Fourier Analysis for Social Choice

MSc Thesis *(Afstudeerscriptie)*

written by

Frank FEYS
(born August 3rd, 1988 in Hasselt, Belgium)

under the supervision of **Prof. Dr. Ronald de Wolf**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
June 17th, 2015	Dr. Alexandru Baltag (*chair*)
	Prof. Dr. Jan van Eijck
	Dr. Ulle Endriss
	Prof. Dr. Hans Peters
	Prof. Dr. Ronald de Wolf

Institute for Logic, Language and Computation

Abstract

Social choice theory studies mathematically the processes involved when groups of people make choices. There are a number of beautiful and astonishing qualitative results in this area, for example Arrow's Theorem about the non-existence of ideal voting schemes [1], and the Gibbard-Satterthwaite Theorem [57, 116] about the manipulation of elections. These classical theorems have had tremendous impact on the field of social choice.

Recently, there has been a sequence of stronger, quantitative versions of such theorems, due to Gil Kalai, Ehud Friedgut, Elchanan Mossel, and others [68, 88]. These results depend on the theory of Fourier analysis on the Boolean cube.

In this thesis, we seek to connect the at first seemingly disparate realms of social choice and Fourier analysis on the Boolean cube.

The first goal of this thesis is to study the aforementioned strengthened theorems. The second goal is to make them more accessible to researchers working in social choice. On our way to these results, we build up the theory of Boolean analysis, introducing pivotal notions such as *influence* and *noise*. Such concepts are of interest in their own right, as we will try to show by proving classical results such as the KKL Theorem [66] and Friedgut's Theorem [49]. The third goal, then, is to convince the reader that Fourier analysis on the Boolean cube is a worthwhile technique to consider for researchers in social choice. A common theme throughout the thesis is the *impartial culture assumption*: the contentious mathematical assumption that voters vote independently and randomly of one another. In the last chapter, the final goal is achieved: inspired by Daniel Kahneman's work on cognitive biases [67], a new, simple, model to simulate the various biases that show up in small meetings involving sequential voting is introduced.

Keywords: Social Choice, Arrow's Theorem, Gibbard-Satterthwaite Theorem, Fourier Analysis, Voting, Elections

Contents

Chapter 1

Introduction

The topics in this thesis are situated at the interchange of the fields of theoretical computer science and social choice theory. Concretely, the thesis deals with a mathematical technique that is commonly applied in theoretical computer science, but that is now put to use in the field of social choice. This technique is called *Fourier analysis on the Boolean cube.* "Boolean" here refers to the fact that all the input variables involved can only take one of two values, usually denoted as 0 or 1. It should not be surprising that in computer science Boolean structures are numerous. For example, we can think of a computer which is given some input bits (by someone pushing the buttons of the keyboard), and after a computation some output appears on the computer's screen. These are the output bits.

However, over time it has become clear that Boolean structures also prove useful in other, perhaps seemingly distant, areas. Initiated in the late 1980s mainly by the works of Ben-Or and Linial [7], and Kahn, Kalai, and Linial [66], discrete Fourier analysis has been successfully applied in social choice also. The idea is that the bits 0 and 1 here represent the candidates that people may choose from; for example, the two candidates in the United States presidential election.

The use of Fourier analysis on the Boolean cube in the area of social choice has led to many interesting results. One of the key achievements is a strengthening of Arrow's impossibility theorem about the non-existence of ideal voting schemes, obtained by Kalai in 2002 [68]. In Chapter 6 we will give a proof of this result. The main advantage of the analytical approach is that a *quantitative version* of Arrow's Theorem is obtained. Roughly stated, this result says the following: the more we want to avoid Condorcet's paradox, the more the election scheme will look like a dictator.[1] Further research [50, 51, 57, 65] has also led to a quantitative version of the Gibbard-Satterthwaite Theorem, which we will review in Chapter 5.

In the following section, we state the aims of the thesis. After that, we give an informal, historical, overview in which we introduce the key players of this thesis. Then, we examine some assumptions underlying the approach. The most important such assumption used in this thesis is the *impartial culture assumption*, which states that each voter votes at random and, moreover, all voters vote independently of one another. While this assumption is convenient from a mathematical perspective, it should be not surprising that it is somewhat unrealistic. The most troublesome point seems to be that it does not capture any *correlations* between people's voting tendencies. Briefly stated, the problem

[1] At least, under the assumption of an impartial culture; see further on.

is that most quantitative results in this thesis do rely on this assumption, and therefore we should be careful not to over-interpret them. In that section we also relate the impartial culture assumption to a discussion from political science, started by William Riker's *Liberalism Against Populism* [111]. The last section contains an outline of the thesis.

Before we proceed, just one last note about prerequisites. From the second chapter onwards, all chapters are rather formal. This thesis was, however, written to be maximally self-contained. We will mostly be working with discrete (even finite) structures, so a basic knowledge of mathematics should be sufficient. This first chapter stands out compared with the rest of the thesis in the sense that it is informal: the intention is to give the reader some perspective and, most importantly, to make the reader interested in the chapters that follow.

1.1 Aims of the Thesis

The primary aims of this thesis are as follows:

- First, to study some of the most interesting results from social choice that were obtained by means of discrete Fourier analysis; in particular, quantitative versions of Arrow's Theorem and the Gibbard-Satterthwaite Theorem.

- Second, we would like to make these results more accessible to researchers working in social choice.

- Third, we want to convince the reader that Fourier analysis on the Boolean cube is a worthwhile technique to consider for researchers in social choice.

- Fourth and last, to investigate other parts of social choice for which discrete Fourier analysis might be used to obtain results. In the last chapter, we put forward a new notion of *influence*. This notion appears in a natural way in the context of sequential binary voting.

1.2 Informal Overview

Below we will sketch some historical events. Doing so, we will certainly not attempt to be complete. Rather, we will focus on the individuals and events that are relevant for this thesis.

The main references for this section are [21, 39, 103, 118, 77, 97, 120, 82, 128, 17].

1.2.1 What is Social Choice about?

The aim of social choice theory is to study collective decision procedures [77]. Central is the following setup: given a number of persons' preferences (which can be votes, judgements, opinions, etc.), how does one reach a collective preference, the so-called *social preference*? Furthermore, how can this be done in a "fair" way—and, what would "fairness" even mean in this context? One might argue that the outcome should "do justice" to as many voters as possible in order for a choice to be "right". Concretely, given for example that each voter has submitted a ranking of all candidates, how does one arrive at a collective ranking? Such election schemes, returning a ranking when given rankings from all voters, are mathematically represented by a function, called *social welfare function*. In social choice, properties of such functions are studied.

The domain of application of social choice, as well as the different fields it touches upon, are various. Indeed, first note that the "persons" whose preferences are taken into account, need not be actual persons. They might, for example, be computers: think about a search engine aggregating rankings of webpages. Therefore, it is a good idea to speak of "agents" rather than "persons". In the context of economics, those agents might also be companies. Second, social choice is related to numerous other disciplines: originally it was mostly entangled with economics and political science, but over time several other disciplines have joined in. As the link with decision-making and thus "rationality" is quite obvious, also researchers from psychology and sociology have taken interest in results from social choice.

During the past two decades, people from computer science have become increasingly interested in social choice as well. This has led to the field of *computational social choice*. The developments examined in the present thesis should be seen in light of this, although somewhat in parallel: as we will explain below, as the name suggest computational social choice focuses on the *computational* aspects (by using ideas from complexity theory) of social choice, whereas this work does not. In fact, this very thesis can be seen as yet another bridge from social choice to computer science, although it is this time mostly just the technique itself which is transposed from the one onto the other.

1.2.2 Early Beginnings

Genesis. It is conceivable that social choice, in some (primitive) form or another, has been around for hundreds, perhaps even thousands, of years. The earliest known sources date back to the Middle Ages, when Ramon Llull (ca. 1235-1315, Majorca) proposed the aggregation method of pairwise majority voting [77]. This method, which will be particularly important to us, works as follows. Given the rankings of all voters, any pair of candidates are compared pairwise, thus obtaining a collective ranking. For example, suppose there are three voters and three candidates (say, a, b, c), and the votes are as follows:

$$a > b > c, \quad b > a > c, \quad a > c > b.$$

Here the expression $a > b > c$, e.g., means that the voter in question prefers a to b, b to c, and a to c. Notice that a wins each pairwise competition: for example versus b it wins 2 to 1, and it wins 3 to 0 versus c.[2]

An election scheme which is based upon pairwise comparisons (regardless whether it uses the majority rule or another one), is said to satisfy the *Condorcet method*.

18th Century: de Condorcet. It was only in the 18th century that social choice really started establishing itself. The most famous name from that period is undoubtedly Nicolas de Condorcet (1743-1794). Born a French nobleman, de Condorcet was, unlike other persons working in social choice at the time, an established mathematician [103]. Despite his heritage, he became a frontrunner in the French revolution. Several years later, de Condorcet was put in jail, due to a disagreement over a draft of the constitution; although he managed to escape, he was later caught. Ultimately he died in jail under mysterious circumstances.

He published his book *Essay on the Application of Probability Analysis to Majority Decisions* [27] in 1785. In it, he strongly argued for the majority rule. At the time, doing so was revolutionary. In particular, de Condorcet said, when electing two candidates, one must use the majority rule; when electing more than two candidates one should do a pairwise competition after which one, again, applies the majority rule. In general, this reduction of the election into pairwise competitions between any two candidates (whichever rules are afterwards applied on these "subelections"), is called *Condorcet's method*. Among de Condorcet's achievements in social choice, two results stand out.

First, the result which has come to be known as *Condorcet's Jury Theorem* [39], says the following. Suppose a jury of n judges has to determine whether an accused is guilty, and suppose each judge independently makes a correct decision with fixed probability $p > 1/2$. Then, the probability that the plurality rule returns the correct decision increases monotonically in n; furthermore, this probability approaches 1 as n goes to infinity.[3] This result can be seen as an example of "the wisdom of the crowd" [129]: a large group of people that are barely smart, can nonetheless take the good decision *as a group*. That is, the whole is bigger than just the sum of its parts.[4]

[2] We should immediately observe that this procedure need not always work: the *Condorcet paradox*, which we will see in a moment, is an example in which this procedure would not give a valid societal ordering. The example given here just happens to be a lucky case in which it does work.

[3] Using modern probability theory, the proof is rather trivial: just apply the law of large numbers.

[4] Of course, this is a simplistic model. For one thing, it assumes that the members of the jury are independent. Also, there seems no compelling reason to assume all members of the jury would be able to take the "good" decision with probability bigger than 50%. This assumption is rather optimistic. These issues are actually part of a more general, and serious, problem: people taking decisions are usually assumed to be *rational*. However, research by Kahneman and Tversky, started in the late 1960s, in particular has shown this assumption to be unrealistic. People are, in general, not rational. See for example *Thinking, Fast and Slow* by Kahneman [67].

Furthermore, suppose that all members of the jury would instead have a fixed probability

Second, and of major importance to this thesis, is the following "paradox" that is attributed to Condorcet. Suppose again that there are three voters and three candidates a, b, c, and the votes are as follows:

$$a > b > c, \quad b > c > a, \quad c > a > b.$$

Following Condorcet's suggestion, in order to obtain the societal preference, we do a pairwise comparison using the majority rule: a wins 2 to 1 from b, b wins from c, and finally c wins from a. Thus, the societal outcome is

$$a > b > c > a > b > c > a > \dots.$$

This means that we have obtained a cycle! There is no consistent linear order on the candidates reflecting the voters' preferences, a phenomenon now known as *Condorcet's paradox*. If in some case the paradox does not show up, the then top-ranked candidate is called the *Condorcet winner* of the election.

Another notion that is still important today and that goes back to Condorcet, is that of *Condorcet consistency*: any voting rule which selects a candidate given the preferences of all voters, must select the Condorcet winner if it exists.

Besides Condorcet, other well-known 18th century researchers include Jean-Charles de Borda (1733-1799), who introduced the *Borda count*.[5]

Questions. Some questions regarding Condorcet's paradox that arise are:

- Is there any way out? That is, is there perhaps some other "reasonable" rule rather than the majority which, when doing a pairwise contest, manages to avoid the paradox altogether?
- Furthermore, what is the probability[6] of a cycle arising? Does it happen often?

The present thesis will address all of these issues, and more. In fact, the answer to the first question has been known already since Arrow's work from the 1950s; one crucial advantage Boolean analysis will give is answering the second question.

of $q < 1/2$ to make a correct decision. In that case, the probability of the jury as a whole to make a correct decision would converge to 0 when n goes to infinity! That is, the group as a whole would do strictly *worse* than the individuals separately: the "stupidity of the crowd".

[5]This election system works as follows. Suppose, for example, again three voters and three candidates:

$$a > b > c, \quad b > a > c, \quad a > c > b.$$

The method says that each candidate in a first spot gets two points; each in a middle spot gets one point; each in a last spot gets zero points. E.g., a receives $2 + 1 + 2 = 5$ points, while b gets $1 + 2 + 0 = 3$ points. Finally, c gets $0 + 0 + 1 = 1$ point. So in this case a wins.

This system, like many others, clearly does not follow the Condorcet method.

[6]Note that this question does not make sense as stated: we first need a probabilistic model capturing the way the voters are voting. This will lead us to the *impartial culture assumption*.

19th Century. In the 19th century we encounter Charles Dodgson (1832-1898), better known under his pseudonym Lewis Carroll, who wrote the famous book *Alice's Adventures in Wonderland.* He was a man of many talents. Besides summarizing the works of his predecessors, Dodgson also came up with a new principle. One should always elect the Condorcet winner if it exists, he said; further, even if it does not exist, one should elect the candidate which is *closest*[7] to being a Condorcet winner. This rule is called *Dodgson's rule.*

1.2.3 The 20th Century Breakthrough: Kenneth Arrow

Arguably the biggest step forward so far in social choice was taken around 1950, when Kenneth Arrow (1921) published his PhD dissertation called *Social Choice and Individual Values* [1]. The following discussion is based on [120, 82].

What is so special about Arrow's work, why was it so groundbreaking? There are a number of reasons, all of which are related to each other. We give three, each of which constitutes a major contribution. Historically speaking, due to its elegance Arrow's impossibility theorem has gotten the most recognition, although in the long run all three are utterly significant accomplishments, according to Sen [120].

