dependent preferences.

Skiadas (1997a) presents a nonexpected utility model, by considering hypothetical preferences over the set of act-event pairs. His coherence axiom has the same role as the extended Pareto axiom in our setup. However, he used the solvability axiom to be able to apply the Debreu's additive representation theorem. In our paper, we consider the class of conditional vN-M preferences. As a result, we only require the extended Pareto for our representation.

Karni and D. Schmeidler (2007) presents a general model with a preference ordering over the set of unconditional acts. Using the preference order, he defines the set of conditional preferences over the set of all conditional acts. Therefore, to connect the class of conditional preferences, the model needs the existence of the constant-valuation acts. Moreover, the cardinal and ordinal coherence axioms are the main forces behind obtaining the Bayesian updating in his representation. However, in our more restricted domain, we only need the extended Pareto to get our representation.

Finally, Ghirardato (2002), by replacing Savage's sure thing principle by dynamic consistency, obtains a subjective expected utility theory that the conditional preferences are connected through the Bayes rule. However, his representation only holds for the state-independent preferences.

BIBLIOGRAPHY

Ahn, David. S., F. Echenique, and K. Saito (2018). "On Path Independent Stochastic Choice". In: *Theoretical Economics* 13, pp. 61–85.

Aizerman, M. A. and A. V. Malishevski (1981). "Some Aspects of the General Theory of Best Option Choice". In: *Automation and Remote Control* 42, pp. 65–83.

Anscombe, F. J. and R. J. Aumann (1963). "A Definition of Subjective Probability". In: *The Annals of Mathematics and Statistics* 34, pp. 199–205.

Arrow, K. J. (1963). *Social Choice and Individual Values.* 12. Yale Univ. Press.

– (1974). "Optimal Insurance and Generalized Deductibles". In: *Scandinavian Actuarial Journal* 1, pp. 1–42.

Baucells, M. and L. S. Shapley (2008). "Multiperson utility". In: *Games and Economic Behavior* 62, pp. 329–347.

Billot, A. et al. (2005). "Probabilities as Similarity-Weighted Frequencies". In: *Econometrica* 73, pp. 1125–1136.

Borgers, T. and Y. Choo (2017a). "A Counterexample to Dhillon". In: *Social Choice Welfare.* 48, pp. 837–843.

– (2017b). "Revealed Relative Utilitarianism". Working Paper.

Cerreia-Vioglio, S. et al. (2018). "Multinomial logit processes and preference discovery: inside and outside the black box". Working Paper.

Chateauneuf, A. and J. U. Jaffray (1989). "Some Characterizations of Lower Probabilities and other Monotone Capacities Through the Use of Möbius Inversion". In: *Mathematical Social Sciences* 17, pp. 263–283.

Cook, P. J. and D. A. Graham (1997). "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities". In: *Quarterly Journal of Economics* 91, pp. 143–156.

Dekel, E., B. Lipman, and A. Rustichini (2001). "Representing preferences with a unique subjective state space". In: *Econometrica* 69, pp. 891–934.

Dekel, E., B. Lipman, A. Rustichini, and T. Sarver (2007). "Representing preferences with a unique subjective state space: a corrigendum". In: *Econometrica* 75, pp. 591–600.

Dereze, J. (1987). "Decision Theory with Moral Hazard and State-dependent Preferences". In: *Essays on Economic Decisions Under Uncertainty.* Cambridge University Press.

Dereze, J. and A. Rustichini (1999). "Moral Hazard and Conditional Preferences". In: *Journal of Mathematical Economics* 31, pp. 159–181.

Dhillon, A. (1998). "Extended Pareto Rules and Relative Utilitarianism". In: *Social Choice Welfare* 15, pp. 521–542.

Dhillon, A. and J. F. Mertens (1997). "An Impossibility Theorem with von Neumann-Morgenstern Preferences". In: *Economics Letters* 56, pp. 305–309.

– (1999). "Relative Utilitarianism". In: *Econometrica* 67, pp. 471–498.

Dilworth, R. P. (1950). "A Decomposition Theorem for Partially ordered sets". In: *Annals of Mathematics* 51, pp. 161–166.

Echenique, F. and K. Saito (2018). "General Luce Model". In: *Theoretical Economics* 13, pp. 61–85.

Edelman, P. H. and R. E. Jamison (1985). "The Theory of Convex Geometries". In: *Geometriae Dedicata* 19, pp. 247–270.

Fishburn, P. C. (1973). "A Mixture-Set Axiomatization of Conditional Subjective Expected Utility". In: *Econometrica* 41, pp. 1–25.

Ghirardato, P. (2002). "Revisiting Savage in a Conditional World". In: *Economic Theory* 20, pp. 83–92.

Gilboa, I. and D. Schmeidler (1995). "Case-Based Decision Theory". In: *Quarterly Journal of Economics* 110, pp. 605–639.

– (2012). *Case-Based Predictions: An Axiomatic Approach to Prediction, Classification and Statistical Learning.* 1st ed. World Scientific Publishing Co, Singapore.

Gorno, L. (2016). "Additive Representation for Preferences Over Menus in Finite Choice Setting". In: *Journal of Mathematical Economics* 65, pp. 41–47.

Gul, F. and W. Pesendorfer (2001). "Temptation and Self-Control". In: *Econometrica* 69, pp. 1403–1435.

Harsanyi, J. C. (1955). "Cardinal Welfare, Individual Ethics, and Interpersonal Comparisons of Utility". In: *Journal of Political Economy* 63, pp. 309–321.

Hylland, A. (1980). "Aggregation Procedure for Cardinal Preferences: A Comments". In: *Econometrica* 48, pp. 539–542.

Kalai, E. and N. Megiddo (1980). "Path Independent Choices". In: *Econometrica* 48, pp. 781–784.

Kalai, E. and N. Schmeidler (1977). "Aggregation Procedure for Cardinal Preferences: A Formulation and Proof of Samuelson's Impossibility Conjecture". In: *Econometrica* 45, pp. 1431–1438.

Karni, E. (1985). *Decision making under uncertainty: the case of state-dependent preferences.* Cambridge: Harvard University Press.

Karni, E. (2006). “Subjective Expected Utility Theory Without States of the World”. In: *Journal of Mathematical Economics* 42, pp. 325–342.

Karni, E. and D. Schmeidler (2007). “Foundations of Bayesian theory”. In: *Journal of Economic Theory* 132, pp. 167–188.

– (2016). “An Expected Utility Theory for State-Dependent Preferences”. In: *Theory and Decision* 81, pp. 467–478.

Karni, E., D. Schmeidler, and K. Vind (1983). “On State Dependent Preferences and Subjective Probabilities”. In: *Econometrica* 51, pp. 1021–1031.

Kopylov, I. (2009). “Finite Additive Utility Representations for Preferences Over Menus”. In: *Journal of Economic Theory* 144, pp. 354–374.

– (2018). “Combinatorial Models of Subjective States”. Working Paper.

Koshevoy, G. A. (1999). “Choice Functions and Abstract Convex Geometries”. In: *Mathematical Social Sciences* 38, pp. 35–44.

Kreps, D. (1979). “A Representation Theorem for ‘Preference for Flexibility’ ”. In: *Econometrica* 47, pp. 565–578.

Luce, R. D. (1959). *Individual Choice Behavior a Theoretical Analysis*. John Wiley and Sons.

Luce, R. D. and D. H. Krantz (1971). “Conditional Expected Utility”. In: *Econometrica* 39, pp. 253–271.

Machina, M. J. and R. P. Parks (1981). “On Path Independent Randomized Choice”. In: *Econometrica* 49, pp. 1345–1347.

May, K. O. (1952). “A set of Independent, Necessary and Sufficient Conditions for Simple Majority Decision”. In: *Econometrica* 20, pp. 680–684.

Nehring, K. (2001). “Preference for Flexibility in a Savage Framework”. In: *Econometrica* 67, pp. 101–120.

Plott, C. R. (1973). “Path Independence, Rationality, and Social Choice”. In: *Econometrica* 41, pp. 1075–1091.

Ramsey, F. (1931). “Truth and Probability”. In: *The Foundations of Mathematics and Other Logical Essays*. Routledge and Kegan Paul Ltd.

Richter, M. and A. Rubinstein (2018). “Back to Fundamentals: Convex Geometry and Economic Equilibrium”. Working Paper.

Savage, L. J. (1954). *The Foundations of Statistics*. John Wiley and Sons.

Sen, A. K. (1977). “Social Choice Theory: A Re-Examination”. In: *Econometrica* 45, pp. 53–89.

Shapley, L. S. and M. Shubik (1982). “Preferences and Utility”. In: *Game Theory in the Social Sciences: Concepts and Solutions*. Ed. by M. Shubik. 3rd ed. Cambridge, MA: MIT Press. Chap. 4.

Shmaya, E. and L. Yariv (2007). “Foundations for Bayesian Updating”. Working Paper, Caltech.

Skiadas, C. (1997a). “Conditioning and Aggregation of Preferences”. In: *Econometrica* 65, pp. 347–367.

– (1997b). “Subjective Probability under Additive Aggregation of Conditional Preferences”. In: *Journal of Economic Theory* 76, pp. 242–271.