Initiating Axiomatic Reasoning. This is likely the most important reason. The basic idea behind Arrow's revolutionary approach is as follows: first, let us think about *what it actually is* we want out of a voting rule, i.e., what attractive properties do we want a voting rule to satisfy? Those properties are then translated into the language of mathematics, as *axioms.* At first, when no property has been insisted on, we just have the class of all voting rules. After that, we add an axiom (a desirable property we would like our voting rules to have), and we end up with a smaller class of voting rules, namely those ones satisfying the property. In each subsequent step, we do the same, each time adding an axiom. This will eventually lead to a voting rule which satisfies a number of these properties we had liked our "ideal" voting system to have. Of course, we would like to add as many desirable properties as possible. However, in each step the class of voting rules satisfying the properties shrinks. The question is, therefore, how long this process can go on before we end up with the empty set.

Whereas before, people such as Condorcet, Borda, etc., had mostly studied properties of particular voting methods (i.e., *in isolation*, or just comparing a few of them, focussing on a small number of specific properties), it was Arrow who first realized the merit to abstract away from those specific instances, to define the general concept of a social welfare function, and to formulate their properties *as axioms.* Amartya Sen calls the emergence of the systematic use of the axiomatic reasoning[8] in social choice "...totally constructively a major

[7] "Closest" here means "the candidate that minimizes the number of swaps of adjacent candidates in order it become a Condorcet winner".

[8] The study of different axiomatizations in mathematics, as part of *formalism*, was introduced in the 19th and 20th century by Frege, Hilbert, Russell, Whitehead, Tarski, etc.

game turner..." [120].

Presenting the Framework. Arrow introduced the framework that has since been used in social choice. In brief, it goes as follows. Each voter ranks all candidates completely, in any way she likes[9], from highest preference to lowest, say. (We will see this mathematically in more detail in Chapter 6, but for now we insist on keeping the discussion informal.) A *voting rule* is then a mechanism which, given any possible way the voters voted, produces one ranking of all candidates, the *societal outcome.* Arrow himself called it "collective rationality" at the time [120]. This very simple but precise notion lies at the origin of all subsequent research in social choice.

Arrow's Impossibility Theorem. By its very name, Arrow's impossibility theorem insinuates some kind of pessimism: something's existence is impossible, hence it cannot exist [82].

Consider the following three desirable properties:

- *Pareto condition.* If all voters rank some candidate higher than another one, then also the election scheme should do so.
- *Independence of irrelevant alternatives.* The societal ranking of two alternatives depends only on the relative rankings of those alternatives, and on no others.[10]
- *Non-dictatoriality.* There is no dictator, i.e., there is no candidate whose ranking always coincides with the societal outcome.

Sometimes a fourth axiom is added. In this thesis we will usually not mention it, simply because we already assume right off the bat that each possible ranking is available to the candidates.

- *Universal domain.* Each voter can pick any ranking, in the sense that no ranking is a priori excluded.

Even these seemingly very mild conditions of reasonableness cannot be simultaneously satisfied by any voting rule: that is what Arrow's impossibility theorem says. Amartya Sen calls the impossibility theorem a "staggeringly

[120, 125, 148]. Logicians at the time were taking a *meta-perspective*: they were starting to study mathematics itself. Initially, their aim was to provide mathematics with a firm foundation upon which all subsequent developments could be built. They had to develop new mathematical techniques for doing so, however, simply because nobody had ever done it before. The axiomatic method can be seen as an exponent of this process [136]. In light of this it is perhaps not a coincidence that Arrow had taken a logic course taught by one of the pioneers of modern logic, Alfred Tarski (1901-1983), while being an undergraduate at the City College of New York [128].

[9] This assumption, that each voter can pick any ranking she likes, may not be as innocent as it looks. Some people in social choice in fact take it as an extra axiom, called the *universal domain axiom.* We will later see the reason for that, but in brief: omitting it can under some circumstances help to avoid the impossibility theorem.

[10] In equivalent terms: the election scheme follows the Condorcet method.

unbelievable, astonishing, theorem" [120]. Arrow was awarded the Nobel Prize in 1972.

Interestingly, Arrow's original proof contained a minor error [12]. A vast number of proofs have appeared since. Until Gil Kalai's proof from 2002 [68], most, if not all, were in essence similar in nature: it is all about defining and manipulating profiles[11] in a "smart" way. Kalai's proof is, to our knowledge, the first proof to take a quantitative approach.

1.2.4 Following Arrow

Ever since Arrow's innovation, a great deal of research in social choice theory has sought ways to cope with the impossibility theorem.

A great deal of discussion takes place on the level of *interpretation*. Here a crucial point seems to be that one should consider *the domain of application* of Arrow's social choice function. Several researchers have argued that ordinal preferences as such do not satisfactorily capture at least some of society's choice procedures. For example, Amartya Sen (born 1933), winner of the Nobel Prize in 1998, has done plenty of research in this area, emphasizing that when applying the Arrovian framework in the case of welfare economics (rather than voting theory per se), ordinal preferences alone are not adequate for making fair social choices: further information besides ordinal preferences is necessary. See for example Sen's Nobel Prize lecture [119], that has the telling title *The Possibility of Social Choice*. These discussion, spanning from economics to political theory, are appealing in their own right. See, e.g., [120, 122, 117] for more in-depth discussions.

1.2.5 Link with Computer Science

In the early 1990s the idea emerged to link social choice with computer science—in particular, computational complexity—and moreover to exploit this connection [4]. Hence the field of *computational* social choice arose.[12] The basic idea is that, while Arrow's Theorem is relevant, it is an entirely mathematical, theoretical, construct; our increasingly technological society advances the use of computers more and more, and therefore we should also take practical considerations into account. For example, is it easy to compute the winner using some particular election scheme? It turns out that deciding whether a given candidate would be selected by Dodgson's rule in a given voting situation, is computationally hard.[13] There are several other voting rules for which the same is true. One tool, also from theoretical computer science, that can be used to resolve this, is approximation [103, 19].

The idea of using complexity theory is particularly relevant in the context of the manipulation of voting schemes. The Gibbard-Satterthwaite Theorem, which one should think of as an analogon of Arrow's Theorem, roughly speaking

[11] A *profile* is the collection of all voter's preferences; we will be more precise further on.

[12] According to [17], the name "computational social choice" was explicitly used for the first time only in 2006.

[13] In the language of computer science we call this NP-hard.

says that voting manipulation is always possible.[14] The suggestion is that computational hardness can then be used to resist strategic manipulation, the idea being that if the manipulator's computer cannot handle the computations, he is forced to abandon his manipulating attempts altogether [4, 103, 17, 39]. One of the problems with this approach, however, is that computational hardness is a worst-case notion; it does not tell us anything about typical instances of the problem. The real question is thus: is manipulation easy or hard, *on average*? In Chapter 5 we review the recent developments. The results are remarkable.

Computer scientists have proposed other methods as well to circumvent the Gibbard-Satterthwaite Theorem. One of the ideas involves using randomized voting rules. Pass and Birrell [9] show that some of the common voting rules can be well-approximated by such randomized schemes, and in this way they obtain an "approximately strategy-proof" rule.

1.2.6 Brief History of Fourier Analysis on the Boolean Cube

Fourier analysis on the Boolean cube studies Boolean functions via analytic means, and in particular via the Fourier expansion [97]. Roughly speaking the crucial idea is that some interesting properties can be very easily inferred from the Fourier coefficients. Therefore, to know the Fourier coefficients is to know those interesting properties. We will see plenty of evidence for this statement throughout the thesis, starting from Chapter 2.

Fourier analysis of real-valued Boolean functions was first studied by Walsh, around 1923 [147]. Boolean analysis as a mathematical discipline developed steadily over the years. One of the major advancements is Bonami's work in the seventies. She proved an important hypercontractivity result [16].[15]

It is interesting to note that, for a relatively long time, the use of Boolean analysis *in computer science* seems to have developed in a largely independent way compared to the above evolution. A Boolean function was originally called "switching function" by engineers; in the late 1930s they first realized the usefulness of studying such functions (see, e.g., [91, 123]). It was only in 1959 when Golomb [59] recognized the connection with Walsh's work. As a result, the study of then-called "Fourier-Walsh analysis" flourished in the early 1970s, when several symposia took place.

After a relative silence, the use of Boolean analysis in theoretical computer science was reinvigorated by Kahn, Kalai, and Linial in 1988 [66], when they published *The Influence of Variables on Boolean Functions*. This highly influential article includes what is now generally known as the *Kahn-Kalai-Linial Theorem*. Roughly speaking, this theorem says that for any binary voting scheme which gives each candidate equal probability of winning, there exists a voter having disproportionally high influence. We will prove this theorem in Subsection 3.3.3.

[14] In the sense that there will always be voters having an incentive to misrepresent their votes. We deal with the Gibbard-Satterthwaite Theorem in much more detail in Chapter 5.

[15] We will review this result in Subsection 3.3.1.

As already noted above, many concepts had been studied first in the mathematics (or other) literature before they eventually found their way to the field of theoretical computer science. For example, the notion of *influence* was first introduced by the geneticist Penrose [101] (and independently by several others, such as Banzhaf [2], and Coleman [23]) in 1946; only in 1985 Ben-Or and Linial [7] introduced the concepts to the area of theoretical computer science. As another example, the *total influence* of Boolean functions, which we will see in Subsection 3.1.3, has long been studied in combinatorics.[16]

A very important concept is *noise stability*, which we will introduce in Section 3.2. This concept is easily explained in the language of voting theory: if all voters vote randomly, and if for each voter there is a fixed, small, probability that the vote is misrecorded, then what is the probability that the amalgamation of all these errors impacts the outcome of the voting process? Interestingly, although Kahn-Kalai-Linial [66] studied noise sensitivity (without giving it a name) already in 1988, the term "noise sensitivity" was coined only ten years later, by Benjamini-Kalai-Schramm [8]. A centerpiece result regarding noise stability and voting is the *Majority is Stablest Theorem* [86]. Although this theorem lies beyond the scope of this thesis, it certainly deserves mentioning. Roughly, it says the following: among all unbiased functions with small influences (which is a condition any "reasonable" voting rule should satisfy), the majority function has the largest noise stability.

In 2002, Gil Kalai obtained a Fourier-theoretic proof of Arrow's Theorem [68]. This result was particularly important in that it not only shed a different light on the impossibility theorem, but it also managed to strengthen it. We will review Kalai's proof in Subsection 4.2.3. Additionally, once again showing its high impact, this paper eventually led to a strengthening of the Gibbard-Satterthwaite Theorem as well. The most up to date result is due to Mossel and Rácz [88], from 2012. Its rather technical proof is, unfortunately, way beyond the scope of this thesis. Nevertheless, in Chapter 5 we will discuss the theorem and its consequences.

1.3 The Impartial Culture Assumption

1.3.1 Motivation

When mathematically modeling voting processes, we need a model for the way the voters are voting. How can we do this? In the absence of any other information, that is to say "in the abstract", perhaps the best we can do is the following: *the preferences of the voters are uniformly random and independent.* This assumption is called the *impartial culture assumption* and is abbreviated by ICA.

More generally, a *culture* [140] is an assumption about how voters vote, i.e., the distribution of their votes. Other cultures are considered in the literature. We just mention a few without going into them. Besides ICA,

[16]The reason is that it happens to be equivalent to edge-boundary size for subsets of the Hamming cube.

another basic model from the literature is the *impartial anonymous culture assumption* [70, 56]. The latter is based on the presentation of voter preferences by anonymous profiles, in which the names of the voters are neglected. It assumes that each resulting anonymous profile class is equally probable. A recent approach [36] introduces the *impartial, anonymous, and neutral culture assumption*.

The impartial culture assumption is a well-known supposition in social choice theory. Although first proposed by Guilbaud [61] in 1952, it was first formalized only by Garman and Kamien [52] in 1968.

A possible way to think about the impartial culture assumption is as follows. In an election one can imagine there is some subset of voters who have already determined for whom to vote; this group's votes we could just factor in as a constant into the election function. Having dealt with the decided voters we remain with the undecided voters, the so-called *swing voters*. Perhaps these voters' preferences can be thought of as being independent and uniformly at random.

From a purely mathematical perspective the impartial culture assumption is before all just an assumption which is *convenient*: it makes computations and analysis a lot easier. In other words it makes the mathematics manageable. In particular the assumption the voters' votes be independent facilitates the calculations involving probabilities tremendously. Of course, this motive is merely pragmatic and opportunistic. It does not say a thing about *why* this mathematical model would be correct; in fact it completely ignores this question altogether.

Before going into a discussion from political science related to the impartial culture assumption, one issue about ICA we should be very clear about right from the start is the following. In Chapter 6 we prove Arrow's Theorem using Fourier analysis. As will then become apparent, in that proof we calculate some probability under the assumption of ICA—in fact, this will be the main idea of the proof. However, the end result (namely the classical version of Arrow's Theorem) is completely *independent* of the underlying model how the voters vote. Having said that, several other results in this thesis do depend on the impartial culture assumption. When we want to calculate some occurrence probability of obtaining a *cycle*, for example, we necessarily need a probabilistic voting model. In this thesis we will then take ICA as the model. Unavoidably, this limits the scope of interpretation of those results. Indeed, perhaps par for the course, the impartial culture assumption is a somewhat unrealistic model. For the purposes of this thesis, it is in this context important to note that both the quantitative version of Arrow's Theorem (see Subsection 4.2.3), as well as the quantitative version of the Gibbard-Satterthwaite Theorem (see Subsection 5.3.1), do depend on ICA.

1.3.2 Liberalism Against Populism

It should not be surprising that the impartial culture assumption is rather controversial. An interesting and influential (though contentious) account is

Liberalism Against Populism [111], a book by the American political scientist William H. Riker (1920-1993) published in 1982. In this book Riker develops an "antidemocratic" interpretation of results from social choice, especially Arrow's Theorem. We used quotation marks here because in general it is not quite clear what the term *antidemocratic* exactly means: it can mean several things. Furthermore, in political contexts this word is used in a derogatory way to belittle political opponents. We will explain further on what is meant by the term.

Because of their relevance for this thesis as well as to illustrate that the often abstract and theoretical considerations in this thesis also relate to discussions about the real world, we briefly sketch Riker's viewpoints and those of his perhaps greatest opponent, Gerald L. Mackie. Our observations are based on [111, 78, 79, 75, 149, 80, 26, 24].