Von-Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Weymark, J. A. (1991). “A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism”. In: *Interpersonal Comparisons of Well-Being*. Ed. by Jon Elster and John E. Roemer. Cambridge University Press, pp. 255–320.

Appendix A

APPENDIX TO CHAPTER 1

A.1 Möbius Operator

The main technical tool for section 1.5 is the Möbius inversion formula [1]. Let $(X, \succsim)$ be a finite partial ordered set. We define the *Möbius* function $\mu : X \times X \to \mathbb{R}$ as follows:

We set $\mu(x, y) = 0$ whenever $y \not\succsim x$, and $\mu(x, x) = 1$ for all $x \in X$. Then, we define other values inductively as follows:

$$\mu(x, y) = -\sum_{y \succ z \succsim x} \mu(x, z).$$

By above definition, we have:

$$\sum_{y \succsim z \succsim x} \mu(x, z) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases}$$

Let $(X)^{\mathbb{R}}$ denote the set of all functions from X to $\mathbb{R}$. Then, the ***Möbius operator*** $\Phi : (X)^{\mathbb{R}} \to (X)^{\mathbb{R}}$ is defined by $\Phi(f)(x) = \sum_{x \succsim y} f(y)$. Möbius inversion formula guarantees that the Möbius operator is bijective and the inverse is $\Phi^{-1}(g)(x) = \sum_{x \succsim y} \mu(x, y) g(y)$.

Theorem 12. *(Möbius inversion formula) Let $(X, \succsim)$ be a finite partial ordered set and μ be its Möbius function. Let $f, g : X \to \mathbb{R}$. Then*

$$g(x) = \sum_{x \succsim y} f(y)$$

implies that

$$f(x) = \sum_{x \succsim y} \mu(y, x) g(y).$$

[1] For an application of the technique in Kreps' setting, check Nehring (2001). For a complete study of the concept, see Chateauneuf and Jaffray (1989).

Proof. To be complete, we add the proof.

$$
\begin{aligned}
\sum_{x \gtrsim y} \mu(y,x) g(y) &= \sum_{x \gtrsim y} \mu(y,x) \big(\sum_{y \gtrsim z} f(z) \big) \\
&= \sum_{x \gtrsim y \gtrsim z} \mu(y,x) f(z) \\
&= \sum_{x \gtrsim z} \big(\sum_{x \gtrsim y \gtrsim z} \mu(y,x) \big) f(z) = f(x).
\end{aligned}
$$

$\square$

Appendix B

APPENDIX TO CHAPTER 2

B.1 Proof of Theorem 4

The following two lemmas are the central ideas behind the proof. They help us to first define the function w and then extend it from the binary sets to any finite-cardinality sets.

Lemma 5. *Select X as any nonempty set. Let X^* denote the set of all nonempty finite subsets of X. Consider two functions $f_1, f_2 : X^* \to \mathbb{R}^n$ that satisfy the strict weighted averaging axiom. Select four points a, b, c, d in the space X^* such that $a \cup b = c \cup d$ and $a \cap b = c \cap d = \varnothing$. If $\forall\ x \in \{a, b, c, d\}$ $f_1(x) = f_2(x)$ and not all $\{f_1(a), f_1(b), f_1(c), f_1(d)\}$ are on a same line, then $f_1(a \cup b) = f_2(a \cup b)$.*

Proof. Since f_1 satisfies the strict weighted averaging axiom and $a \cup b = c \cup d$ and $a \cap b = c \cap d = \varnothing$, thus $f_1(a \cup b)$ is on the line connecting $f_1(a), f_1(b)$. Also, since $a \cup b = c \cup d$, $f_1(a \cup b) = f_1(c \cup d)$ should be on the line connecting $f_1(c)$ and $f_1(d)$. But $\{f_1(a), f_1(b), f_1(c), f_1(d)\}$ are not collinear, thus the line connecting $f_1(a)$ and $f_1(b)$ and the line connecting $f_1(c)$ and $f_1(d)$ can only intersect at most at a single point. But $f_1(a \cup b)$ is on the both lines, hence this point must be the unique intersection of them.

Similarly, the same is true for f_2. This means $f_2(a \cup b)$ must be the unique intersection of the line passing through $f_2(a), f_2(b)$ and the line passing through $f_2(c), f_2(d)$. But since $\forall x \in \{a, b, c, d\}$ $f_1(x) = f_2(x)$, $f_2(a \cup b)$ should be the unique intersection of the line passing through $f_1(a), f_1(b)$ and the line passing through $f_1(c), f_1(d)$. But we have already shown that $f_1(a \cup b)$ is also the unique intersection of the line passing through $f_1(a), f_1(b)$ and the line passing through $f_1(c), f_1(d)$. Thus, $f_1(a \cup b) = f_2(a \cup b)$.

□

Lemma 6. *Assume that $\{x, y, z\}$ are three points in X such that $f(x), f(y), f(z)$ are not collinear. Let f satisfy the strict weighted averaging axiom and $f(\{x, y, z\}) = a_1 f(x) + a_2 f(y) + a_3 f(z)$, then a_1/a_2 must be* *independent of the choice of z, as long as* $f(x), f(y), f(z)$ *are not collinear. Moreover, if* $f(\{x,y\}) = \lambda f(x) + (1-\lambda)f(y)$, *then* $\frac{a_1}{a_2} = \frac{\lambda}{1-\lambda}$.

Proof. Since $f(x), f(y), f(z)$ are not collinear, they should be affinely independent. Hence, a_1, a_2, a_3 are uniquely defined.

By the strict weighted averaging axiom, there exists $\lambda_1 \in (0,1)$ such that $f(\{x,y,z\}) = \lambda_1 f(\{x,y\}) + (1-\lambda_1)f(\{z\})$. Again by the strict weighted averaging axiom there exists $\lambda \in (0,1)$ such that $f(\{x,y\}) = \lambda f(x) + (1-\lambda)f(y)$. Hence, $f(\{x,y,z\}) = \lambda_1(\lambda f(x) + (1-\lambda)f(y)) + (1-\lambda_1)f(\{z\})$. By affinely independence of $f(x), f(y), f(z)$, we should have $a_1 = \lambda_1\lambda$ and $a_2 = \lambda_1(1-\lambda)$. This means that $\frac{a_1}{a_2} = \frac{\lambda}{1-\lambda}$, which means that a_1/a_2 is independent of the choice of z, as long as $f(x), f(y), f(z)$ are not collinear. □

B.1.1 Proving the necessary and the uniqueness part

Assume that the weight function w exists. Therefore, $f(A) = \frac{\sum_{x\in A} w(x)f(x)}{\sum_{x\in A} w(x)}$. It shows that if $A \cap B = \varnothing$, then $f(A \cup B) = \frac{\sum_{x\in A\cup B} w(x)f(x)}{\sum_{x\in A\cup B} w(x)} = (\frac{\sum_{x\in A} w(x)}{\sum_{x\in A\cup B} w(x)})(\frac{\sum_{x\in A} w(x)f(x)}{\sum_{x\in A} w(x)}) + (\frac{\sum_{x\in B} w(x)}{\sum_{x\in A\cup B} w(x)})(\frac{\sum_{x\in B} w(x)f(x)}{\sum_{x\in B} w(x)})$. By defining $\lambda = \frac{\sum_{x\in A} w(x)}{\sum_{x\in A\cup B} w(x)}$, we have $f(A \cup B) = \lambda f(A) + (1-\lambda)f(B)$. Thus, the strict weighted averaging axiom satisfied.

Regarding the uniqueness of w, assume that there exist two w_1, w_2 such that $f(A) = \frac{\sum_{x\in A} w_1(x)f(x)}{\sum_{x\in A} w_1(x)} = \frac{\sum_{x\in A} w_2(x)f(x)}{\sum_{x\in A} w_2(x)}$. Since the range of f is not a subset of a line, there exist at least three elements $x, y, z \in X$ such that $f(x), f(y), f(z)$ are not collinear. Thus, they are affinely independent. Hence, $f(\{x,y,z\}) = a_1 f(x) + a_2 f(y) + a_3 f(z)$ has a unique solution a_1, a_2, a_3. Hence, there should be an α such that $w_1(p)/w_2(p) = \alpha\ \forall p \in \{x,y,z\}$. We will show that for all other point $r \in X$, $w_1(r)/w_2(r) = \alpha$.

Select a point $r \in X$, based on the assumption on $\{x,y,z\}$, there should be at least two points u, v in $\{x,y,z\}$ such that $f(r), f(u), f(v)$ are not collinear. Without loss of generality, assume that $\{u,v\} = \{x,y\}$. Since $f(r), f(x), f(y)$ are affinely independent, $f(\{x,y,r\}) = b_1 f(x) + b_2 f(y) + b_3 f(r)$ where b_1, b_2, b_3 are unique. Therefore, there exists β such that $w_1(p)/w_2(p) = \beta\ \forall p \in \{x,y,r\}$. But notice that $\alpha = w_1(x)/w_2(x) = \beta$. Hence, we should have $w_1(r)/w_2(r) = \alpha$ and this is what we wanted to prove.