Riker questions the role of *voting* in democracy. Which purpose does it serve? There are two major views. According to the first one, the *liberal* view, citizens vote only in order to control (meaning select, punish, replace) elected officials. In this viewpoint voting can be seen as just a *negative liberty*: it yields liberty from coercion, especially by the state. A second viewpoint is the *populist* view. It was initiated by Jean-Jacques Rousseau (1712-1778), and states that citizens vote to establish the will of the electorate. The elected officials are in this respect a direct extension of the will of the people. As the title of his book suggests, Riker claims that the liberal viewpoint is the only valid one: "the people" cannot rule as a corporate body. Quoting Riker [111], "The function of voting is to control officials, *and no more*" (emphasis in original). To argue for his cause, Riker uses arguments from social choice theory. Voting as a populist means of representation, he says in [111], is

- *inaccurate*, because different voting systems yield different outcomes from exactly the same profile of individual voters' preferences;
- *meaningless*, since the outcome of voting is always manipulable by the Gibbard-Satterthwaite Theorem (see Chapter 5), and moreover it is impossible to distinguish manipulated from unmanipulated outcomes because of the unknowability of private intentions underlying public actions.

In summary: according to Riker populist democracy is *incoherent*. Riker's position can be viewed as antidemocratic since, as believed by him, voting is the central act of democracy (see page 5 in [111]):

> *(...) I want to point out that the coherence depends on the fact that all democratic ideas are focused on the mechanism of voting. All the elements of the democratic method are means to render voting practically effective and politically significant, and all the elements of the democratic ideal are moral extensions and elaborations of the features of the method that make voting work. Voting, therefore, is the central act of democracy, and this fact makes apparent the immediate relevance of the theory of social choice.*

To support the first claim, Riker uses Arrow's Theorem. The main idea is that because of this theorem no method of aggregating individuals' transitive preference orderings can guarantee a collective preference ordering that is transitive: we can always get a *cycle* in some instances (we will see this in detail in Chapter 6). In that case it is unclear which outcome should be elected as being the winner. Furthermore, Riker argues that cycles in reality are almost universal. But even leaving the possibility of cycles aside, Riker finds it problematic in itself that different voting rules can yield different election outcomes. Interestingly, some of Riker's arguments for the inaccuracy assume the impartial culture assumption. For example, he provides a summary table concerning the likelihood of a cycle, under assumption of ICA. According to this table, cycles are ubiquitous. We will elaborate on this discussion in Subsection 4.3.2.

In his book *Democracy Defended* [79] from 2003 as well as in [78], Mackie vigorously responds to Riker's criticisms. In his analysis Riker provided lots of examples from American politics which he thought illustrated his points. An example worth mentioning because of its historical importance is the 1860 election of president Lincoln. Back then the major controversy in the US was the extent to which slavery would be allowed in the vast new territories, the so-called "free soil" question. Riker used this specific example to show that according to different voting rules a different president would have been elected,[17] and that under pairwise majority comparison (which we will call a *Condorcet election* based on the majority rule in the next chapter) there was a cycle.

Briefly stated, in response to Riker's book Mackie argues that

- cycles are empirically improbable, as evidence shows;
- different reasonable election procedures "often" yield the same results;
- though manipulative voting can arise, the people's genuine preferences emerge over time;
- much of Riker's empirical evidence is either speculative, mistaken, or misinterpreted.

In conclusion, Mackie claims voting can be both accurate as well as meaningful, and Riker's arguments are no reason to reject the populist viewpoint.

Mackie argues that Riker's viewpoints are antidemocratic. What does he mean exactly by that? On page 3 in *Democracy Defended* [79], we read:

[17] We should be a bit more precise here. In the US, each voter elects his unique favorite presidential candidate. However, what Riker did was, using political arguments, to speculate about *how the US voters would have voted had they instead all submitted a complete ranking of the four candidates*, rather than just one—in the terminology from Subsections 4.1.1 and 5.2.1, Riker thus imagined the US' means of voting was modelled as a social welfare function rather than just a social choice function. Then, using this (speculative) new information, he argued that another president would have been elected: often, though not always, Stephen A. Douglas.

> *Riker calls populist any democratic theory which depends on a systematic connection between the opinion or will of the citizens and public policy, and liberalist any democratic theory which requires only that voting result in the random removal of elected officials. Riker rejects populist democracy as infeasible, and offers his liberalist democracy in its place. What almost everyone means by democracy is what Riker calls populist democracy; and, I shall argue, Riker's liberalist alternative fails, descriptively and normatively. Thus, I am tempted to label his doctrine antidemocratic. I believe that it is antidemocratic in consequence, whether or not it is antidemocratic in spirit. But to use such a label throughout this volume would be tendentious. To call his doctrine antipopulist, though, is to beg the question in his favor: the word populism has many negative connotations, and I do not mean to defend such things as Peronism, short-sighted policy, or mob rule. Since Riker's claim is that in the political sphere the rational individual opinions or desires of citizens cannot be amalgamated accurately and fairly, it is apt to describe his doctrine as one of democratic irrationalism. Riker's irrationalist doctrine emphasizes principled failings of democracy and recommends a constitutionalist libertarianism and the substitution of economic markets for much of political democracy (Riker and Weingast 1988).*

A relevant and related question is: how omnipresent are cycles *actually*, in practice? Several researchers in the literature (see e.g. [54, 55, 36, 140, 106], and Chapter 4 in [76]) have explored this question. The short answer is: perhaps quite surprisingly, they are rare. We will come back to this issue in Subsection 4.3.2.

Summarizing this subsection: the impartial culture assumption is first and foremost a mathematically convenient yet unrealistic model for the way voters are voting. Given quantitative (probabilistic) results in social choice, one should always keep in mind which theoretical model underlies it, and interpret results appropriately in light of that. Riker's account shows that, furthermore, we should be careful to interpret results from formal sciences in an appropriate way when applying it to the social sciences. Indeed, as mentioned, a good part of his argument unduly relies on the rather unrealistic impartial culture assumption.

1.4 Outline of the Thesis

CHAPTER 2. We introduce the two main characters of this thesis: the Boolean function and its Fourier decomposition. Some basic theorems, such as Plancherel's Theorem and Parseval's Theorem, are proved. We end the chapter by making the connection with social choice theory.

CHAPTER 3. The notions of influence and noise, paramount for the theory of

Boolean analysis, are presented. The Hypercontractive Inequality is stated. This is a famous inequality from Boolean analysis. We prove only a special case, called Bonami's Lemma. This special case is, however, already sufficiently strong to imply two interesting theorems: the Kahn-Kalai-Linial Theorem, and Friedgut's Theorem. The latter theorem, for example, says that in a two-candidate election with low total influence there is always a small coalition of voters which are able to determine the election with high probability.

CHAPTER 4. In this chapter we come to one of the main results from this thesis: Arrow's Theorem. Using the results we will already have established by then, its proof will be remarkably easy. The main idea, due to Gil Kalai [68] is to calculate the probability of having a Condorcet cycle under ICA. What is more, this proof method leads to a strengthening of Arrow's Theorem for three candidates. Roughly stated, the more we want to avoid Condorcet's paradox, the more the election scheme will look like a dictator.

CHAPTER 5. We state the Gibbard-Satterthwaite Theorem. This theorem concerns the manipulation of elections. Initiated by the quantitative version of Arrow's Theorem obtained by Kalai, follow-up research [65, 88] has led to a quantitative version of the Gibbard-Satterthwaite Theorem as well. The proofs are heavily inspired by Boolean analysis, but the strongest results obtained to date involve also several other advanced mathematical techniques which are beyond the scope of this thesis. For that reason we limit ourselves to giving an overview of the current state of affairs, and its implications for social choice.

CHAPTER 6. Inspired by Kahneman's *Thinking, Fast and Slow* [67], in this last chapter we come up with a simple model to simulate the various biases that show up in small meetings in which people have to vote, one after the other, "yes" or "no" for a given proposal.

Chapter 2

Fourier Analysis on the Boolean Cube

In this chapter, starting from the very basics we develop Fourier analysis for real Boolean functions. This chapter will stand out in the sense that its focus will be quite mathematical. Still, we try to provide for some intuitions too.

It should be noted first that more generally one could develop the theory of Fourier analysis for any finite Abelian group. In that sense our treatment is not the most general one in the mathematical sense. Nevertheless, for our purposes only the case $\mathbb{Z}_2^n$ is important so we will focus solely on the latter.

The set $\{-1,1\}^n$, where $n \geq 1$ is a natural number, is called a *Boolean cube*. Likewise, any set of the form $\{0,1\}^n$ is called a Boolean cube. In fact these are just two representations of the same thing. In computer science TRUE is commonly represented by 1 while FALSE is represented by 0. However, in this thesis we will usually let TRUE be represented by -1 while FALSE be represented by 1. That is, 0 corresponds to 1, whereas 1 corresponds to -1; the reasons for doing so will become apparent later. One should think of -1 and 1 as real numbers, so that any element of $\{-1,1\}^n$ can be thought of as an element of $\mathbb{R}^n$, a *vector* in the linear algebraic sense. We will freely pass between these representations of the Boolean cube. *Usually* it is most convenient for us to use the $\{-1,1\}^n$ representation, so all definitions and theorems in this chapter will feature this representation, although one should keep in mind that everything can similarly be done for $\{0,1\}^n$ as well.

Definition 2.1

> A *function on the Boolean cube* is a function from $\{-1,1\}^n$ to $\mathbb{R}$ where n is some natural number.

We also say that a function is a *Boolean function* to say it is a function on the Boolean cube. A function on $\{-1,1\}^n$ which takes values in $\{-1,1\}$ is called *Boolean-valued*.

All the definitions and results from this chapter are based on [97, 126, 63, 95, 96, 28]. Unless specifically mentioned, all our results come from these resources.

2.1 Introduction

Given a Boolean function, we would like to find an elegant way to represent that function. The Fourier expansion will establish exactly that. The underlying idea is very simple: we want to write a Boolean function as a multilinear polynomial. (*Multilinear* means that no variable appears squared, cubed, or as any higher power, i.e., each variable has local degree one everywhere.) The following examples should clarify this idea.

Example. Consider the function $\mathrm{Maj}_3 : \{-1,1\}^3 \to \{-1,1\}$ which takes the majority of the three bits, so e.g. $\mathrm{Maj}_3(1,-1,1) = 1$ while $\mathrm{Maj}_3(-1,-1,1) = -1$. It is easy to verify that for all $x \in \{-1,1\}^3$ we have

$$\mathrm{Maj}_3(x) = \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 - \frac{1}{2}x_1x_2x_3.$$

The latter polynomial is the Fourier expansion of Maj_3.

Example. Consider the function $\min_3 : \{-1,1\}^3 \to \{-1,1\}$ taking the minimum of three bits. Since -1 corresponds to TRUE and 1 to FALSE, this function represents the logical OR. Then it is easy to verify that for all $x \in \{-1,1\}^3$ we have

$$\min_3(x) = -\frac{3}{4} + \frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{4}x_3 + \frac{1}{4}x_1x_2 + \frac{1}{4}x_2x_3 + \frac{1}{4}x_1x_3 + \frac{1}{4}x_1x_2x_3,$$

so this polynomial is the Fourier expansion of $\min_3$.

In fact we will see later that not only every Boolean function has such a polynomial representation, but that representation is in fact *unique.*

2.2 Finding the Fourier Expansion

How did we find those polynomials from the examples? There is an easy way: interpolation. For example take again the first example, the function Maj_3. First we want to "make it right" on the coordinate $(-1,-1,1)$, on which the function should be -1. Well, the polynomial

$$(-1)\left(\frac{1-x_1}{2}\right)\left(\frac{1-x_2}{2}\right)\left(\frac{1+x_3}{2}\right)$$

does just that: it is -1 on $(-1,-1,1)$ and 0 on any other coordinate. Now it only remains to do this for the other coordinates, after which we just have to add the polynomials we had obtained; after some calculations we then indeed obtain a multilinear polynomial.

As an example, consider the function $\max_2 : \{-1,1\}^2 \to \{-1,1\}$, which takes the maximum of the bits. Its Fourier expansion is obtained via

$$(+1)\left(\frac{1+x_1}{2}\right)\left(\frac{1+x_2}{2}\right) + (+1)\left(\frac{1+x_1}{2}\right)\left(\frac{1-x_2}{2}\right)$$
$$+(+1)\left(\frac{1-x_1}{2}\right)\left(\frac{1+x_2}{2}\right) + (-1)\left(\frac{1-x_1}{2}\right)\left(\frac{1-x_2}{2}\right),$$

which after some calculations can be seen to be equal to $\frac{1}{2}+\frac{1}{2}x_1+\frac{1}{2}x_2-\frac{1}{2}x_1x_2$.

We thus have the following:

THEOREM 2.2

Every $f:\{-1,1\}^n\to\mathbb{R}$ can be expressed as a multilinear, real, polynomial in the variables $x=(x_1,\ldots,x_n)$. That is, for each $S\subseteq[n]$ there is a $\lambda_S\in\mathbb{R}$ such that for all elements $x\in\{-1,1\}^n$ we have

$$f(x)=\sum_{S\subseteq[n]}\lambda_S x^S,$$

where with x^S we mean $\prod_{i\in S}x_i$ if S is nonempty, and where $x^{\emptyset}$ is 1. Moreover, this representation is unique.

Proof. The existence we already explained. Uniqueness will follow by an easy linear algebra argument that we will see later. ■

We then define the Fourier expansion formally:

DEFINITION 2.3

Let $f:\{-1,1\}^n\to\mathbb{R}$. For each $S\subseteq[n]$, the coefficient λ_S from the statement of Theorem 2.2 is denoted by $\widehat{f}(S)$ and is called *the Fourier coefficient of f on S*. The right-hand side of the expression

$$f(x)=\sum_{S\subseteq[n]}\widehat{f}(S)x^S$$

is called the *Fourier expansion of f*.