B.1.2 Proving the sufficiency part

First, in order to define the function w, fix an element $x_0 \in X$ and put $w(x_0) = 1$. Based on the strict weighted averaging axiom for any $y \in X \backslash \{x_0\}$ such that $f(y) \neq f(x_0)$, we have a unique $\lambda \in (0,1)$ such that $f(\{x_0, y\}) = \lambda f(x_0) + (1-\lambda) f(y)$. Let define $w(y) = \frac{1-\lambda}{\lambda}$.

To define the weight for any other $y \in X \backslash \{x_0\}$ with $f(y) = f(x_0)$, we fix another point $z_0 \in X \backslash \{x_0\}$ such that $f(z_0) \neq f(x_0)$. Since $f(x_0) = f(y)$, we should have $f(y) \neq f(z_0)$. By using the strict weighted averaging axiom, we know that there exists a unique $\lambda \in (0,1)$ such that $f(\{z_0, y\}) = \lambda f(z_0) + (1-\lambda) f(y)$. Since the weight on z_0 has already been defined, we define the weight of y such that $\frac{w(y)}{w(z_0)} = \frac{1-\lambda}{\lambda}$. Thus, $w(y) = w(z_0) \times \frac{1-\lambda}{\lambda}$.

In the rest of this section, we are going to prove that w satisfies the representation of the theorem. It means that by defining $f^*(A) = \frac{\sum_{x \in A} w(x) f(x)}{\sum_{x \in A} w(x)}$, we should have $f^*(A) = f(A)$.

First, in Step 1 we prove that the representation holds for any three points, as long as the three points under f are not collinear. In Step 2, we prove that the representation holds for any two points. In Step 3, (which is not necessary, and we provide it for its simplicity to capture the main ideas of the main part) we prove that for three points the representation holds. Finally, in Step 4, by using induction on the cardinality of subsets of X, we show that the representation holds for any subset of X.

Step 1: for any three points r, s, t such that $f(r), f(s), f(t)$ are not collinear, we have $f(\{r, s, t\}) = a_1 f(r) + a_2 f(s) + a_3 f(t)$, where a_i are unique. Note that it is enough to prove that $\frac{a_1}{a_2} = \frac{w(r)}{w(s)}$, because in the same way, we can also get $\frac{a_2}{a_3} = \frac{w(s)}{w(t)}$, $\frac{a_3}{a_1} = \frac{w(t)}{w(r)}$. There are two cases:

Case 1: If x_0, r, s are such $f(x_0), f(r), f(s)$ are not collinear then $f(\{x_0, r, s\}) = b_1 f(x_0) + b_2 f(r) + b_3 f(s)$. Based on Lemma 6, we know that $\frac{a_1}{a_2} = \frac{b_2}{b_3}$. But Again using the Lemma 6 and the way we define w, we know that $\frac{b_1}{b_2} = \frac{1}{w(r)}$, $\frac{b_1}{b_3} = \frac{1}{w(s)}$ which means that $\frac{b_2}{b_3} = \frac{w(r)}{w(s)}$. Hence, we have $\frac{a_1}{a_2} = \frac{w(r)}{w(s)}$.

Case 2: If x_0, r, s are such that $f(x_0), f(r), f(s)$ are collinear, in this case both $\{f(x_0), f(r), f(t)\}$ and $\{f(x_0), f(s), f(t)\}$ are not collinear. By the same technique as the first case, we get that $\frac{a_1}{a_3} = \frac{w(r)}{w(t)}$ and $\frac{a_3}{a_2} = \frac{w(t)}{w(s)}$. Hence, it means that $\frac{a_1}{a_2} = \frac{a_1}{a_3} \times \frac{a_3}{a_2} = \frac{w(r)}{w(t)} \times \frac{w(t)}{w(s)} = \frac{w(r)}{w(s)}$, which is what we wanted to prove.

Step 2: Assume that $r, s \in X$. We want to show that $f^*(\{r, s\}) = f(\{r, s\})$. If $f(r) = f(s)$, then it is true. If $f(r) \neq f(s)$, then by the richness condition, there exists an element $t \in X$ such that $\{f(t), f(r), f(s)\}$ are not collinear. Based on Step 1, we know that $f(\{t, r, s\}) = f^*(\{t, r, s\})$, also we have $f(t) = f^*(t), f(r) = f^*(r)$, and $f(s) = f^*(s)$. Notice that, based on the strict weighted averaging axiom, $f(\{r, s\})$ is on the line connecting $f(r)$ and $f(s)$. Also, it is on the line connecting $f(\{t, r, s\})$ and $f(t)$. The reason is that by the strict combination axiom, there exist a $\lambda \in (0, 1)$ such that $f(\{t, r, s\}) = \lambda f(t) + (1 - \lambda) f(\{r, s\})$, which means that $f(\{r, s\})$ is on the line connecting $f(\{t, r, s\})$ and $f(t)$. Similarly, everything holds for f^* which means that $f^*(\{r, s\})$ is on the line connecting $f^*(r)$ and $f^*(s)$ and also it is on the line connecting $f^*(\{t, r, s\})$ and $f^*(t)$. Since $\{f(t), f(r), f(s)\}$ are not collinear the intersection of two line can have at most one solution and since $f(\{t, r, s\}) = f^*(\{t, r, s\})$, $f(t) = f^*(t)$, $f(r) = f^*(r)$, and $f(s) = f^*(s)$ then by a similar argument as Lemma 5, we should have a unique intersection, which satisfies $f^*(\{r, s\}) = f(\{r, s\})$. This is what we wanted to prove.

Step 3: (This part is the tricky part, and we provide it to capture the main ideas. We will use the same technique in Step 4) We are going to prove that for all three point r, s, t we have $f^*(\{r, s, t\}) = f(\{r, s, t\})$.There are two separate cases to be considered.

Case 1: If $f(r), f(s), f(t)$ are not collinear, then by Step 1, it is correct.

Case 2: Assume that $f(r), f(s), f(t)$ are collinear. If all of them are the same, then by strict weighted averaging axiom $f^*(\{r, s, t\}) = f(\{r, s, t\})$. Hence, assume that they are not all the same.

Without loss of the generality, assume that $f(s) \neq f(r), f(s) \neq f(t)$. Based on the richness condition of f, we should have a point $v \in X$ such that not all $f(v), f(r), f(s)$, and $f(t)$ are collinear. Note that, $f(v), f(r), f(s)$ are not collinear. Similarly, $f(v), f(s), f(t)$ are not collinear. Based on Case 1, we know that $f^*(\{v, r, s\}) = f(\{v, r, s\})$ and $f^*(\{v, s, t\}) = f(\{v, s, t\})$. Also, we know that $f^*(v) = f(v)$, $f^*(r) = f(r)$, $f^*(s) = f(s)$, and $f^*(t) = f(t)$. Using the strict weighted averaging axiom, we know that $f(\{v, r, s, t\})$ is on the intersection of the line passing through $f(\{v,r,s\})$ and $f(t)$, and the line passing through the $f(\{v,s,t\})$ and $f(r)$. Also, note that not all of $f(\{v,r,s\})$, $f(\{v,s,t\})$, $f(r)$, and $f(t)$ are collinear, since otherwise $f(v)$ must be on the line connecting $f(r)$ and $f(s)$. Similarly, we have the same properties for $f^*(\{v,r,s,t\})$. Based on the argument of the Lemma 5, we have $f(\{v,r,s,t\}) = f^*(\{v,r,s,t\})$.

By using the strict weighted averaging axiom, we know that $f(\{r,s,t\})$ is on the line passing through $f(\{v,r,s,t\})$ and $f(v)$, since there exists $\lambda \in (0,1)$ such that $f(\{v,r,s,t\}) = \lambda f(\{r,s,t\}) + (1-\lambda) f(v)$. Again, by using the strict weighted averaging axiom, we know that $f(\{r,s,t\})$ is on the line passing through $f(\{r,s\})$ and $f(t)$. Also, the same holds for f^*. Moreover, we have $f(\{v,r,s,t\}) = f^*(\{v,r,s,t\})$, $f(\{r,s\}) = f^*(\{r,s\})$, $f(v) = f^*(v)$, and $f(t) = f^*(t)$. Also, not all $f(\{v,r,s,t\})$, $f(\{r,s\})$, $f(v)$, and $f(t)$ are on a same line, since otherwise $f(v)$, $f(r)$, $f(s)$, and $f(t)$ are collinear which is not correct. As a result, based on the argument of lemma 5, we have $f^*(\{r,s,t\}) = f(\{r,s,t\})$. The latter is what we wanted to prove.

Step 4 (The main Step): Up to here, we prove that for any $A \in X^*$ if $|A| \leqslant 3$ then $f^*(A) = f(A)$. To complete the proof, we will use an induction on the cardinality of A. Assume that for all $A \in X^*$ with $|A| \leqslant k$ we have $f^*(A) = f(A)$. We are going to show that for all $A \in X^*$ with $|A| = k+1$, we have $f^*(A) = f(A)$.

Fix a subset A with $|A| = k+1$. Assume that $A = \{x_1, \ldots, x_{k+1}\}$. There are two separate cases to be considered.

Case 1: Assume that not all $\{f(x_i)\}_{i=1}^{k+1}$ are collinear. Note that, by the induction hypothesis, $\forall x \in A$ and $\forall\ B \in 2^{A\backslash\{x\}}$ we have $f(B) = f^*(B)$. Define $line(f(x), f(A\backslash\{x\}))$ as the line passing through $f(x)$ and $f(A\backslash\{x\})$ for the case where $f(x) \neq f(A\backslash\{x\})$. However, if $f(x) = f(A\backslash\{x\})$, then define it as the single point $f(x)$.

If there exists $x \in A$ such that $f(x) = f(A\backslash\{x\})$, then based on the strict weighted averaging axiom , there exists $\lambda \in (0,1)$ such that $f(A) = \lambda f(x) + (1-\lambda) f(A\backslash\{x\}) = f(x)$. Similarly, $f^*(A) = f^*(x)$. But, we know that $f(x) = f^*(x)$, which means that $f(A) = f^*(A)$.

If $\forall\ x \in A\ f(x) \neq f(A\backslash\{x\})$, then there exist $x, y \in A$ such that

$f(x), f(A\backslash\{x\}), f(y), f(A\backslash\{y\})$ are not collinear. Beacuse otherwise all $f(x_i)$ are on the $f(x), f(A\backslash\{x\})$, which cannot be correct since we assumed that not all $\{f(x_i)\}_{i=1}^{k+1}$ are collinear. Considering $x, y \in A$ such that $f(x), f(A\backslash\{x\}), f(y), f((A\backslash\{y\})$ are not collinear, based on the strict weighted averaging axiom we know that $f(A)$ is on $line(f(x), f(A\backslash\{x\}))$. Also, it must be on $line(f(y), f(A\backslash\{y\}))$. Similarly, by the strict weighted averaging axiom when applied to f^*, we know that $f^*(A)$ is on $line(f^*(x), f^*(A\backslash\{x\}))$. Moreover, it must be on the $line(f^*(y), f^*(A\backslash\{y\}))$.

Since (1) $f(x) = f^*(x)$, $f(A\backslash\{x\}) = f^*(A\backslash\{x\})$, $P(y) = P^*(y)$, $P(A\backslash\{y\}) = P^*(A\backslash\{y\})$ and, (2) not all $f(x), f(A\backslash\{x\}), f(y)$, and $f((A\backslash\{y\})$ are collinear, based on the Lemma 5, we have $f^*(A) = f(A)$. Hence in the case that not all $\{f(x_i)\}_{i=1}^{k+1}$ are collinear, we showed that $f^*(A) = f(A)$.

Case 2: Assume that $\{f(x_i)\}_{i=1}^{k+1}$ are collinear. Without loss of generality, assume that $f(x_1), f(x_{k+1})$ are the two extreme points on the line that contains them, which means that all other points are between these two.

If $f(x_1) = f(x_{k+1})$,then all $\{f(x_i)\}_{i=1}^{k+1}$ are the same. Using the strict weighted averaging axiom, it shows that $f(A) = f(x_1) = f^*(x_1) = f^*(A)$.

If $f(x_1) \neq f(x_{k+1})$, based on the richness condition of the aggregation rule f, we can select a point $y \in X\backslash A$ such that not all $f(y), f(x_1)$, and $f(x_{k+1})$ are collinear. Based on the previous Case 1, we know that $f(y, x_1, \ldots, x_k) = f^*(y, x_1, \ldots, x_k)$, since we have proved that f and f^* are coincided for any $k + 1$ not collinear points. Similarly, we have $f(y, x_2, \ldots, x_{k+1}) = f^*(y, x_2, \ldots, x_{k+1})$.

Using the strict weighted averaging axiom, $f(\{y, x_1, \ldots, x_{k+1}\})$ is on the $line(f(\{y, x_1, \ldots, x_k\}), f(x_{k+1}))$. It is also on the $line(f(\{y, x_2, \ldots, x_{k+1}\}), f(x_1))$. Also, not all $f(\{y, x_1, \ldots, x_k\})$, $f(x_{k+1}))$, $f(\{y, x_2, \ldots, x_{k+1}\})$, $f(x_1))$ are collinear, since $f(\{y, x_1, \ldots, x_k\})$ cannot be on $line(f(x_1), f(x_{k+1}))$ otherwise $f(y)$ must be on that line which is not correct. Similarly, everything holds for the f^*.

Since, $f(\{y, x_1, \ldots, x_k\}) = f^*(\{y, x_1, \ldots, x_k\})$, $f(x_{k+1}) = f^*(x_{k+1})$, $f(\{y, x_2, \ldots, x_{k+1}\}) = f^*(\{y, x_2, \ldots, x_{k+1}\})$, and $f(x_1) = f^*(x_1)$ then again by using Lemma 5, we get $f(\{y, x_1, \ldots, x_{k+1}\}) = f^*(\{y, x_1, \ldots, x_{k+1}\})$.

The point $f(\{x_1, \ldots, x_{k+1}\})$ is on the $line(f(x_1), f(x_k))$. It is also on the

$line(f(y), f(\{y, x_1, \ldots, x_{k+1}\}))$, since by the strict weighted averaging axiom $f(\{y, x_1, \ldots, x_{k+1}\}) = \lambda f(y) + (1-\lambda) f(\{x_1, \ldots, x_{k+1}\})$ for some $\lambda \in (0,1)$. Similarly, the same holds for f^*. Finally, since (1) $f(y) = f^*(y)$, $f(\{y, x_1, \ldots, x_{k+1}\}) = f^*(\{y, x_1, \ldots, x_{k+1}\})$, $f(x_1) = f^*(x_1)$ and, (2) $f(x_{k+1}) = f^*(x_{k+1})$ and $f(x_1), f(x_{k+1})$, and $f(y)$ are not collinear, then using the same types of arguments in Lemma 5, we get $f(A) = f^*(A)$. This is what we wanted to prove.

Hence, for all $A \in X^*$ with cardinality $k+1$, we have $f(A) = f^*(A)$. Based on the induction, we have $f(A) = f^*(A)$ for all $A \in X^*$. This completes the proof.

□

B.2 Proof of Theorem 5

There are couple of steps in the proof. **Defining the weak order:**

Step 1: First, we define a binary relation $\succcurlyeq$ over every two different elements $x, y \in X$ by:

Case 1: If $f(x) \neq f(y)$, we define $x \succcurlyeq y \iff f(\{x, y\}) \neq f(y)$.

Case 2: If $f(x) = f(y)$, then by the strong richness condition, we select another point $z \in X$, such that $f(\{x, z\}) \notin \{f(x), f(z)\}$. Hence, we have $f(z) \neq f(y)$. In this case, we define $x \succcurlyeq y \iff f(\{z, y\}) \neq f(y)$.

To obtain reflexivity, for any $x \in X$, we define $x \succcurlyeq x$.

Step 2: We prove that $\succcurlyeq$ is a weak order. The reflexivity and the completeness are trivial. We only need to establish the transitivity. Assume that $x \succcurlyeq y, y \succcurlyeq z$. We will show that $x \succcurlyeq z$.

The proof is by contradiction. Therefore, assume that $z \succ x$.

Case 1: Assume that $f(x), f(y), f(z)$ are non-collinear. Since $z \succ x$, based on the way we defined $\succcurlyeq$, we have $f(x, z) = f(z)$.

Consider the coalition $\{x, y, z\}$. By using the weighted averaging axiom over the sub-coalitions $\{x, z\}$ and $\{y\}$, the vector $f(\{x, y, z\})$ should be on the line joining $f(y)$ and $f(\{x, z\})$ (which is the same as $f(z)$). Similarly, by considering the sub-coalitions $\{x, y\}$ and $\{z\}$, $f(\{x, y, z\})$ should be on the line passing through $f(\{x, y\})$ and $f(z)$. Since $f(\{x, y\}) \neq f(y)$ and $f(\{x, y\}), f(y), f(z)$

are non-collinear, we have $f(\{x,y,z\}) = f(z)$. However, by considering the sub-coalitions $\{y,z\}$, $\{x\}$ and the fact that $f(\{y,z\}) \neq f(z)$, this cannot be happen. Therefore, $x \succcurlyeq z$.

Case 2: Assume that $f(x), f(y), f(z)$ are collinear. By using the strong richness condition, we can select a point $u \in X$ such that $f(u)$ is not on the line passing through $f(x), f(y), f(z)$, and also $f(\{u,x\}) \notin \{f(x), f(u)\}$ (this means that $x \sim u$). First, using Case 1, by considering the coalitions $\{u,x,y\}$, $\{u,x,z\}$, we have $u \succcurlyeq y$ and $z \succ u$. Since $\{u,y,z\}$ are non-collinear, by using case 1, we have $z \succ u, u \succcurlyeq y \Rightarrow z \succ y$. But this is a contradiction. Therefore, $x \succcurlyeq z$.

The main part: proving $f(A) = f(M(A, \succcurlyeq))$.

Up to here, we show that $\succcurlyeq$ is a weak order. Next, we will show that for any coalition $A \in X^*$ we have $f(A) = f(M(A, \succcurlyeq))$.

We use the letter H for the highest-ordered elements of A, and L for the rest. In other words, $H := M(A, \succcurlyeq),\ L := A \backslash H$. The proof is by a double-induction on the cardinality of H and L. In Step 1, we will show that if $x \in X$ and $L \in X^*$ are such that $\forall z \in L:\ x \succ y$, then we should have $f(\{x\} \cup L) = f(x)$.