So for example we have

$$\widehat{\mathrm{Maj}_3}(\emptyset)=0,\quad \widehat{\mathrm{Maj}_3}(\{1\})=\widehat{\mathrm{Maj}_3}(\{2\})=\widehat{\mathrm{Maj}_3}(\{3\})=\frac{1}{2},\quad \widehat{\mathrm{Maj}_3}(\{1,2,3\})=-\frac{1}{2}.$$

2.3 The Importance of the Fourier Expansion in a Slogan

One could ask what is the point of having the Fourier expansion of a Boolean function. In computer science applications, the Boolean function often represents some combinatorial object, e.g. the operation of a circuit, the concept class in machine learning, a set system in extremal combinatorics, or, which will be particularly interesting for us, perhaps a voting rule. It turns out a lot of interesting combinatorial properties of Boolean functions get encoded by what the Fourier coefficients are. For example in the context of voting we will see the important notion of *influence* of a particular voter on the voting scheme; this highly interesting combinatorial quantity is nicely expressible in terms of the Fourier coefficients, as we will see. Therefore, to know the Fourier coefficients is to know a lot, if not all, of interesting properties of the Boolean function at hand.

The rest of this thesis should be read with the following slogan in mind, the value and importance of which the reader hopefully will become convinced in due course:

> *The Fourier coefficients of a function bear a lot of information about interesting properties regarding that function.*

This thesis as a whole—using Boolean analysis to investigate properties from social choice theory—can be seen as one big chunk of evidence supporting this slogan.

2.4 Connection with Linear Algebra

Each monomial $x^S = \prod_{i \in S} x_i$, where $S \subseteq [n]$, is itself a Boolean function: it is -1 if and only if the number of S-variables being -1 is odd.

DEFINITION 2.4

For $S \subseteq [n]$ we let $\chi_S : \{-1,1\}^n \to \{-1,1\} : x \mapsto \prod_{i \in S} x_i$, called the *parity function on S*. If $S = \{i\}$ we write this also as χ_i and call it the *i-th dictator function.*

The function $x \mapsto -\chi_i$, where $i \in [n]$, is also called the *i-th anti-dictator function.*

With the notation from Definition 2.4 we can write the result from Theorem 2.2 as

$$f = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S$$

for each Boolean function f; i.e., f is a linear combination of the functions $\chi_S, S \subseteq [n]$. This makes us think about the situation in linear algebra terms. Particularly, do the functions $\{\chi_S \mid S \subseteq [n]\}$ form a basis? We will see that they do.

First, let V be the function space consisting of all Boolean functions, where addition and scalar multiplication are pointwise, so $V = \{f \mid f : \{-1,1\}^n \to \mathbb{R}\}$.

Note that any Boolean function can be thought of as a vector in $\mathbb{R}^{2^n}$: indeed, just take its truth table (for some arbitrary but fixed order of the coordinates) and stack them into a column vector. This shows that V and $\mathbb{R}^{2^n}$ are isomorphic. Therefore the dimension of V is 2^n. But there are exactly 2^n parity functions, and Theorem 2.2 shows that they span the space, so they must be linearly independent as well. Thus they form a basis of V, called the *Fourier basis.* This finishes the proof of Theorem 2.2.

In the natural way we have an inner product on V too:

Definition 2.5

For $f, g : \{-1,1\}^n \to \mathbb{R}$ we let $\langle f, g\rangle \stackrel{\text{def}}{=} 2^{-n} \sum_{x\in\{-1,1\}^n} f(x)g(x)$.

The 2^{-n} factor was put there only in order to get an "average" value, since there are in total precisely 2^n terms. It is just a rescaling. In this way V becomes an inner product space.

Note that we can equivalently state the definition as follows:

$$\langle f, g\rangle = \mathop{\mathrm{E}}_{x\sim\{-1,1\}^n}[f(x)\cdot g(x)]$$

where the notation $x \sim \{-1,1\}^n$ denotes that x is drawn uniformly at random, i.e., $x_1, \ldots, x_n$ are independently ± 1 with probability $\frac{1}{2}$ each. In what follows, unless stated otherwise, all expectations are with respect to such a uniform distribution on $\{-1,1\}^n$. We often simply write $\mathrm{E}_x[f(x)g(x)]$ or even $\mathrm{E}[f(x)g(x)]$ for brevity.

If $f, g : \{-1,1\}^n \to \{-1,1\}$ are Boolean-valued, the inproduct $\langle f, g\rangle$ in a way measures how similar the functions f and g are. It measures their "correlation": if e.g. $f = g$ then $\langle f, g\rangle = 1$, and if $f = -g$ then $\langle f, g\rangle = -1$. More generally by a very easy calculation it follows that

$$\langle f, g\rangle = 1 - 2\Pr_x[f(x) \neq g(x)]$$

where again $x \sim \{-1,1\}^n$. Consequently, the closer to 1 the quantity $\langle f, g\rangle$ is, the more the two functions agree.

Definition 2.6

For $f, g : \{-1,1\}^n \to \mathbb{R}$ we put $\mathrm{dist}(f, g) \stackrel{\text{def}}{=} \Pr_x[f(x) \neq g(x)]$, which we call the *fractional Hamming distance of f and g*. If $\varepsilon > 0$ and $\mathrm{dist}(f, g) \leq \varepsilon$, then we say f and g are *ε-close.*

We have

$$\langle f, g\rangle = 1 - 2\,\mathrm{dist}(f, g). \tag{2.1}$$

As usual in an inner product space, for any $f : \{-1,1\}^n \to \mathbb{R}$ we have the 2-*norm of* f which is defined as $||f||_2 \stackrel{\text{def}}{=} \sqrt{\langle f, f\rangle}$. Note that by the above Boolean-valued functions have norm 1, i.e., they are unit vectors in this function space.

Lemma 2.7

The set $\{\chi_S \mid S \subseteq [n]\}$ is an orthonormal basis of the inner product space $(V, \langle\cdot,\cdot\rangle)$, that is, we have

$$\langle \chi_S, \chi_T\rangle = \delta_{S,T} = \begin{cases} 1 & \text{if } S = T \\ 0 & \text{if } S \neq T \end{cases}$$

for all $S, T \subseteq [n]$.

Proof. If $S = T$, then clearly $\langle \chi_S, \chi_S \rangle = \mathrm{E}[\chi_\emptyset] = \mathrm{E}[1] = 1$. For the other part, suppose $S \neq T$. Note that $\chi_S \cdot \chi_T = \chi_{S \Delta T}$, where $S \Delta T$ is the symmetric difference of S and T. Therefore the lemma follows immediately by the following claim: for any $A \subseteq [n]$ with $A \neq \emptyset$ we have $\sum_{x \in \{-1,1\}^n} \chi_A(x) = 0$.

To show this, pick any $a \in A$. For any $x \in \{-1,1\}^n$, let $x' \in \{-1,1\}^n$ be the same as x except at spot a it has the other bit, i.e. $x'_a = -x_a$ and $x'_i = x_i$ for all $i \neq a$. Clearly $\chi_A(x) = 1$ if and only if $\chi_A(x') = -1$, so taking the sum over all $x \in \{-1,1\}^n$ gives 0 as a result.

Another proof is as follows. If $S \neq \emptyset$, then by independence of all x_i with $i \in S$ we get $\mathop{\mathrm{E}}_x[\prod_{i \in S} x_i] = \prod_{i \in S} \mathop{\mathrm{E}}_x[x_i] = 0$, as $\mathop{\mathrm{E}}_x[x_i] = (-1)\,\frac{1}{2} + (+1)\,\frac{1}{2} = 0$ for all i. ∎

Note that Lemma 2.7 implies the uniqueness from Theorem 2.2 that we still had to show.

The following theorem is important since it gives a concrete formula to calculate the Fourier coefficients.

Theorem 2.8

For $f : \{-1,1\}^n \to \mathbb{R}$ and $S \subseteq [n]$ we have $\widehat{f}(S) = \langle f, \chi_S \rangle$.

Proof. Using Lemma 2.7 and writing f in terms of the Fourier basis we have

$$\langle f, \chi_S \rangle = \langle \sum_{T \subseteq [n]} \widehat{f}(T) \chi_T, \chi_S \rangle = \sum_{T \subseteq [n]} \widehat{f}(T)\, \langle \chi_T, \chi_S \rangle = \widehat{f}(S),$$

proving the theorem. ∎

2.5 Theorems of Plancherel and Parseval

The following two theorems are very important and will be regularly used.

Theorem 2.9

*(*Plancherel's Theorem*) We have*

$$\langle f, g \rangle = \sum_{S \subseteq [n]} \widehat{f}(S)\, \widehat{g}(S)$$

for all $f, g : \{-1,1\}^n \to \mathbb{R}$.

Proof. We write f and g as linear combinations of the Fourier basis vectors. Using linearity of the inner product we then get

$$\langle f, g \rangle = \langle \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S \,, \sum_{T \subseteq [n]} \widehat{g}(T) \chi_T \rangle = \sum_{S,T \subseteq [n]} \widehat{f}(S) \widehat{g}(T) \langle \chi_S, \chi_T \rangle = \sum_{S \subseteq [n]} \widehat{f}(S) \widehat{g}(S).$$

∎

THEOREM 2.10

> (PARSEVAL'S THEOREM) *For* $f : \{-1,1\}^n \to \mathbb{R}$ *we have*
>
> $$||f||_2^2 = \langle f, f\rangle = \mathrm{E}_x[f^2] = \sum_{S\subseteq[n]} \widehat{f}(S)^2.$$
>
> *In particular, we have*
>
> $$\sum_{S\subseteq[n]} \widehat{f}(S)^2 = 1$$
>
> *for any Boolean-valued function, i.e. a function* $f : \{-1,1\}^n \to \{-1,1\}$.

Proof. This follows immediately from Plancherel's Theorem by taking $f = g$. ■

Therefore, the squares of the Fourier coefficients of a Boolean-valued function always add up to 1: e.g. in the first example from above, where we considered Maj_3, we indeed have $4 \cdot (1/2)^2 = 1$.

NOTE ON NOTATION. So far we have focused on the Boolean cube being represented by $\{-1,1\}^n$. But what if we instead consider it to be of the form $\{0,1\}^n$?

Let us consider a function $f : \{0,1\}^n \to \mathbb{R}$. By defining $\chi_S : \{0,1\}^n \to \{-1,1\}$ by $\chi_S(x) \stackrel{\text{def}}{=} (-1)^{\sum_{i\in S} x_i}$ in this case, we instantly see that all results in this chapter from Subsection 2.4 onwards remain true. Specifically we can write the Fourier expansion of $f : \{0,1\}^n \to \mathbb{R}$ again as

$$f = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S.$$

Note that $\chi_S(x) = \prod_{i\in S}\chi(x_i)$ where $\chi(b) = (-1)^b$ for $b \in \{0,1\}$. This explains why mathematically it is natural to have $-1 \in \{-1,1\}$ correspond to $1 \in \{0,1\}$ (representing TRUE) and $1 \in \{-1,1\}$ correspond to $0 \in \{0,1\}$ (representing FALSE).

2.6 Fourier Weights

Consider a Boolean-valued function. The squares of its Fourier coefficients then sum up to 1. We can think about these nonnegative numbers as a probability distribution, although we prefer to call them *weights*:

DEFINITION 2.11

> Given $f : \{-1,1\}^n \to \mathbb{R}$, the *weight of* f *on* $S \subseteq [n]$ is $\widehat{f}(S)^2$.

These weights contain a great deal of combinatorial information of the Boolean-valued function at hand, and will therefore play an extremely important role in what follows.

Intuitively one can think of f's weight at S as a quantitative measure indicating how important the set of coordinates S is to the function f; for a Boolean-valued function the weight is a number belonging to $[0, 1]$.

DEFINITION 2.12

> Given $f : \{-1, 1\}^n \to \mathbb{R}$, the *mean of* f on $S \subseteq [n]$ is $\mathrm{E}_x[f(x)]$. We say f is *unbiased* if its mean is 0.

For Boolean-valued functions the mean measures how "biased" the function is towards $+1$ or -1. By Theorem 2.8 it follows that the mean of a function $f : \{-1, 1\}^n \to \mathbb{R}$ can be nicely read off from its Fourier coefficients: it equals $\widehat{f}(\emptyset)$, and we have again an example the emergence of the slogan from earlier. For the example where we considered $\min_3$, it is clear that its mean is $(+1)\,\frac{1}{8} + (-1)\,\frac{7}{8} = -\frac{3}{4}$, which is equal to $\widehat{\min_3}(\emptyset)$.

Now we give another example illustrating the ample importance of the Fourier coefficients, and more distinctly the weights. Given a function $f : \{-1, 1\}^n \to \mathbb{R}$, consider the random variable $f(x)$ where $x \sim \{-1, 1\}^n$. Furthermore let $\mathrm{Var}(f)$ be the *variance* of the random variable $f(x)$, i.e., $\mathrm{Var}(f) = \mathrm{E}_x[f(x)^2] - \mathrm{E}_x[f(x)]^2$. For a Boolean-valued function this number is in $[0, 1]$. It measures how "spread out" a function is, that is, how much it varies; for example if f is constant then its variance is 0, while f being unbiased implies that its variance is maximally, namely 1. For Boolean-valued functions variance measures unbiasedness, an important property a function might have. Parseval's Theorem together with $\mathrm{E}_x[f(x)]^2 = \widehat{f}(\emptyset)^2$ immediately implies that $\mathrm{Var}(f) = \sum_{S \neq \emptyset} \widehat{f}(S)^2$. The variance of a function is hence nicely expressible in terms of the Fourier weights: once again we have found that an interesting quantity (the variance) becomes apparent just by looking at the Fourier coefficients.

DEFINITION 2.13

> Let $f : \{-1, 1\}^n \to \mathbb{R}$. For $0 \le k \le n$, the *weight of f at degree k* is
>
> $$\mathrm{W}^k[f] \stackrel{\text{def}}{=} \sum_{|S|=k} \widehat{f}(S)^2.$$
>
> We also put $\mathrm{W}^{\le k}[f] \stackrel{\text{def}}{=} \sum_{i=0}^{k} \mathrm{W}^i[f]$.

Note that Parseval's Theorem, that is Theorem 2.10, implies that for any function $f : \{-1, 1\}^n \to \{-1, 1\}$ we have

$$\mathrm{W}^0[f] + \mathrm{W}^1[f] + \ldots + \mathrm{W}^n[f] = 1. \tag{2.2}$$

The weight of a function up to a degree is a good measure of the complexity of a Boolean-valued function. Its total weight is, as we know, equal to 1, but the question is at what degrees those weights are mainly present. In some way, the higher the degree the weight is at, the more complex the Boolean function is. Functions with much of their weight at low degrees are in many ways more simple; e.g., they can be efficiently learned by computational learning algorithms. See for example Section 4.2 in [81].

An example illustrating this idea which will be essential for us is the following:

Lemma 2.14

Let $f : \{-1,1\}^n \to \{-1,1\}$. If f has all its weight on degree 1, then f is a dictator or an anti-dictator function. In symbols, if $\mathrm{W}^1[f] = 1$ then $f = \pm\chi_i$ for some $i \in [n]$.

Proof. We have $\sum_{|S|=1} \widehat{f}(S)^2 = 1$, so by Parseval's Theorem we know that $\widehat{f}(S) = 0$ for all $S \subseteq [n]$ with $|S| \neq 1$. Therefore f is of the form $f(x) = a_1x_1 + \ldots + a_nx_n$ for some reals $a_1, \ldots, a_n$ satisfying $a_1^2 + \ldots + a_n^2 = 1$. It is clear that without loss of generality we may suppose each a_i is different from 0 (otherwise we can just eliminate them).

Note that

$$f(\mathrm{sgn}(a_1), \ldots, \mathrm{sgn}(a_n)) = |a_1| + \ldots + |a_n| \geq 0,$$

so, as f is Boolean-valued, we have $|a_1| + \ldots + |a_n| = 1$. Thus,

$$1 = \sum_{i=1}^{n} a_i^2 \leq (\max_i |a_i|) \sum_{i=1}^{n} |a_i| = \max_i |a_i|,$$

so there is an $i \in [n]$ such that $a_i = \pm 1$. Then, $f = \pm\chi_i$. ∎

2.7 Notions from Social Choice Theory

2.7.1 Boolean Functions as Voting Rules

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean-valued function. Then there is a simple interpretation of f as a *voting rule* (also known as *social choice function* in the social choice literature) for n voters who have to choose from two candidates -1 and 1. Indeed, an $x = (x_1, \ldots, x_n) \in \{-1,1\}^n$ can interpreted as meaning that voter i prefers alternative x_i over alternative $-x_i$. In that case $f(x)$ is the winner of the election specified by this voting rule, given that the voters voted according to $x = (x_1, \ldots, x_n)$.

We now give a list of some voting rules:

- *Majority Rule.* This rule is defined by $\mathrm{Maj}_n(x) \stackrel{\mathrm{def}}{=} \mathrm{sgn}(x_1 + \ldots + x_n)$. Here we have to assume n is odd, since we want to leave the sign of 0 undefined, reflecting the fact that in this thesis we do not want to go into tie-breaking. Furthermore, even when n is even, we call a Boolean-valued function *a majority function* if, whenever $x_1 + \ldots + x_n \neq 0$, we have $f(x) = \mathrm{sgn}(x_1 + \ldots + x_n)$.

- *Unanimity Rule.* This rule is defined by

$$\mathrm{AND}_n(x) \stackrel{\text{def}}{=} \begin{cases} -1 & \text{if } x = (-1, \ldots, -1) \\ 1 & \text{else} \end{cases}.$$

  Suppose a committee has to decide on passing a resolution. Being in favor of the proposal is indicated by -1 while being against the proposal is indicated by 1. In this selection scheme the selection committee passes the resolution only if there is unanimity over accepting it.

- *At-Least-One Rule.* Dually from AND_n, it is defined by

$$\mathrm{OR}_n(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x = (1, \ldots, 1) \\ -1 & \text{else} \end{cases}.$$

  In this selection scheme the committee passes the resolution whenever there is at least one person in favor of the proposal.

- *k-Junta Rule.* A *k-junta*, where $k \in [n]$, is a Boolean-valued function that depends on at most k coordinates. In fact this is not one rule, but a set of rules; in that sense it is as a matter of fact more a property of an election rule rather than an election rule itself.

- *Dictator Rule.* It is defined by $\chi_i(x) \stackrel{\text{def}}{=} x_i$ where $i \in [n]$. I.e., using our previous notation, χ_i is just $\chi_{\{i\}}$.

  This rule plays an especially important rule in social choice, as we will see in the next chapters. An *anti-dictator rule* is defined by $x \mapsto -x_i$ for any $i \in [n]$. Anti-dictators do not arise in practice, although for theoretical purposes it is important to have this notion; in fact from the perspective of Boolean analysis dictators and anti-dictators are very much alike.

- *Weighted Majority Rule.* A *weighted majority function* is a Boolean function for which there are $a_i \in \mathbb{R}$ such that

$$f(x) = \mathrm{sgn}(a_0 + a_1 x_1 + \ldots + a_n x_n)$$

  for all $x \in \{-1, 1\}^n$. The reals $a_0, a_1, \ldots, a_n$ have to be chosen in such a way that $a_0 + a_1 x_1 + \ldots + a_n x_n = 0$ for no x; this is necessary as we want to leave the sign of 0 undefined.

  Note that $\mathrm{Maj}_n, \mathrm{AND}_n, \mathrm{OR}_n$, as well as a dictator or anti-dictator rule are all weighted majority rules; this is easy to see by appropriately picking the a_i's.

  The idea behind this election scheme is that certain voters are more important and should thus count more. For example in the European

Union votes of representatives of different countries are weighted according to the number of people they represent. E.g., Italy has about six times as many inhabitants as Belgium, so the coefficient a_i corresponding to the Italian representative would be about six times higher than the coefficient corresponding to a Belgian representative.

Since weighted majority rules are commonly seen in practice, they are very important. It is interesting to investigate their properties using Boolean analysis. We will come back to this later.

- *The US Electoral College Rule.* Roughly speaking, the US president is chosen as follows. We assume there are just two candidates. There are 50 states, and each state chooses their preferred candidate by a plurality vote. Once this has been done, a weighted majority rule is applied to these 51 states' votes to finally elect the president. The coefficient a_i corresponding to state i is proportionate to state i's number of inhabitants.

- *The Tribes Rule.* This rule is quite artificial and is probably not used in practice. However, from the perspective of Boolean analysis it has proven to be an interesting function, e.g. to find counterexamples. Also regarding influences it is interesting; we will come back to this point later. For this rule $n = s \cdot w$, and the n voters are divided into s subgroups, each of which is called a *tribe* and is of size w. In any such tribe the unanimity rule is applied. Then the at-least-one rule is applied on the s votes. That is, a proposal under this rule is accepted if and only if at least one tribe is unanimously in favor of the proposal. It is denoted by $\mathrm{Tribes}_{s,w}$. One can show that if s, w are such that $s \sim (\ln 2)2^w$, then the scheme is approximately unbiased.

In the social choice literature a voting rule which—as all the above examples—always elects exactly one winner, is often called a *resolute* voting rule. Since we do not allow for multiple winners of an election, we always assume voting rules to be resolute and just call them "voting rules".

2.7.2 Properties of Voting Rules

We want to have a way of comparing the voting rules from the previous subsection. We compare them by checking whether they have the following desirable properties.

- *Monotonicity.* If somebody changes their vote, it can only change the outcome in their favor. That is, we have $f(x) \leq f(y)$ whenever $x \leq y$ pointwise.

- *Oddness.* If everyone reverses their votes, the outcome is reversed; or equivalently, the voting rule does not depend on the names of the candidates. This means that we have $f(-x) = -f(x)$ for all x. Usually this notion is called *neutrality* in the social choice literature.

- *Symmetry.* The voting rule does not depend on the identity of the voters, i.e., $f(x^\pi) = f(x)$ for all x and each permutation π of the voters. This is usually called *anonimity* in the social choice literature.

- *Unanimity.* If there is unanimity on the proposal, then the outcome reflects this: $f(-1, \ldots, -1) = -1$ and $f(1, \ldots, 1) = 1$. This is usually called *the Pareto condition* in the social choice literature.

- *Unbiasedness.* In the range of the voting rule are equally many -1's as 1's. That is, $\mathrm{E}_x[f(x)] = 0$, the mean of f is zero.

The majority rule satisfies all of these properties, and it is the only rule of the list that we have seen to do so. For example the unanimity rule does not satisfy oddness and unbiasedness; the US electoral college rule does not satisfy symmetry; the tribes rule does not satisfy oddness and symmetry, and, in general, neither unbiasedness. The dictator rule satisfies all properties except symmetry.

For 2-party elections the majority rule is the unique preferred rule:

Theorem 2.15

*(*May's Theorem*) A voting rule $f : \{-1, 1\}^n \to \{-1, 1\}$ for two alternatives is symmetric, monotone, and odd if and only if n is odd and $f = \mathrm{Maj}_n$.*

Proof. The right-to-left implication is immediate. We assume $f : \{-1, 1\}^n \to \{-1, 1\}$ is symmetric, monotone, and odd. By Lemma 2.16 following this theorem, we know f can be expressed as $f(x) = \mathrm{sgn}(a_0 + x_1 + \ldots + x_n)$, where a_0 is an integer such that $a_0 + \sum_i x_i \neq 0$ for all x.

Let D_n be the set of all values that the function $x \mapsto \sum_i x_i$ can take; observe that $D_n = \{\pm n, \pm(n-2), \pm(n-4), \ldots, \pm(n - 2\lceil n/2 \rceil)\}$. Then by the above we have $a_0 \notin D_n$.

- Case 1: n is even. Let $x = (1, \ldots, 1, -1, \ldots, -1)$, where both 1 and -1 appear $n/2$ times. Then we have $\sum_i x_i = 0$. Therefore, as f is odd, we get
$$- \mathrm{sgn}(a_0) = -f(-x) = f(x) = \mathrm{sgn}(a_0).$$
If $a_0 > 0$ or $a_0 < 0$ we get a contradiction, so we must have $a_0 = 0$. However, for n even we have $D_n = \{n, n-2, \ldots, 2, 0, -2, \ldots, -(n-2), -n\}$. But $0 = a_0 \notin D_n$, so we get a contradiction. Consequently n is odd.

- Case 2: n is odd. Let $x = (1, \ldots, 1, 1, -1, \ldots, -1)$, where 1 appears $(n+1)/2$ times and -1 appears $(n-1)/2$ times. Oddness of f gives us
$$- \mathrm{sgn}(a_0 - 1) = -f(-x) = f(x) = \mathrm{sgn}(a_0 + 1).$$
This implies $a_0 \in]-1, 1[$. As a_0 is an integer we obtain $a_0 = 0$. Thus, $f = \mathrm{Maj}_n$.

■

LEMMA 2.16

> *Let $f : \{-1,1\}^n \to \{-1,1\}$. Then f is symmetric and monotone if and only if f can be expressed as a weighted majority function with $a_1 = a_2 = \ldots = a_n = 1$ and a_0 an integer.*

Proof. For each n, let D_n be the set of values that the function $x \mapsto \sum_i x_i$ can take. It suffices to show the implication from left to right, since the other one is obvious. If $f \equiv -1$ is constantly -1, we can simply pick $a_0 = -(n+1)$; similarly if $f \equiv 1$ we can pick $a_0 = n+1$. So from now on we may assume f is nonconstant.

Claim: There are $x, y \in \{-1,1\}^n$ such that x and y differ on only one coordinate, and $f(x) = -1$ and $f(y) = 1$.

To show the claim, suppose by contradiction that it is not true. Then for all x, y differing on only one coordinate we have $f(x) = f(y)$. Putting $c = f(-1,\ldots,-1)$ it is easy to see that this implies $f(x) = c$ for all x, a contradiction.

Using the claim, let $b, y \in \{-1,1\}^n$ such that they differ on one coordinate, and also $f(b) = -1$ and $f(y) = 1$. By monotonicity we have $b \leq y$. Put $a_0 \overset{\text{def}}{=} -(\sum_i b_i + 1)$. We claim that $f(x) = \operatorname{sgn}(\sum_i x_i - (\sum_i b_i + 1))$ for all $x \in \{-1,1\}^n$. Pick $t \in \{-1,1\}^n$. There are two cases.

First, if the number of 1's in $t \geq$ the number of 1's in y, then let t' be a "rearrangement" (formally, permutation) of the coordinates of t in some way such that $t' \geq y$. Then, by symmetry, $f(t') = f(t)$. Monotonicity implies $f(t') \geq f(y) = 1$, so $f(t) = 1$. But $\sum_i t_i = \sum_i t'_i \geq \sum_i y_i = \sum_i b_i + 2$, so $\operatorname{sgn}(\sum_i t_i - (\sum_i b_i + 1)) = 1$.

Second, if the number of 1's in $t <$ the number of 1's in y, a similar argument leads to $\operatorname{sgn}(\sum_i t_i - (\sum_i b_i + 1)) = -1$. ■

2.7.3 Definition of the Impartial Culture Assumption

We need a model for the way the voters are voting. How can we do this? In the absence of any other information, that is to say "in the abstract", perhaps the best we can do is the following:

DEFINITION 2.17

> (IMPARTIAL CULTURE ASSUMPTION) The preferences of the voters are uniformly random and independent.

We abbreviate the impartial culture assumption by ICA.

More generally, a *culture* [140] is a assumption about how voters vote, i.e., the distribution of their votes. Below we will see that in the literature also other cultures are considered.

Chapter 3

Influence and Noise

The goal of this chapter is to introduce the concepts of *influence* and *noise*. After that we will be able to state the *Hypercontractive Inequality*. This is a renowned result that has had some remarkable consequences in theoretical computer science. We only prove a special case of it. Surprisingly, even this special case is already sufficiently strong to imply two consequences that are most relevant for us: the *Kahn-Kalai-Linial Theorem* and *Friedgut's Theorem*. Roughly stated, in the context of social choice the former of these theorems says that there is always a small group of voters with high influence.

This definitions and results from this chapter are based mainly on [97, 126, 63, 95, 96, 28]. Unless specifically mentioned, all our results come from these sources.

3.1 Influence of a Voter

3.1.1 Definition and Some History

We turn to a fundamental question: *does one's vote make a difference?* The following notion is appropriate to answer this question.

First we need some notation. Given a voter $i \in [n]$ and $x \in \{-1,1\}^n$, we denote with $x^{\oplus i}$ the string $(x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n)$.

Definition 3.1

> The *influence* of a voter $i \in [n]$ on $f : \{-1,1\}^n \to \{-1,1\}$ is
>
> $$\mathrm{Inf}_i[f] \stackrel{\text{def}}{=} \Pr_x[f(x) \neq f(x^{\oplus i})],$$
>
> the probability under the impartial culture assumption that voter i determines the election's outcome.

Note that this quantity equals

$$\Pr_{x \setminus x_i}[f(x_1, \ldots, x_{i-1}, -1, x_{i+1}, \ldots, x_n) \neq f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n)],$$

where the probability is taken over all the $n-1$ other voters' independent and random votes, so it is indeed the probability (over everybody else's votes) that voter i makes the difference.