In Step 2, we show that for a given coalition $H \in X^*$, where all elements of H are in the same equivalence class, and for all $L \in X^*$, if for all $x \in H,\ y \in L : x \succ y$, then we have $f(H \cup L) = f(H)$. Using these two steps, we will finish the proof.

Step 1: Fix an element $x \in X$. By induction on the cardinality of L, where $\forall y \in L,\ x \succ y$, we prove that $f(\{x\} \cup L) = f(x)$.

We have already proved the case where $|L| = 1$. Assume that for all $|L| \leqslant k$ the result is correct. We will show that for all L with $|L| = k+1$, the result is also correct.

Fix a coalition L with $|A| = k+1$ and such that $\forall y \in L,\ x \succ y$. Assume that $A = \{y_1, \ldots, y_{k+1}\}$.

If for all $y \in L : f(y) = f(x)$, then using the weighted averaging axiom $f(\{x\} \cup L) = f(x)$, which is what we wanted to prove. Similarly, if $f(L) = f(x)$, we have $f(\{x\} \cup L) = f(x)$.

Therefore, consider the case that not all of them are the same and $f(L) \neq f(x)$.

Using our definition of the *line* in the proof of Theorem 4, for each $y \in L$ we consider $line(f(\{x\} \cup (L \backslash \{y\})), f(y))$. Using the weighted averaging axiom, for all $y \in L$, $f(\{x\} \cup L) \in line(f(\{x\} \cup L \backslash \{y\}), f(y))$. By using the induction hypothesis $f(\{x\} \cup L \backslash \{y\}) = f(x)$. Therefore, $\forall y \in L$, $f(\{x\} \cup L) \in line(f(x), f(y))$.

Similar to the proof of Theorem 4, we consider two separate cases.

Case 1: Consider the case where there exist two elements $y_1, y_2 \in L$ such that $f(x), f(y_1), f(y_2)$ are non-collinear. We use the same technique as in the proof of Theorem 4.

We know that $f(\{x\} \cup L) \in line(f(x), f(y_1))$ as well as $f(\{x\} \cup L) \in line(f(x), f(y_2))$. Moreover, we know that $f(x), f(y_1), f(y_2)$ are non-collinear. Therefore, the intersection of the two lines should be $f(x)$. This shows that $f(\{x\} \cup L) = f(x)$.

Case 2: If all the vectors $f(x), f(y_1), \ldots, f(y_{k+1})$ are collinear. In this case, the idea is to add a point $x' \in X$ such that $x' \sim x$ and $f(x')$ is not on the line containing all $f(x), f(y_1), \ldots, f(y_{k+1})$. This is possible because of the strong richness condition. By using the transitivity of the $\succcurlyeq$, $\forall y \in L: \ x' \succ y$.

Fix a point $y_0 \in L$ such that $f(y_0) \neq f(x)$. This is possible since we already assumed that not all $f(y)$,with $y \in L$, are the same as $f(x)$.

Consider the coalition $\{x\} \cup \{x'\} \cup L$. By using the weighted averaging axiom and the sub-coalitions $\{x, L \backslash \{y_0\}\}$, $\{x', y_0\}$, we have $f(\{x\} \cup \{x'\} \cup L) \in line(f(\{x, L \backslash \{y_0\}\}), f(\{x', y_0\}))$. Using the induction hypothesis, $f(\{x, L \backslash \{y_0\}\}) = f(x)$ and $f(\{x', y_0\}) = f(x')$. Therefore, $f(\{x\} \cup \{x'\} \cup L) \in line(f(x), f(x'))$.

Next, we show that $f(\{x\} \cup \{x'\} \cup L) \neq f(x)$. Since $x \sim x'$ and $f(x) \neq f(x')$, we have $f(\{x, x'\}) \neq f(x')$. Moreover, based on the way we selected the point x', $f(\{x, x'\})$ is not on the line containing $f(x)$ and $\{f(y) | y \in L\}$. Consider the partition of $\{x\} \cup \{x'\} \cup L$ into $\{x, x'\}$ and L. Based on the choice of the L, at the beginning of Step 1, $f(L) \neq f(x)$. Since $f(\{x, x'\}) \neq f(x)$, $f(L) \neq f(x)$, and $f(x), f(x'), f(L)$ are non-collinear, we have $f(\{x\} \cup \{x'\} \cup L) \neq f(x)$.

Finally, by partitioning $\{x\} \cup \{x'\} \cup L$ into $\{x'\}$ and $\{x\} \cup L$, the weighted averaging axiom results in $f(\{x\} \cup \{x'\} \cup L) \in line(f(\{x\} \cup L), f(x'))$. Therefore, $f(\{x\} \cup L)$ is on the line joining $f(\{x\} \cup \{x'\} \cup L)$ and $f(x')$. However, we have already shown that $f(\{x\} \cup \{x'\} \cup L)$ is on the line passing through $f(x)$ and $f(x')$. Thus, $f(\{x\} \cup L)$ should be on the line joining $f(x)$ and $f(x')$. However, the only intersection of $line(f(x), f(x'))$ and the line containing all the points $f(x), f(y_1), \ldots, f(y_{k+1})$ is the point $f(x)$. Thus, $f(\{x\} \cup L) = f(x)$, which completes the proof.

Step 2: In this step, by using induction on the cardinality of the set H, in which all elements have the same order, we show that for any coalition L if all elements of the set L have lower orderings compared to the elements of H, then we should have $f(H \cup L) = f(H)$.

Fix a set L. Based on Step 1, we know that for any $x \in X$ such that $\forall y \in L : x \succ y$, we should have $f(\{x\} \cup L) = f(x)$. This is the starting point of our induction. Assume that for all $|H| = k$, we have $f(H \cup L) = f(H)$. We will show that for any $|H| = k + 1$, we have $f(H \cup L) = f(H)$.

For any $x \in H$, by the weighted averaging axiom over the sub-coalitions $\{x\} \cup L$ and $H \backslash \{x\}$, we have $f(H \cup L) \in line(f(\{x\} \cup L), f(H \backslash \{x\}))$. Based on step 1, we know that $f(\{x\} \cup L) = f(x)$. Therefore, $f(H \cup L) \in line(f(x), f(H \backslash \{x\}))$. Similarly, by the weighted averaging axiom over the coalition H and its sub-coalitions $\{x\}, H \backslash \{x\}$, we should have $f(H) \in line(f(x), f(H \backslash \{x\}))$. Consider two cases:

Case 1: Consider the case in which not all members of $\{f(x) | x \in H\}$ are collinear. Hence, there should be at least two elements $x, y \in H$ that $f(x), f(H \backslash \{x\}), f(y)$, and $f(H \backslash \{y\})$ are not collinear. Therefore, the intersection of the lines joining $f(x), f(H \backslash \{x\})$ and the line joining $f(y), f(H \backslash \{y\})$ can have at most one intersection. Since $f(H)$ is on both lines, the unique intersection should be $f(H)$. But $f(H \cup L)$ is also on both lines. Hence, we should have $f(H \cup L) = f(H)$, which completes this case.

Case 2: Consider the case where all members of the set $\{f(x) | x \in H\}$ are on a line. By using the strong richness condition, there exists an element $x' \in X$ such that $f(x')$ is not on that line. We consider the coalition $\{x'\} \cup H \cup L$. By using the weighted averaging axiom over the sub-coalitions $\{x'\} \cup L$ and H, we should have $f(\{x'\} \cup H \cup L)$ on the line joining $f(\{x'\} \cup L)$ and $f(H)$. By the induction hypothesis, $f(\{x'\} \cup L) = f(x')$. Hence, we should have $f(\{x'\} \cup H \cup L) \in line(f(x'), f(H))$.

Similarly, by partitioning the set $\{x'\} \cup H \cup L$ into $H \cup L$ and $\{x'\}$, we have $f(\{x'\} \cup H \cup L) \in line(f(x'), f(H \cup L))$.

Select an element $x_1 \in H$. By partitioning the coalition $\{x'\} \cup H \cup L$ between the sub-coalitions $\{x_1, x'\}$ and $(H \backslash \{x_1\}) \cup L$, using the weighted averaging axiom, we obtain $f(\{x'\} \cup H \cup L) \neq f(\{x'\})$.

Finally, using (1) $f(\{x'\} \cup H \cup L) \in line(f(x'), f(H))$, and (2) $f(\{x'\} \cup H \cup L) \in line(f(x'), f(H \cup L))$, and (3) $f(\{x'\} \cup H \cup L) \neq f(x')$, we have $line(f(x'), f(H)) = line(f(x'), f(H \cup L))$. But the intersection of the last line with the line containing all the elements of H, can have at most one intersection. Therefore, $f(H \cup L) = f(H)$, which completes the proof.

Completing the proof:

Consider a coalition H where all elements have the similar order. We consider any two disjoint sub-coalitions $H_1, H_2 \in H$, where $f(H_1) \neq f(H_2)$. Using the same technique of the previous part, we have $f(H_1 \cup H_2) \neq f(H_1)$.

By using the result of Theorem 4, we can get the appropriate representation in each equivalence class. Also by using the result of the previous part, $f(A) = f(M(A, \succcurlyeq))$. The combination of these two results completes the proof. $\square$

B.3 Proof of Theorem 6

The following two lemmas help us prove the theorem.