Given $f : \{-1,1\}^n \to \{-1,1\}$, a vector $x \in \{-1,1\}^n$ and $i \in [n]$, we say that the coordinate i is *pivotal* (for f) on x if

$$f(x_1, \ldots, x_{i-1}, -1, x_{i+1}, \ldots, x_n) \neq f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n).$$

Consequently the influence of i on f is the probability that i is pivotal when all other voters are voting independently and uniformly. The influence of i on a function in some way quantifies how "pertinent" i is for the function.

It is interesting to mention that the notion of influence was independently invented by several people. In 1967 John Banzhaf, a lawyer, invented the notion of $\mathrm{Inf}_i[f]$ while taking legal action on behalf of two towns [2]. Banzhaf wanted to show in an objective manner that the Nassau County Board's voting system was unfair. The Nassau County (New York) Board used a weighted majority voting system to make its decisions. The six towns got different weights, based on their population. Particularly, the voting rule being used was $f : \{0,1\}^6 \to \{-1,1\}$ defined by

$$f(x) = \mathrm{sgn}(-58 + 31x_1 + 31x_2 + 28x_3 + 21x_4 + 2x_5 + 2x_6),$$

where $\mathrm{sgn}(x)$ denotes the sign of the nonzero real number x. Here $x_i = 1$ means town i is in favor of the proposal, and $f(x) = 1$ means the society accepts the proposal when the towns vote according to x. Banzhaf argues that weighted voting does not allocate voting power among legislators in proportion to the population each represents, because voting power is not proportional to the number of votes a legislator may cast. He claims it would be more accurate to think of voting power as the ability of a legislator, by his vote, *to affect* the passage or defeat of a measure. Under ICA, that is precisely what we called the influence of a voter on a voting scheme. It is therefore also known as the *Banzhaf power index.* It is easy to compute that $\mathrm{Inf}_5[f] = \mathrm{Inf}_6[f] = 0$ in this case, i.e., towns five and six have zero influence. Banzhaf argues (page 339 in [2]):

> *It is hard to conceive of any theory of representative government which could justify a system under which the representatives of three of the six municipalities "represented" are allowed to attend meetings and cast votes, but are unable to have any effect on legislative decisions. Yet this is exactly what occurs now in Nassau County. (...) No changes in the voting of any or all of the three smallest representatives will have anything other than a persuasive effect on the outcome of any proposal. They may as well stay home, except for their ability to persuade the more powerful legislators.*

Just to give an example, since $x = 1^n$ and $x = 1^n \oplus e_i$ (where e_i is the vector with all zeroes and a 1 at index i) are the only pivotal inputs, we have $\mathrm{Inf}_i[\mathrm{AND}_n] = 2^{-(n-1)}$ for every i, which is very small. Also, $\mathrm{Inf}_j[\chi_i]$ equal to the maximal value of 1 if $j = i$, and the minimal value of 0 otherwise. Further, it is clear that, given any odd n, for each i, $\mathrm{Inf}_i[\mathrm{Maj}_n] = \binom{n-1}{\frac{n-1}{2}}/2^{n-1}$, as a voter is only decisive if exactly half of the other $n-1$ voters are in favor while the other half are opposed; using Stirling's approximation this can easily seen to be equal to

$$\frac{\sqrt{2/\pi}}{\sqrt{n}} + O(n^{-\frac{3}{2}}). \tag{3.1}$$

Let us restrict to unbiased functions for now. Is it then possible to have even less influence than $O(1/\sqrt{n})$? It turns out it is: if s, w are such that $s \sim (\log 2)2^w$ (implying that the scheme is approximately unbiased) with $n = sw$, then for all i,

$$\mathrm{Inf}_i[\mathrm{Tribes}_{s,w}] \sim \frac{\log(n)}{n}.$$

This quantity can be computed by keeping in mind that for a voter to be influential in this scheme, everybody in the voter's tribe has to be in favor while all other tribes are not unanimous. The influence of any fixed voter in the tribes rule is thus tremendously smaller than $O(1/\sqrt{n})$, since $\log(n)$ is exponentially smaller than $\sqrt{n}$. In 1985, Ben-Or and Linial conjectured this scheme is a worst-case scenario: there is no unbiased function such that all voters have influence smaller than $\Omega(\log(n)/n)$. In 1988 the conjecture was settled, when Kahn, Kalai, and Linial proved the following theorem [66], now known as the *Kahn-Kalai-Linial (KKL) Theorem*:

Theorem 3.2

(Kahn-Kalai-Linial Theorem) *If $f : \{-1,1\}^n \to \{-1,1\}$ is an unbiased Boolean-valued function, then there exists an $i \in [n]$ such that $\mathrm{Inf}_i[f] \geq \Omega(\log(n)/n)$.*

Besides an analytical proof, two proofs that only use combinatorial techniques are known: one by Falik and Samorodnitsky [41], and another very similar one by Rossignol [114]. In Subsection 3.3.3 we will prove the KKL Theorem.

3.1.2 General Definition of Influence

Given a function $f : \{-1,1\}^n \to \{-1,1\}$, how can one calculate $\mathrm{Inf}_i[f]$ for $i \in [n]$? We turn to this question now. First, we define the notion of *influence* in general:

Definition 3.3

Let $f : \{-1,1\}^n \to \mathbb{R}$ and $i \in [n]$. We define the *influence of i on f* to be

$$\mathrm{Inf}_i[f] \stackrel{\text{def}}{=} \mathrm{E}_x\left[\left(\frac{f(x) - f(x^{\oplus i})}{2}\right)^2\right].$$

Note that for each function $f : \{-1,1\}^n \to \{-1,1\}$ and vector x, it holds that

$$\frac{f(x) - f(x^{\oplus i})}{2}$$

is 0 if the coordinate i is not pivotal on x, and ± 1 if i is pivotal on x. Therefore,

$$\left(\frac{f(x) - f(x^{\oplus i})}{2}\right)^2$$

is the 0/1-indicator that i is pivotal on x. Thus, for f a Boolean-valued function this new definition is in correspondence with Definition 3.1.

A useful corollary is the following.

THEOREM 3.4

If $f : \{-1,1\}^n \to \mathbb{R}$, then $\mathrm{Inf}_i[f]$ is equal to the sum of the squares of all Fourier coefficients which contain i. That is, $\mathrm{Inf}_i[f] = \sum_{S \ni i} \widehat{f}(S)^2$.

Proof. Let f_i be defined as $f_i(x) \stackrel{\text{def}}{=} f(x^{\oplus i})$ for all x. Then, $\mathrm{Inf}_i[f] = \frac{1}{4}\,\mathrm{E}_x[(f(x) - f_i(x))^2]$. Let us think a bit about the expression $f(x) - f_i(x)$. Since $f_i(x)$ just means flipping the i-th input bit of x and then applying f, the Fourier expansion of f_i is the same as the one from f, except that in each term in which the i-th variable occurs we have to add a minus sign. Thus in the Fourier expansion of $f - f_i$ all terms not involving the i-th variable get cancelled, and the other ones get doubled. Therefore we have

$$f - f_i = 2 \sum_{S \ni i} \widehat{f}(S) \chi_S.$$

Accordingly, Parseval's Theorem (Theorem 2.10) allows us to conclude that $\mathrm{Inf}_i[f] = \sum_{S \ni i} \widehat{f}(S)^2$. ∎

To calculate the influence of a voter on a function, it thus suffices to compute the Fourier expansion, after which we square and add all coefficients belonging to terms that involve the i-th variable.

Going back to the second example given in Section 2.1 where we were considering the function $\min_3$, the influence of i is $(1/4)^2 + (1/4)^2 + (1/4)^2 + (1/4)^2 = 1/4$; this is indeed correct, as a voter can only make a difference in this scheme when the other two voters vote 1, which happens in exactly 1 of the 4 possible cases.

For monotone voting schemes the formula gets even nicer (this result and proof are based on Lecture 12 of [95]):

THEOREM 3.5

Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone. Then we have $\mathrm{Inf}_i[f] = \widehat{f}(i)$.

Proof. The most elementary proof is by double counting. Call two strings n-bit strings a *pair* (in the literature also called *edge of the hypercube*) if and only if they are identical except on exactly one coordinate. Note that there are precisely 2^{n-1} pairs.

By Definition 3.1, $\mathrm{Inf}_i[f]$ represents the fraction of pairs such that $f(y) = -f(y^{\oplus i})$.

On the other hand, since $\widehat{f}(i) = \mathrm{E}_x[x_i\, f(x)]$ we have that $\widehat{f}(i)$ is equal to the fraction of pairs such that $f(x^{i=-1}) = -1$ and $f(x^{i=1}) = 1$, minus the number of pairs such that $f(y^{i=-1}) = 1$ and $f(y^{i=1}) = -1$. (Here $y^{i=1}$ is the string which is like y but which has a 1 in its i-th coordinate, and $y^{i=-1}$ is defined similarly.) But this last case cannot happen because f is monotone. Therefore, both numbers count exactly the same thing, and are thus equal. ∎

Opting for a voting rule that is monotone and symmetric inherently makes all voters have equal, small, influence:

THEOREM 3.6

Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be monotone and symmetric. Then it holds that $\mathrm{Inf}_i[f] \le \frac{1}{\sqrt{n}}$, *for all* $i \in [n]$.

Proof. For each $j \in [n]$, we have $1 = \sum_{S\subseteq[n]} \widehat{f}(S)^2 \ge \sum_{i=1}^{n} \widehat{f}(i)^2 = n\,\widehat{f}(j)^2$; the last step is true because $\widehat{f}(i) = \widehat{f}(i')$ for all $i, i' \in [n]$, by symmetry. An application of Theorem 3.5 then gives the desired result. ■

3.1.3 Total Influence

The total influence of a function is simply the sum of all influences:

DEFINITION 3.7

Let $f : \{-1, 1\}^n \to \mathbb{R}$. The *total influence of* f is $\mathrm{I}[f] \overset{\text{def}}{=} \sum_{i=1}^{n} \mathrm{Inf}_i[f]$.

We give another useful interpretation of this quantity. To do this, let $f : \{-1, 1\}^n \to \{-1, 1\}$. For each $x \in \{-1, 1\}^n$, let $\mathrm{s}(f, x)$ be the number of pivotal voters on x; we call this the *sensitivity* of f on x. Note that

$$\mathrm{s}(f, x) = \sum_{i=1}^{n} 1_{[f(x)\neq f(x^{\oplus i})]},$$

where 1_A denotes the indicator function of the event A. Taking the expected value of $\mathrm{s}(f, x)$ over a uniform x, after exchanging sum and expectation, we get

$$\mathrm{E}_x\Big[\sum_{i=1}^{n} 1_{[f(x)\neq f(x^{\oplus i})]}\Big] = \sum_{i=1}^{n} \mathrm{E}_x\big[1_{[f(x)\neq f(x^{\oplus i})]}\big] = \sum_{i=1}^{n} \Pr_x\big[f(x) \neq f(x^{\oplus i})\big].$$

This is precisely the total influence of f. Therefore, $\mathrm{I}[f] = \mathrm{E}_x[\mathrm{s}(f, x)]$, i.e., *the total influence equals the average sensitivity.*

Consider the function $\chi_{[n]}$, which is just the parity on all coordinates. The total influence of this function is maximal; it equals n, as the number of pivotal voters on $x = \mathrm{s}(f, x) = n$, for all x. A trivial example attaining the minimal possible value of 0 total influence is a constant function: every voter has influence 0. We also have

$$\mathrm{I}[\mathrm{Maj}_n] = \sqrt{\frac{2}{\pi}}\sqrt{n} + O(n^{-\frac{1}{2}}).$$

This follows immediately from equation (3.1) in Subsection 3.1.1 for n odd; for n even one can show that we still have that the influence of an n-variable majority function is $\sqrt{\frac{2}{\pi}}\sqrt{n} + O(n^{-\frac{1}{2}})$.

In *Du contrat social ou Principes du droit politique* [115], a book written in 1762, the French Enlightenment philosopher Rousseau argues that, when doing an election, one should always try to maximize the number of voters that agree with the outcome. Given a fixed voting scheme f, let $A(x)$ be the number of voters that agree with the outcome x when the election scheme in use is f. Note that

$$A(x) = \sum_{i=1}^{n} \Big(\frac{1}{2} + \frac{1}{2} x_i f(x)\Big).$$

Indeed, the quantity inside the sum is 0 if $x_i \neq f(x)$, and 1 if $x_i = f(x)$, i.e., if voter i agrees with the outcome x. Therefore, we have

$$\mathrm{E}_x[A(x)] = \frac{n}{2} + \frac{1}{2}\sum_{i=1}^{n} \mathrm{E}_x[x_i f(x)] = \frac{n}{2} + \frac{1}{2}\sum_{i=1}^{n} \widehat{f}(i),$$

where we used that $\mathrm{E}_x[x_i f(x)] = \langle \chi_i, f\rangle = \widehat{f}(i)$.

Following Rousseau, we want to maximize this quantity. We can, of course, do this by just maximizing the quantity $\sum_{i=1}^{n} \widehat{f}(i)$. For monotone functions f by Theorem 3.5 this just means maximizing $\mathrm{I}[f]$.

Theorem 3.8

Among all functions $f : \{-1,1\}^n \to \{-1,1\}$, the sum $\sum_{i=1}^{n} \widehat{f}(i)$ is maximized by the majority functions. In particular, for each monotone function f we have $\mathrm{I}[f] \leq \sqrt{\frac{2}{\pi}}\sqrt{n} + O(n^{-\frac{1}{2}})$.

Proof. By the calculation preceding this theorem, we have

$$\sum_{i=1}^{n} \widehat{f}(i) = \sum_{i=1}^{n} \mathrm{E}_x[x_i f(x)] = \mathrm{E}_x[(x_1 + \ldots + x_n)\, f(x)].$$

Since $f(x)$ is either -1 or 1, this quantity is upper bounded by $\mathrm{E}_x[|x_1+\ldots+x_n|]$, with equality if and only if $f(x) = \mathrm{sgn}(x_1 + \ldots + x_n)$ for all x for which $x_1 + \ldots + x_n \neq 0$, i.e., if and only if f is a majority function. The second claim is an immediate consequence of the discussion preceding this theorem. ∎

This theorem once again shows that for a 2-candidate voting scheme the majority rule is best.

Lastly we have the following result.

Theorem 3.9

For each $f : \{-1,1\}^n \to \mathbb{R}$ we have $\mathrm{I}[f] = \sum_{S\subseteq[n]} |S|\widehat{f}(S)^2$.

Proof. By Theorem 3.4, $\mathrm{I}[f] = \sum_{i=1}^{n} \sum_{S\ni i} \widehat{f}(S)^2$. By inspection we see that the latter is precisely $\sum_{S\subseteq[n]} |S|\widehat{f}(S)^2$. ∎

3.2 Noise Stability

3.2.1 Introduction

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a voting rule used in an election. We can imagine there might be some misreading by the computer that registers the voters' preferences, or human error by the people who do the counting. Suppose $x_1, \ldots, x_n$ are the "true" votes of the voters, i.e., their sincere preferences. Then let $y_1, \ldots, y_n$ be the votes as they are recorded by the voting computer; it is possible that something went wrong (for example miscounting) and some y_i's may be different from the corresponding true vote x_i. We want to investigate the quantity $\Pr[f(x) = f(y)]$. But to do this of course we need to give y a probability distribution. For x we assume the impartial culture assumption, as usual. We do this as follows.

Definition 3.10

Let $x \in \{-1,1\}^n$ be fixed and $\rho \in [0,1]$. If y is a random variable taking values in $\{-1,1\}^n$ such that

$$y_i = \begin{cases} x_i & \text{with probability } \rho \text{ (independently for all } i) \\ \pm 1 & \text{uniformly with probability } 1-\rho \end{cases},$$

then we say y is *ρ-correlated to x* and denote this by $y \sim \mathcal{N}_\rho(x)$.

We should think of ρ as being close to 1, symbolizing that, hopefully, the voting computer's propensity to make errors is small. It follows immediately that

$$\Pr[y_i = x_i] = \rho + \frac{1}{2}(1-\rho) = \frac{1}{2} + \frac{1}{2}\rho, \quad \Pr[y_i \neq x_i] = \frac{1-\rho}{2} = \frac{1}{2} - \frac{1}{2}\rho.$$

Note that this even makes sense for $\rho \in [-1,0]$, so we might as well take these equations as the general definition.

Definition 3.11

Let $x \sim \{-1,1\}^n$ be uniformly distributed, and $\rho \in [-1,1]$. Then if for a random variable $y = (y_1, \ldots, y_n)$ taking values in $\{-1,1\}^n$ it holds that all y_i's are independent, and

$$\Pr[y_i = x_i] = \frac{1}{2} + \frac{1}{2}\rho,$$

then we say that (x,y) is a *ρ-correlated pair*, denoted by $(x,y)_\rho$.

Note that $\Pr[y_i = 1] = 1/2$ for each i. Hence each y_i, and because of that also y, is uniformly distributed as well; therefore the above definition is symmetric, in the sense that (x,y) is a ρ-correlated pair if and only if (y,x) is a ρ-correlated pair. Fix an i. It is plain that x_i and y_i are in general *not* independent: as a matter of fact, we have

$$\mathrm{E}[x_i y_i] = \mathrm{E}[x_i y_i \,|\, y_i = x_i]\Pr[y_i = x_i] + \mathrm{E}[x_i y_i \,|\, y_i \neq x_i]\Pr[y_i \neq x_i],$$

and this is equal to $(+1)\left(\frac{1}{2}+\frac{1}{2}\rho\right) + (-1)\left(\frac{1}{2}-\frac{1}{2}\rho\right) = \rho$. For $\rho \neq 0$ it follows that x_i and y_i are not independent; for $\rho = 0$ they are.

An equivalent way of stating Definition 3.11 would be to say that (x, y) is a ρ-correlated pair if and only if

- x_i and y_i are uniformly distributed, for each i;
- x_i and y_i have correlation ρ, for each i;
- All pairs (x_i, y_i) are independent.

3.2.2 Definition of Noise Stability

Using the notion of ρ-correlated pair, we can define noise stability:

Definition 3.12

For $f : \{-1, 1\}^n \to \mathbb{R}$ and $\rho \in [-1, 1]$, we let

$$\mathrm{Stab}_\rho[f] \stackrel{\text{def}}{=} \mathrm{E}_{(x,y)_\rho}[f(x)f(y)]$$

and call it the *noise stability of f at ρ*.

Note the presence of the impartial culture assumption in this definition.

For example, $\mathrm{Stab}_\rho[+1] = 1$, which is maximal; this case is however not interesting. We also have $\mathrm{Stab}_\rho[\chi_i] = \rho$ for any i: dictators have noise stability ρ. This is easy to see using the following.

If $f : \{-1, 1\}^n \to \{-1, 1\}$ is Boolean-valued, we can view f as an election rule and interpret this quantity further: we then have

$$\mathrm{Stab}_\rho[f] = \Pr[f(x) = f(y)] - \Pr[f(x) \neq f(y)] = 1 - 2\Pr[f(x) \neq f(y)].$$

So, in that case, the noise stability is a number in $[-1, 1]$; moreover, it is close to 1 if and only if $\Pr[f(x) \neq f(y)]$ is close to 0, i.e., if and only if the election rule f is unlikely to have its outcome altered due to the flipped votes (symbolized by the ρ-correlated pair (x, y)).

For an election scheme it is beneficial to have a high noise stability at ρ, given that ρ is reasonably close to 1.

What about the noise stability of the majority rule? By using the multidimensional Central Limit Theorem it is not so hard to show that, given a fixed ρ, we have

$$\mathrm{Stab}_\rho[\mathrm{Maj}_n] \to \frac{2}{\pi}\arcsin(\rho) \tag{3.2}$$

for $n \to +\infty$, and n odd. Since the derivative of $\rho \mapsto \arcsin(\rho)$ goes to plus infinity when ρ approaches 1 from below, we can say the following: supposing the number of voters n is very large, given a ρ very close to 1 (meaning the quality of the computer recording the votes is very good), if we were to further increase ρ even by a tiny bit (meaning we would improve the computer's quality slightly), the probability under the impartial culture assumption that we would get a "wrong" election outcome due to computer errors when using the majority rule, shrinks *tremendously*.

An result worth mentioning in this context is the *Majority is Stablest Theorem*. We will not give the precise formulation, but roughly speaking it says the following: among all unbiased functions with small influences, the majority function has the largest noise stability. The first proof was given by Mossel-O'Donnell-Oleszkiewicz [86].

3.2.3 Computing Noise Stability: the Noise Operator

Very similar to the beginning of Subsection 3.1.2, we ask ourselves the following question: given a function $f : \{-1,1\}^n \to \{-1,1\}$ and $\rho \in [-1,1]$, how can one compute $\mathrm{Stab}_\rho[f]$? We use the notion of *noise operator* to do so.

Definition 3.13

Let $\rho \in [-1,1]$. We let the *noise operator* T_ρ *associated with* ρ be the functional mapping the function $f : \{-1,1\}^n \to \mathbb{R}$ into the function $\mathrm{T}_\rho f : \{-1,1\}^n \to \mathbb{R}$ defined by

$$\mathrm{T}_\rho f(x) \stackrel{\text{def}}{=} \mathrm{E}_{y \sim \mathcal{N}_\rho(x)}[f(y)].$$

Note that $T_\rho f$ *itself* is a function, for each given $f : \{-1,1\}^n \to \mathbb{R}$.

Let us imagine ρ is not too far from 1. Then, at any fixed point x, the average of f over all y close to x is taken, and this is $\mathrm{T}_\rho f(x)$. In that way we can think about $\mathrm{T}_\rho f$ as a flattened version[1] of f.

The following theorem is crucial.

[1]In fact, this flattening of functions is also called *smoothing*, and this concept is of great importance in the field of computer vision [130, 25, 141]. The basic idea is that, when one for example takes a picture, inevitably a lot of things go wrong: light fluctuations, quantization effects, etc. Indeed, images are not perfect. All of those effects are called "noise": it is anything in the image that we are not interested in, but which is present nevertheless. A naive way to enhance a picture by reducing noise is to just take lots of them and then take their average. While very effective, it is highly inefficient. The next best thing is to simply modify the pixels of the image based on some fixed function of a *local neighborhood* of any given pixel. In computer vision so-called *filters* are used for smoothing. An intuitive explanation why this actually works is that the variance of noise in the average is smaller than the variance of the pixel noise (when assuming Gaussian noise with mean zero). Thus, averaging diminishes noise. Peaks get flattened. In the application to computer vision, for example the Gaussian filter (a so-called *low-pass* filter) makes high-frequency components from the image, called *artifacts*, smaller.

THEOREM 3.14

Let $f : \{-1,1\}^n \to \mathbb{R}$. *Then the Fourier expansion of* $\mathrm{T}_\rho f$ *is*

$$\mathrm{T}_\rho f(x) = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S) x^S.$$

Proof. First, note that the function T_ρ is *linear*: $\mathrm{T}_\rho(f+g) = \mathrm{T}_\rho(f) + \mathrm{T}_\rho(g)$ and $\mathrm{T}_\rho(\lambda f) = \lambda \mathrm{T}_\rho(f)$, for all $f, g : \{-1,1\}^n \to \mathbb{R}$ and $\lambda \in \mathbb{R}$. Since each function is expressible as a linear combination of the basis functions χ_S, $S \subseteq [n]$, it suffices to prove the theorem for any such χ_S. That is, it suffices to check that $\mathrm{T}_\rho x^S = \rho^{|S|} x^S$.

Take $S \subseteq [n]$. By definition of the noise operator and by independence of the y_i's, we have

$$\mathrm{T}_\rho x^S = \mathrm{E}_{y \sim \mathcal{N}_\rho(x)} \big[\prod_{i \in S} y_i \big] = \prod_{i \in S} \mathrm{E}[y_i] = \prod_{i \in S} (x_i \rho + 0(1-\rho)) = \rho^{|S|} x^S.$$

■

Finally we get a formula for the noise stability.

THEOREM 3.15

For $f : \{-1,1\}^n \to \mathbb{R}$, *we have*

$$\mathrm{Stab}_\rho[f] = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S)^2 = \sum_{k=0}^{n} \rho^k \, \mathrm{W}^k[f].$$

Proof. First, note that $\mathrm{Stab}_\rho[f]$ equals

$$\mathrm{E}_{(x,y)_\rho}[f(x)f(y)] = \mathrm{E}_x[f(x) \, \mathrm{E}_{y \sim \mathcal{N}_\rho(x)}[f(y)]] = \mathrm{E}_x[f(x) \, \mathrm{T}_\rho f(x)] = \langle f, \mathrm{T}_\rho f \rangle.$$

An application of Plancherel's formula, Theorem 2.9, together with Theorem 3.14, results in $\mathrm{Stab}_\rho[f] = \sum_{S \subseteq [n]} \widehat{f}(s) \widehat{\mathrm{T}_\rho f}(S) = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S)^2$.

The other equality follows directly from the first one and Definition 2.13. ■

This theorem shows that the high-degree coefficients are reduced in force when calculating noise stability. Indeed, for an $S \subseteq [n]$ with $|S|$ big, the number $\rho^{|S|}$ is very small. So, *individually* they are small indeed. However, note that there is a great deal of such terms: $\binom{n}{t} \sim (n/t)^t$ of them, so we cannot simply conclude that it is the low-degree Fourier coefficients which are most important for the noise stability.

3.3 The Hypercontractive Inequality

As mentioned in the introduction of this chapter, the Hypercontractive Inequality [16, 60, 6] is a deep mathematical result. It has some remarkable applications in the field of theoretical computer science. We give the formulation now.

First we needs some notation. Let X be a random variable. For each real $p \geq 1$, we define the *p-norm of X* by

$$||X||_p \stackrel{\text{def}}{=} \sqrt[p]{\mathrm{E}[|X|^p]}.$$

To start, note that this definition is in correspondence with the norm as we defined it for Boolean functions right after Definition 2.6. Recall that Hölder's inequality (see, e.g., [100]), adapted to our setting, states that

$$||fg||_1 \leq ||f||_p \, ||g||_q \tag{3.3}$$

for all Boolean functions f, g and reals $p, q > 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$. Using Hölder's inequality it is not hard to see that $|| \cdot ||_p$ is monotone nondecreasing in p, i.e., $|| \cdot ||_q \geq || \cdot ||_p$ if $q \geq p \geq 1$.

3.3.1 Statement and Proof

In general, the theorem reads as follows:

Theorem 3.16

> (Hypercontractive Inequality) *Let P be a multilinear polynomial over the real numbers of degree d in n variables. Let $x_1, \ldots, x_n \sim \{-1, 1\}$ be uniformly distributed and independent, and consider the random variable F defined by $P(x_1, \ldots, x_n)$. Then for all $q \geq p \geq 1$ we have*
>
> $$||F||_q \leq \left(\sqrt{\frac{q-1}{p-1}}\right)^d ||F||_p.$$

Because of the comment preceding the statement, the quintessence of this inequality is that it gives a bound on $||F||_q$ *from above*, and *in terms of* $||F||_p$. Thus, for a low-degree polynomial P, $||F||_q$ is not much bigger than $||F||_p$.

For our purposes the special case in which $q = 4$ and $p = 2$ suffices. This special case of the general theorem is also called the *Bonami Lemma* [97]. What is remarkable about this special case, is that it is by itself already strong enough to prove a lot of theorems from the theory of analysis of Boolean functions. For example, we will show the Kahn-Kalai-Linial (KKL) Theorem and Friedgut's Theorem below. Both are corollaries of the Bonami Lemma. In the next chapter, about Arrow's Theorem, the Bonami Lemma will be used to prove the Friedgut-Kalai-Naor (FKN) Theorem.

Theorem 3.17

(Hypercontractive Inequality in case $(q,p) = (4,2)$, Bonami's Lemma) Let P be a multilinear polynomial over the real numbers of degree d in n variables. Let $x_1, \ldots, x_n \sim \{-1, 1\}$ be uniformly distributed bits. Consider the random variable F defined by $P(x_1, \ldots, x_n)$. Then we have

$$\mathrm{E}[F^4] \leq 9^d \, \mathrm{E}[F^2]^2.$$

In other words, for any $f : \{-1,1\}^n \to \mathbb{R}$ of degree at most d, it holds that

$$||f||_4 \leq \sqrt{3}^d ||f||_2.$$

The proof of this limited version is not particularly hard. The idea is to just proceed by induction on n and use some easy bounds, e.g., the Cauchy-Schwarz inequality. The details are as follows.