Lemma 7. *Given any two linearly independent vectors v_1, v_2 in $\mathbb{R}^n$, there exists a neighborhood of v_1 that any vector in that neighborhood is linearly independent of v_2. More generally, given any m vectors $\{v_1, \ldots, v_m\}$ such that v_1 is not in the linear space generated by the rest of the points, then there exists a neighbor of v_1 such that any point in that neighborhood is not in span$(\{v_2, \ldots, v_m\})$.*

Proof. Since $K = \text{span}(\{v_2, \ldots, v_m\})$ is a closed set that is disjoint from the vector v_1, the distance between v_1 and K should be nonzero. Hence, there exists a neighborhood of v_1 (for example the ball with radios $dist(v_1, K)/2$ around v_1) disjoint from K. As a result, any point in that neighborhood is not in $\text{span}(\{v_2, \ldots, v_n\})$. $\square$

Lemma 8. *Let $v_1, v_2 \in \mathbb{R}^n$ be two linearly independent vectors and $v = \alpha v_1 + (1-\alpha)v_2$, for some $\alpha \in [0,1]$, is a vector between v_1, v_2. If the vectors $v_n = \alpha_n v_1 + (1-\alpha_n)v_2^n$ are such that $\alpha_n \in [0,1]$, $v_2^n \to v_2$, and $v_n \to v$, then $\alpha_n \to \alpha$.*

Proof. We prove it by contradiction. If it is not the case, there exists a subsequence α_{n_k} of α_n and some $\epsilon > 0$, such that $\forall\ n_k :\ \ \alpha_{n_k} \notin B_\epsilon(\alpha)$. Based on compactness of $[0,1]$, there exist a subsequence $\alpha_{n_{k_j}}$ of α_{n_k} that is convergent to some $\beta \in [0,1]$. Since $\alpha_{n_k} \notin B_\epsilon(\alpha)$, we have $\beta \neq \alpha$. Based on the assumption of the lemma, since the sequence v_n is convergent to v, the subsequence $v_{n_{k_j}}$ also converges to v. Similarly, $v_2^{n_{k_j}}$ converges to v_2. Hence, $v_{n_{k_j}} = \alpha_{n_{k_j}} v_1 + (1-\alpha_{n_{k_j}})v_2^{n_{k_j}} \to \beta v_1 + (1-\beta)v_2$ and $v_{n_{k_j}} \to v$. As a result, $\beta v_1 + (1-\beta)v_2 = v = \alpha v_1 + (1-\alpha)v_2$. However, since v_1, v_2 are linearly independent, α and β should be the same, which is a contradiction. The contradiction shows that $\alpha_n \to \alpha$. $\square$

Using the lemmas mentioned above, we will complete the proof. Based on Theorem 2.4, there exist a unique weak order $\succcurlyeq$ and a weight function $w : X \to \mathbb{R}_{++}$ such that for any $A \in X^*$

$$f(A) = \frac{\sum\limits_{x \in M(A,\succcurlyeq)} w(x)f(x)}{\sum\limits_{x \in M(A,\succcurlyeq)} w(x)}.$$

Let $x \in X$ be any given point. We need to prove that the weight function is continuous around x and any point close enough to x has the same order, respect to the weak order $\succcurlyeq$, as x.

To complete the proof, assume that $x_n \in X$ and $x_n \to x$. We are going to prove that:

1) $w(x_n) \to w(x)$,

2) $\exists N \in \mathbb{N}$ such that for all $n > N : x_n \sim x$.
Proving these two completes the proof.

Based on the strong richness condition, there should be a point $y \in X$ such that (1) $f(x), f(y)$ are linearly independent, and (2) $f(\{x,y\}) = \frac{w(x)f(x)+w(y)f(y)}{w(x)+w(y)}$, which means that $x \sim y$. The reason is that by the strong richness condition, there should be at least two other points y, z with the same order as x, such that not all of $f(x), f(y)$, and $f(z)$ are collinear. This means that $f(x)$ and at least one of $f(y)$ or $f(z)$ should be linearly independent. Without loss of generality, we assume that $f(x)$ and $f(y)$ are linearly independent.

Given any two points $a, b \in X$, we define the function $\mathbf{1}_a(b)$ as follows:

$$\mathbf{1}_a(b) = \begin{cases} 1 & \text{if } b \succcurlyeq a, \\ 0 & \text{Otherwise.} \end{cases}$$

Consider the sequence of vectors $f(\{x_n, y\})$. By Theorem 5, we have $f(\{x_n, y\}) = \frac{\mathbf{1}_y(x_n)w(x_n)f(x_n)+\mathbf{1}_{x_n}(y)w(y)f(y)}{\mathbf{1}_y(x_n)w(x_n)+\mathbf{1}_{x_n}(y)w(y)} = \frac{\mathbf{1}_y(x_n)w(x_n)}{\mathbf{1}_y(x_n)w(x_n)+\mathbf{1}_{x_n}(y)w(y)} f(x_n) + \frac{\mathbf{1}_{x_n}(y)w(y)}{\mathbf{1}_{x_n}(y)w(y)+\mathbf{1}_y(x_n)w(x_n)} f(y)$. Based on continuity of the aggregation rule f, $f(x_n) \to f(x)$ and $f(\{x_n, y\}) \to f(\{x, y\})$. Since $f(x)$ and $f(y)$ are linearly independent, all conditions of Lemma 8 are satisfied. Hence, we have $\frac{\mathbf{1}_{x_n}(y)w(y)}{\mathbf{1}_y(x_n)w(x_n)+\mathbf{1}_{x_n}(y)w(y)} \to \frac{w(y)}{w(x)+w(y)}$ and $\frac{\mathbf{1}_y(x_n)w(x_n)}{\mathbf{1}_y(x_n)w(x_n)+\mathbf{1}_{x_n}(y)w(y)} \to \frac{w(x)}{w(x)+w(y)}$. Since both $w(x)$ and $w(y)$ are strictly positive, we should have $\mathbf{1}_y(x_n) \to 1$ and similarly $\mathbf{1}_{x_n}(y) \to 1$. This means that for large n, $x_n \sim y$. Since $y \sim x$, for large n we have $x_n \sim x$. This complete part 2 of the proof.

For the part 1, since we have already proved that for large n, $x_n \sim x \sim y$, the convergence $\frac{\mathbf{1}_{x_n}(y)w(y)}{\mathbf{1}_y(x_n)w(x_n)+\mathbf{1}_{x_n}(y)w(y)} \to \frac{w(y)}{w(x)+w(y)}$ becomes $\frac{w(y)}{w(x_n)+w(y)} \to \frac{w(y)}{w(x)+w(y)}$. This means that $w(x_n) \to w(x)$, which proves that w is continuous at x.

Proving part 1 and 2 complete the proof.

$\square$

B.4 Proof of Proposition 9

There are a couple of steps to prove the result.

Step 1: Assume that all signals arrive at time 1. By using Corollary 3, there exists a unique (up to multiplication) weight function $w : X^* \to \mathbb{R}_{++}$, such that for all $A \in X^*$, $f(A,1) = \frac{\sum_{x\in A} w(x)f(x)}{\sum_{x\in A} w(x)}$. By using the uniqueness of w and the stationarity axiom, for any constant time shift c and for all $A \in X^*$ we have:

$$f(A,c) = \frac{\sum\limits_{x\in A} w(x)f(x)}{\sum\limits_{x\in A} w(x)}.$$

Consider two signals $x_0, y_0 \in X$, where $f(x_0) \neq f(y_0)$. Let the timing of x_0, y_0 be $T_{\{x_0,y_0\}}(x_0) = 1, T_{\{x_0,y_0\}}(y_0) = 2$. Using the strict weighted averaging axiom, there exists a $\lambda \in (0,1)$ where $f(\{x_0,y_0\}, T_{\{x_0,y_0\}}) = \lambda f(x_0) + (1-\lambda)f(y_0)$. We define q such that $\frac{1-\lambda}{\lambda} = q \times \frac{w(y_0)}{w(x_0)}$.

In the rest of the proof, we show that these choices of w, q attain the representation of Proposition 9.

Step 2: We show that for any signal $z \in X$, the representation holds for the coalition $\{x_0, z\}$ and for the timing function $T_{\{x_0,z\}}(x_0) = 1, T_{\{x_0,z\}}(z) = 2$.

Case 1: Consider any signal $z \in X$, such that $\{f(x_0), f(y_0), f(z)\}$ are not collinear. We form the coalition $\{x_0, y_0, z\}$ with the timing $T_{\{x_0,y_0,z\}}(x_0) = 1, T_{\{x_0,y_0,z\}}(y_0) = 2$, $T_{\{x_0,y_0,z\}}(z) = 2$. Using the strict weighted averaging axiom, by considering the sub-coalitions $\{x_0\}$ and $\{y_0, z\}$ and the fact that y_0 and z has the same timing, Lemma 6, in the proof of Theorem 4, shows that the representation holds for the coalition $\{x_0, z\}$ with the timing $T_{\{x_0,z\}}(x_0) = 1, T_{\{x_0,z\}}(z) = 2$.

Case 2: Consider any signal $z \in X$, such that $\{f(x_0), f(y_0), f(z)\}$ are collinear. By the richness condition, there exists a signal $z' \in X$ such that $\{f(x_0), f(y_0), f(z), f(z')\}$ are not collinear. We consider the timing $T_{\{x_0,y_0,z,z'\}}(x_0) = 1, T_{\{x_0,y_0,z,z'\}}(y_0) = 2$, $T_{\{x_0,y_0,z,z'\}}(z) = 2, T_{\{x_0,y_0,z,z'\}}(z') = 2$. The representation holds for the sub-coalitions $\{x, y, z'\}$ (by Case 1) and $\{y, z, z'\}$ (since all have the same timing). Thus, by applying Lemma 6 first on $\{y, z, z'\}$ and then on $\{x, y, z'\}$, we can show that the representation holds for the coalition $\{x_0, z\}$ with the timing $T_{\{x_0,z\}}(x_0) = 1, T_{\{x_0,z\}}(z) = 2$.

Step 3: We show that the representation holds for any two signals $u, v \in X$ with the timing function $T_{\{u,v\}}(u) = 1, T_{\{u,v\}}(v) = 2$.

Case 1: If $\{f(x_0), f(u), f(v)\}$ are non-collinear, then we consider the timing function $T_{\{x_0,u,v\}}(x_0) = 1, T_{\{x_0,u,v\}}(u) = 1, T_{\{x_0,u,v\}}(v) = 2$. By applying Lemma 6 twice on $\{x_0, u\}$ and $\{x_0, v\}$ with their corresponding timings, we can show that the representation should holds for $u, v \in X$ with the timing function $T_{\{u,v\}}(u) = 1, T_{\{u,v\}}(v) = 2$.

Case 1: If $\{f(x_0), f(u), f(v)\}$ are collinear, then by the richness condition, there exists a signal $z \in X$ such that $\{f(x_0), f(u), f(v), f(z)\}$ are not collinear. Consider the timing function $T_{\{x_0,u,v,z\}}(x_0) = 1, T_{\{x_0,u,v,z\}}(u) = 1, T_{\{x_0,u,v,z\}}(v) = 2, T_{\{x_0,u,v,z\}}(z) = 2$. By applying Lemma 6 for the sub-coalition $\{x_0, u, z\}$ and their corresponding timing, shows that the representation holds for $\{u, z\}$ and their timing $T_{\{u,z\}}(u) = 1, T_{\{u,z\}}(z) = 2$. Then, by considering the coalition $\{u, v, z\}$ and their corresponding timing, Lemma 6 shows that the representation holds for $\{u, v\}$ and the timing function $T_{\{u,v\}}(u) = 1, T_{\{u,v\}}(v) = 2$.

Step 4: In this step, we show that given any $t \in \mathbb{N}$, the representation holds for any two signals $u, v \in X$ and the timing function $T_{\{u,v\}}(u) = 1, T_{\{u,v\}}(v) = t$. The proof is by induction on t. By Step 3, the representation holds for $t = 2$. Assume that the representation holds for all $t < k$ with $k > 3$. We will show that it also holds for $t = k$.

Case 1: If $f(u) \neq f(v)$, then we consider a signal $z \in X$ such that $\{f(u), f(v), f(z)\}$ are not collinear. Let the timing function be $T_{\{u,v,z\}}(u) = 1, T_{\{u,v,z\}}(v) = k + 1, T_{\{u,v,z\}}(z) = k$.

Consider any $w(u, 1), w(v, k), w(z, k-1) \in (0, 1)$ such that $f(\{u, v, z\}, T_{u,v,z}) = w(u, 1)f(u) + w(v, k)f(v) + w(z, k-1)f(z)$. By Lemma 6, induction hypothesis, and the stationarity axiom, we have $\frac{w(v,k)}{w(u,1)} = \frac{w(v,k)}{w(z,k-1)} \times \frac{w(z,k-1)}{w(u,1)} = (q\frac{w(v)}{w(z)})(q^{k-2}\frac{w(z)}{w(u)}) = q^{k-1}\frac{w(v)}{w(u)}$. Thus, the representation holds.

Case 1: If $f(u) = f(v)$, then we consider two signals $z, z' \in X$ such that $\{f(u), f(v), f(z), f(z')\}$ are not collinear (which is possible by the richness condition). Let the timing function be $T_{\{u,v,z,z'\}}(u) = 1, T_{\{u,v,z,z'\}}(v) = k + 1, T_{\{u,v,z,z'\}}(z) = k, T_{\{u,v,z,z'\}}(z') = k$. By the uniqueness part of Theorem 4 and the induction hypothesis, the representation still holds in this case.

Step 4: Finally, for any coalition $A \in X^*$ and any timing function T_A, the uniqueness of Theorem 4 and Step 4 establish that the representation should hold with q, w.

Appendix C

APPENDIX TO CHAPTER 3

C.1 Proof of Theorem 8

Assume that the aggregation rule $f : X^* \to \mathcal{R}^m$ satisfies the minimal agreement condition and $v \in \mathbb{R}^m$ is the direction on which all agents agree.

Consider two disjoint coalitions $A, B \in X^*$ with the corresponding cardinal utilities $u_A \in U_A$ and $u_B \in U_B$. Assume that $u_{A\cup B} \in U_{A\cup B}$ is a cardinal utility that represents the preference ordering of the union $A \cup B$. If $u_A = u_B$, then the result is trivial. Hence, consider the case $u_A \neq u_B$.

First, by using the Farkas' Lemma, we show that the extended Pareto is equivalent to $u_{A\cup B} \in \text{Cone}^\circ(u_A, u_B)$ (which $\text{Cone}^\circ(u_A, u_B)$ denotes the interior of the cone generated by u_A and u_B).

If $u_{A\cup B} \in \text{Cone}^\circ(u_A, u_B)$, then there exist $\alpha, \beta > 0$ such that $u_{A\cup B} = \alpha u_A + \beta u_B$. Therefore, for any $x, y \in L$ if $u_A \cdot x \geqslant u_A \cdot y$ and $u_B \cdot x \geqslant u_B \cdot y$, then $u_{A\cup B} \cdot x \geqslant u_{A\cup B} \cdot y$. Similarly, if $u_A \cdot x > u_A \cdot y$ and $u_B \cdot x \geqslant u_B \cdot y$, then $u_{A\cup B} \cdot x > u_{A\cup B} \cdot y$. This proves that the preference ordering of $A \cup B$ satisfies the extended Pareto axiom.

For the other side, if the utility of the union $u_{A\cup B} \notin \text{Cone}^\circ(u_A, u_B)$, then $\nexists \alpha, \beta > 0$ such that $u_{A\cup B} = \alpha u_A + \beta u_B$. The Farkas' Lemma guarantees that there exists a vector $z \in \mathbb{R}^m$ such that $z \cdot u_A \geqslant 0, z \cdot u_B \geqslant 0$ and $z \cdot u_{A\cup B} < 0$.

Consider a vector $y \in \mathbb{R}^m$ that is in the interior of L. We select $\lambda > 0$ such that $y + \lambda z \in L$. This is possible since we assume that y is in the interior of L. By defining $x = y + \lambda z$, we get $x - y = \lambda z$. Since $\lambda > 0$ and $z \cdot u_A \geqslant 0, z \cdot u_B \geqslant 0$, we have $u_A \cdot x \geqslant u_A \cdot y$, and $u_B \cdot x \geqslant u_B \cdot y$. But since $z \cdot u_{A\cup B} < 0$, we have $u_{A\cup B}$ and $x < u_{A\cup B} \cdot y$. But by the extended Pareto axiom, this cannot be true. Therefore, $u_{A\cup B} \in \text{Con}^\circ(u_A, u_B)$.

Now consider the intersection of $H = \{x \in \mathbb{R}^m |\ x \cdot v = 1\}$ and $\text{Cone}^\circ(u_A, u_B)$. Since $u_A \cdot v > 0$ and $u_B \cdot v > 0$, there should be a unique $\hat{u}_A \in U_A, \hat{u}_B \in U_B$ both in H. It is trivial that $\text{Con}^\circ(u_A, u_B) = \text{Con}^\circ(\hat{u}_A, \hat{u}_B)$.

Since both $\hat{u}_A.v > 0$ and $\hat{u}_B.v > 0$, the intersection of the interior of the cone generated by them and the linear variety H is the segment $[\hat{u}_A, \hat{u}_B] = \{\lambda \hat{u}_A + (1-\lambda)\hat{u}_B | \ \lambda \in (0,1)\}$. Since $u_{A\cup B} \in \text{Con}^\circ(\hat{u}_A, \hat{u}_B)$, we should have $v.u_{A\cup B} > 0$. Hence, there should be a $\hat{u}_{A\cup B} \in H$ representing $u_{A\cup B}$. Therefore, $\hat{u}_{A\cup B} \in [\hat{u}_A, \hat{u}_B]$. This completes the proof.

□

C.2 Proof of Theorem 9

There are a couple of steps in the proof. Note that for any profile $R \in \mathcal{R}_X$, and for any coalition $A \subseteq X$, R_A denotes the restricted sub-profile of the coalition A.

Step 1: Fix a preference $\hat{r} \in \mathcal{R}_{\overline{x}}$. Using the corollary 8, for any profile $R \in \mathcal{R}_X$ such that $R_1 = \hat{r}$, we can uniquely define a weight function (which depends on the full profile R) $w^R(i) : X \to \mathbb{R}_{++}$ with $w^R(1) = 1$, such that for any coalition $A \subseteq X$ we have:

$$u_H(f(R_A)) = \sum_{i\in A} \left(\frac{w^R(i)}{\sum\limits_{j\in A} w^R(j)} \right) u_H(R_i).$$

First, we show that for any individual $i \in X\backslash\{1\}$ and for any two profiles $R_a, R_b \in \mathcal{R}_X$ with $(R_a)_1 = (R_b)_1 = \hat{r}$ and $(R_a)_i = (R_b)_i$, we have $w^{R_a}(i) = w^{R_b}(i)$. There are two separate cases:

Case 1: If $(R_a)_i = (R_b)_i \neq \hat{r}$, then using $(R_a)_{\{1,i\}} = (R_b)_{\{1,i\}}$ and the result of corollary 8, we should have $w^{R_a}(i) = w^{R_b}(i)$.

Case 2: If $(R_a)_i = (R_b)_i = \hat{r}$, then by considering the definition of the domain $\mathcal{R}_X$, which require the existence of three non-collinear preferences in each profile, there should be a profile $R_c \in \mathcal{R}_X$ and two individual $j_1, j_2 \in X\backslash\{1, i\}$ such that $(R_c)_1 = (R_c)_i = \hat{r}$, $(R_c)_{j_1} = (R_a)_{j_1} \neq \hat{r}$, and $(R_c)_{j_2} = (R_b)_{j_2} \neq \hat{r}$. Using Case 1, we have $w^{R_c}(j_1) = w^{R_a}(j_1)$ and $w^{R_c}(j_2) = w^{R_b}(j_2)$. Since $(R_a)_{\{i,j_1\}} = (R_c)_{\{i,j_1\}}$ and $w^{R_a}(j_1) = w^{R_c}(j_1)$, using corollary 8, we should have $w^{R_a}(i) = w^{R_c}(i)$. Similarly, we have $(R_b)_{\{i,j_2\}} = (R_c)_{\{i,j_2\}}$ and $w^{R_b}(j_2) = w^{R_c}(j_2)$. Therefore, we should have $w^{R_b}(i) = w^{R_c}(i)$. Hence, we have $w^{R_a}(i) = w^{R_b}(i)$.

By considering profiles of the form $R \in \mathcal{R}_X$ with $R_1 = \hat{r}$, we can define the weight function $w : X\backslash\{1\} \times \mathcal{R}_{\overline{x}} \to \mathbb{R}_{++}$ such that $w(i, R_i) = w^R(i)$ for all $i \in X\backslash\{1\}$. By the result of Step 1, this function is well defined. Moreover, we define $w(1, \hat{r}) = 1$.

At this point, for any preference profile $R \in \mathcal{R}_X$ with $R_1 = \hat{r}$ and for any coalition $A \subseteq X$, we have:

$$u_H(f(R_A)) = \sum_{i \in A} \left(\frac{w(i, R_i)}{\sum\limits_{j \in A} w(j, R_j)} \right) u_H(R_i).$$

Step 2: We need to define $w(1, r)$ for all $r \in \mathcal{R}_{\overline{x}}$. We have already fixed the value $w(1, \hat{r}) = 1$. For any $r \in \mathcal{R}_{\overline{x}}\backslash\{\hat{r}\}$, let $R \in \mathcal{R}_X$ be a profile with $R_1 = r$ and $R_2 = \hat{r}$. By corollary 8, there should be a unique function $w^R : X \to \mathbb{R}_{++}$ with $w^R(2) = w(2, \hat{r})$. We define $w(1, r) = w^R(1)$.

Notice that for any two profile $R_a, R_b \in \mathcal{R}_X$ with $(R_a)_1 = (R_b)_1 = r$ and $(R_a)_2 = (R_a)_2 = \hat{r}$, if we normalize the value of the $w^{R_a}(2) = w^{R_b}(2) = w(2, \hat{r})$, then we should have $w^{R_a}(1) = w^{R_b}(1)$. Hence, the value $w(1, r)$ is independent of the choice of the profile R.

At this point the function $w : X \times \mathcal{R}_{\overline{x}} \to \mathbb{R}_{++}$ is fully defined. We only need to show that it works.

Step 3: Select any profile $R \in \mathcal{R}_X$. We need to show that the representation holds with the weight function defined above.

If $R_1 = \hat{r}$, by the result of Step 1 the representation holds. Hence, fix any $\overline{r} \in \mathcal{R}_{\overline{x}}$ where $\overline{r} \neq \hat{r}$. In the rest of the proof we show that the representation holds for any $R \in \mathcal{R}_X$ with $R_1 = \overline{r}$.

Similar to Step 1, using the corollary 8, for any profile $R \in \mathcal{R}_X$ such that $R_1 = \overline{r}$, we can uniquely define a weight function (depending on the full profile R) $w'^R(i) : X \to \mathbb{R}_{++}$ with $w'^R(1) = w(1, \overline{r})$, such that for any coalition $A \subseteq X$ we have:

$$u_H(f(R_A)) = \sum_{i \in A} \left(\frac{w'^R(i)}{\sum\limits_{j \in A} w'^R(j)} \right) u_H(R_i).$$

In the same manner as Step 1, for any two profiles $R_a, R_b \in \mathcal{R}_X$ with $(R_a)_1 = (R_b)_1 = \overline{r}$ and for every individual $i \in X$, we should have $w'^{R_a}(i) = w'^{R_b}(i)$. Hence, by considering profiles of the form $R \in \mathcal{R}_X$ with $R_1 = \overline{r}$, we can define the weight function $w' : X\backslash\{1\} \times \mathcal{R}_{\overline{x}} \to \mathbb{R}_{++}$ such that for all $i \in X\backslash\{1\}$: $w'(i, R_i) = w'^{R}(i)$. By the result of Step 1, this function is well defined. Moreover, we fix $w'(1, \overline{r}) = w(1, r)$.

For every preference profile $R \in \mathcal{R}_X$ with $R_1 = \overline{r}$ and for every coalition $A \subseteq X$, we have:

$$u_H(f(R_A)) = \sum_{i \in A} \left(\frac{w'(i, R_i)}{\sum\limits_{j \in A} w'(j, R_j)} \right) u_H(R_i).$$

To complete the proof, since we have $w(1, \overline{r}) = w'(1, \overline{r})$, it remains to show that for all $i \in X\backslash\{1\}$ and for all $r \in \mathcal{R}_{\overline{x}}$ we have $w(i, r) = w'(i, r)$.

Case 1: Since $\hat{r} \neq \overline{r}$ and $w(1, \hat{r}) = w'(1, \hat{r})$, based on Step 2 we should have $w(2, \hat{r}) = w'(2, \hat{r})$.

Case 2: Assume that $r \neq \hat{r}$ and $i \in X\backslash\{1, 2\}$. Since $N \geqslant 5$, based on definition of $\mathcal{R}_X$, there exist $R_a, R_b \in \mathcal{R}_X$ such that $(R_a)_1 = \hat{r}, (R_a)_2 = \hat{r}, (R_a)_i = r$ and $(R_b)_1 = \overline{r}, (R_b)_2 = \hat{r}, (R_b)_i = r$.

Since $r \neq \hat{r}$, $(R_a)_{\{2,i\}} = (R_b)_{\{2,i\}}$, and by Case 1 $w(2, \hat{r}) = w'(2, \hat{r})$, then we should have $w(i, r) = w'(i, r)$.

Case 3: Assume that $r = \hat{r}$ and $i \in X\backslash\{1, 2\}$. Since $N \geqslant 5$, we can select an individual $j \in X\backslash\{1, 2, i\}$. Based on the definition of $\mathcal{R}_X$, there exist $R_a, R_b \in \mathcal{R}_X$ such that $(R_a)_1 = \hat{r}, (R_a)_i = \hat{r}, (R_a)_j = \overline{r}$ and $(R_b)_1 = \overline{r}, (R_b)_i = \hat{r}, (R_b)_j = \overline{r}$.

Since $r = \hat{r} \neq \overline{r}$, $(R_a)_{\{i,j\}} = (R_b)_{\{i,j\}}$, and by Case 2 $w(j, \overline{r}) = w'(j, \overline{r})$, then we should have $w(i, r) = w'(i, r)$.

Case 4: Finally, assume that $i = 2$ and $r \neq \hat{r}$. Select an individual $j \in X\backslash\{1, 2\}$. We consider profiles $R_a, R_b \in \mathcal{R}_X$ such that $(R_a)_1 = \hat{r}, (R_a)_2 = r, (R_a)_j = \hat{r}$ and $(R_b)_1 = \overline{r}, (R_b)_2 = r, (R_a)_j = \hat{r}$. By Case 3, we have $w(i, \hat{r}) = w'(i, \hat{r})$. Hence, since $r \neq \hat{r}$ and $(R_a)_{\{2,j\}} = (R_b)_{\{2,j\}}$ we should have $w(2, r) = w'(2, r)$.

The last observation completes the proof. $\square$