Proof of Theorem 3.19. The proof goes by induction on n.

The base case $n = 0$ is trivial, as then the random variable F is just a fixed constant.

Finally we show the induction step. Let $n \geq 1$ and let P be the given polynomial of degree d in n variables $x_1, \ldots, x_n$. Since P is multilinear we can write

$$P(x) = x_n \, Q(x_1, \ldots, x_{n-1}) + R(x_1, \ldots, x_{n-1})$$

where Q and R are real polynomials in only $n-1$ variables $x_1, \ldots, x_{n-1}$. Since the degree of P is d, the degree of Q is must be at most $d-1$. The degree of R is at most d.

Let G be the random variable $Q(x_1, \ldots, x_{n-1})$, and H be $R(x_1, \ldots, x_{n-1})$, where all x_i are uniformly distributed and independent.

By simple algebra we calculate that

$$\mathrm{E}[F^4] = \mathrm{E}[(x_n G + H)^4] = \mathrm{E}[x_n^4 G^4 + 4x_n^3 G^3 H + 6x_n^2 G^2 H^2 + 4x_n G H^3 + H^4].$$

From linearity of expectation we obtain

$$\mathrm{E}[F^4] = \mathrm{E}[x_n^4 G^4] + 4\,\mathrm{E}[x_n^3 G^3 H] + 6\,\mathrm{E}[x_n^2 G^2 H^2] + 4\,\mathrm{E}[x_n G H^3] + \mathrm{E}[H^4].$$

Two terms immediately vanish: $\mathrm{E}[x_n^3 G^3 H] = 0$ and $\mathrm{E}[x_n G H^3] = 0$ because x_n is independent of $x_1, \ldots, x_{n-1}$ (and thus also of $G^3 H$ and GH^3) and $\mathrm{E}[x_n] = \mathrm{E}[x_n^3] = 0$. Before we proceed, note that $\mathrm{E}[x_n^2] = \mathrm{E}[x_n^4] = 1$. We will use this below.

We are left with three terms:

- First, $\mathrm{E}[x_n^4 G^4] = \mathrm{E}[x_n^4]\,\mathrm{E}[G^4] = \mathrm{E}[G^4]$ again because of independence. Noting that G is a random variable of polynomial form of degree at most $d-1$ and in only $n-1$ variables, the induction hypothesis implies $\mathrm{E}[x_n^4 G^4] = \mathrm{E}[G^4] \leq 9^{d-1}(\mathrm{E}[G^2]^2)$.

- Second, $\mathrm{E}[x_n^2 G^2 H^2] = \mathrm{E}[x_n^2]\,\mathrm{E}[G^2 H^2] = \mathrm{E}[G^2 H^2]$ by independence. By the Cauchy-Schwarz inequality we get
$$\mathrm{E}[x_n^2 G^2 H^2] = \mathrm{E}[G^2 H^2] \le \sqrt{\mathrm{E}[G^4]}\sqrt{\mathrm{E}[H^4]}.$$
Note that we can apply the induction hypothesis to G (which comes from a polynomial of degree of at most $d-1$) as well as H (which comes from a polynomial of degree of at most d), and we obtain that
$$\mathrm{E}[x_n^2 G^2 H^2] \le 3^{d-1}\,\mathrm{E}[G^2] 3^d\,\mathrm{E}[H^2].$$
- Third, $\mathrm{E}[H^4] \le 9^d\,\mathrm{E}[H^2]^2$ by the induction hypothesis.

Now all that remains is to put everything together. We have

$$\begin{aligned}
\mathrm{E}[F^4] &= \mathrm{E}[x_n^4 G^4] + 4\,\mathrm{E}[x_n^3 G^3 H] + 6\,\mathrm{E}[x_n^2 G^2 H^2] + 4\,\mathrm{E}[x_n G H^3] + \mathrm{E}[H^4] \\
&\le 9^{d-1}(\mathrm{E}[G^2]^2) + 6\,3^{d-1}\,\mathrm{E}[G^2]3^d\,\mathrm{E}[H^2] + 9^d\,\mathrm{E}[H^2]^2 \\
&= 9^{d-1}(\mathrm{E}[G^2]^2) + 2\,9^d\,\mathrm{E}[G^2]\,\mathrm{E}[H^2] + 9^d\,\mathrm{E}[H^2]^2 \\
&\le 9^d(\mathrm{E}[G^2]^2) + 2\,\mathrm{E}[G^2]\,\mathrm{E}[H^2] + \mathrm{E}[H^2]^2) \\
&= 9^d(\mathrm{E}[G^2] + \mathrm{E}[H^2])^2 \\
&= 9^d(\mathrm{E}[G^2] + 2\,\mathrm{E}[x_n G H]\,\mathrm{E}[x_n^2 H^2])^2 \\
&= 9^d\,\mathrm{E}[(G + x_n H)^2]^2 \\
&= 9^d\,\mathrm{E}[F^2]^2
\end{aligned}$$

Notice that the third last step is true because x_n is independent of both GH (implying that $\mathrm{E}[x_n GH] = \mathrm{E}[x_n]\,\mathrm{E}[GH] = 0$) as well as H^2 (so $\mathrm{E}[x_n^2 H^2] = \mathrm{E}[x_n^2]\,\mathrm{E}[H^2] = \mathrm{E}[H^2]$). ■

For any $f : \{-1,1\}^n \to \mathbb{R}$ and natural number m, with $f^{=m}$ we mean the Boolean function $\sum_{S:|S|=m} \widehat{f}(S)\chi_S$, i.e., $f^{=m}$ is obtained by taking only the degree-m terms of the Fourier expansion of f. Since it is of degree precisely m, Bonami's Lemma together with Theorem 3.15 implies

$$\left|\left|\mathrm{T}_{\sqrt{1/3}}(f^{=m})\right|\right|_4 = \left|\left|\left(\sqrt{1/3}\right)^m f^{=m}\right|\right|_4 = \left(\sqrt{1/3}\right)^m ||f^{=m}||_4 \le ||f^{=m}||_2\,. \tag{3.4}$$

This expression hints at the following theorem, which involves the noise operator and can be shown to be equivalent with Theorem 3.16:

Theorem 3.18

*(*Hypercontractivity Theorem*) Let p and q be reals satisfying $1 \le p \le q$, and let ρ be a real such that*

$$\rho \le \sqrt{\frac{p-1}{q-1}}.$$

Then for all $f : \{-1,1\}^n \to \mathbb{R}$ we have

$$||\,\mathrm{T}_\rho\, f||_q \le ||f||_p.$$

One of the nice things about this formulation of hypercontractivity is that it involves the noise operator. Recall from the discussion in Subsection 3.2.3 that adding noise makes functions more flat. More precisely, $\mathrm{T}_\rho f$ will be a function that looks somewhat like f but which is more "spread out", i.e., its peaks are reduced.

More precisely the question is: *how exactly* does noise affect the p-norm of a given Boolean function? The Hypercontractivity Theorem provides an answer (see [29] for a more in-depth discussion). Indeed, the interpretation goes as follows. First, it is not hard to see that $||\,\mathrm{T}_\rho(f)||_p \leq ||f||_p$ for any p and f. Second, even if q is bigger than p, provided a sufficient amount of noise is applied to the function (i.e., $\rho \leq \sqrt{(p-1)/(q-1)}$), then even in that case we still get $||\,\mathrm{T}_\rho\, f||_q \leq ||f||_p$.

3.3.2 Special Case of the Hypercontractivity Theorem

Previously we claimed that the statements from Theorem 3.16 and Theorem 3.18 are in fact equivalent, in the sense that if one of them holds for all p and q, then also the other one holds for all p, q, and vice versa. We will not show this in general, but we will just prove the following special cases which we will explicitly need further on in order to prove the KKL Theorem and Friedgut's Theorem:

Theorem 3.19

The Hypercontractivity Theorem (that is, Theorem 3.18) holds for $(q, p, \rho) = \left(2, \frac{4}{3}, \sqrt{\frac{1}{3}}\right)$. In other words, for each Boolean function $f : \{-1,1\}^n \to \mathbb{R}$, we have

$$||\,\mathrm{T}_{\sqrt{\frac{1}{3}}}\, f||_2 \quad \leq \quad ||f||_{\frac{4}{3}}.$$

In essence, the proof of Theorem 3.19 comes down to an application of Bonami's Lemma (Theorem 3.19). Roughly speaking, the idea is as follows:

- Summing Equation (3.4) over all m suggests that

$$\left|\left|\mathrm{T}_{\sqrt{1/3}}\, f\right|\right|_4 \leq ||f||_2 \tag{3.5}$$

  is true of all $f : \{-1,1\}^n \to \mathbb{R}$.

  Even though this plan of action fundamentally works, it does not work *directly*, in the sense that we would need to use some extra ploys. As it turns out, it is easier just to "repeat" the induction proof from Bonami's Lemma. We do this in Lemma 3.20.

- Then we turn (3.5) into $||\,\mathrm{T}_{\sqrt{1/3}}\, f||_2 \leq ||f||_{\frac{4}{3}}$ for any $f : \{-1,1\}^n \to \mathbb{R}$, using some "tricks" from analysis. This will result in the proof of Theorem 3.19.

We now elaborate on these "tricks" from analysis. In short, they are self-adjointness and Hölder's inequality.

Given an operator T on the space of Boolean functions (meaning that it maps Boolean functions to Boolean functions), we call T *self-adjoint* if

$$\langle Tf, g\rangle = \langle f, Tg\rangle$$

holds for all functions f, g. For example, the associated noise operator T_ρ is self-adjoint, for any $\rho \in [-1, 1]$. This can be seen as follows. From Theorem 3.14, we have

$$\mathrm{T}_\rho\, f = \sum_{S\subseteq[n]} \rho^{|S|}\widehat{f}(S)\chi_S \quad , \quad \mathrm{T}_\rho\, g = \sum_{S\subseteq[n]} \rho^{|S|}\widehat{g}(S)\chi_S.$$

Then, expanding $\langle T_\rho f, g\rangle$ and $\langle f, T_\rho g\rangle$ by using bilinearity of the inner product, and then invoking the identity from Lemma 2.7, we see that $\langle Tf, g\rangle = \langle f, Tg\rangle$.

The other ingredient is, again, Hölder's inequality; see Equation (3.3).

Proof of Theorem 3.19. For brevity we write T for $\mathrm{T}_{\sqrt{1/3}}$. First, we have

$$||\,\mathrm{T}\, f||_2^2 = \langle \mathrm{T}\, f, \mathrm{T}\, f\rangle = \langle f, \mathrm{T}\,\mathrm{T}\, f\rangle = \mathrm{E}_x[f(x)\ \mathrm{T}\,\mathrm{T}\, f(x)] \leq \mathrm{E}_x[|f(x)\ \mathrm{T}\,\mathrm{T}\, f(x)|],$$

where the second equality followed by self-adjointness. The right-hand side is equal to $||f \cdot \mathrm{T}\,\mathrm{T}\, f||_1$. Second, by Holder's inequality, this quantity is at most $||f||_{4/3}\, ||\,\mathrm{T}\,\mathrm{T}\, f||_4$. Third, by Lemma 3.20 (which will be independently proved below), $||\,\mathrm{T}\,\mathrm{T}\, f||_4 \leq ||\,\mathrm{T}\, f||_2$.

Putting everything together, we have $||\,\mathrm{T}\, f||_2^2 \leq ||f||_{4/3}\, ||\,\mathrm{T}\, f||_2$. Of course, we may assume $||\,\mathrm{T}\, f||_2$ to be non-zero. Dividing by $||\,\mathrm{T}\, f||_2$ ends the proof. ∎

The proof of Theorem 3.19 is based on the proof of the $(4/3, 2)$-Hypercontractivity Theorem in [97].

The following lemma, which should be thought of as an alternative form of Bonami's Lemma, is the only thing we still need to show. The proof is given only for completeness and self-containment of the exposition; the proof is really the same as the proof of Bonami's Lemma, so it may very well be skipped. Its proof can also be found in [97].

LEMMA 3.20

> *Let $f : \{-1, 1\}^n \to \mathbb{R}$. Then,*
>
> $$\left|\left|\mathrm{T}_{\sqrt{1/3}}\, f\right|\right|_4 \leq ||f||_2\,.$$

Proof. Equivalently, we prove $\mathrm{E}[(\mathrm{T}_{\sqrt{1/3}}\, f)^4] \leq \mathrm{E}[f^2]^2$, essentially using the same induction on n as in the proof of Bonami's Lemma (Theorem 3.19).

We write T for $\mathrm{T}_{\sqrt{1/3}}$. The base case is trivial. Let $n \geq 1$ and let f be the given Boolean function in n variables $x_1, \ldots, x_n$. We can write

$$f(x) = x_n\, Q(x_1, \ldots, x_{n-1}) + R(x_1, \ldots, x_{n-1})$$

where Q and R are polynomials in only $n-1$ variables $x_1, \ldots, x_{n-1}$. We abbreviate $A \stackrel{\text{def}}{=} Q(x_1, \ldots, x_{n-1})$ and $B \stackrel{\text{def}}{=} R(x_1, \ldots, x_{n-1})$, so $f = x_n A + B$. Using Theorem 3.15 it is easy to see that

$$\mathrm{T}\, f(x) = x_n \frac{1}{\sqrt{3}} \mathrm{T}\, A + \mathrm{T}\, B.$$

Taking the fourth power of this equation gives five terms. Two of them immediately vanish under expectation, for they contain the variable x_n to an odd power (which in expectation is 0, as x_n is independent from A and B since the latter two only involve $x_1, \ldots, x_{n-1}$). Because $\mathrm{E}[x_n^2] = \mathrm{E}[x_n^4] = 1$, we have

$$\mathrm{E}[(\mathrm{T}\, f)^4] = \left(\frac{1}{\sqrt{3}}\right)^4 \mathrm{E}[(\mathrm{T}\, A)^4] + 6\left(\frac{1}{\sqrt{3}}\right)^2 \mathrm{E}[(\mathrm{T}\, A)^2(\mathrm{T}\, B)^2] + \mathrm{E}[(\mathrm{T}\, B)^4].$$

By Cauchy-Schwarz, $\mathrm{E}[(\mathrm{T}\, A)^2(\mathrm{T}\, B)^2] \le \sqrt{\mathrm{E}[(\mathrm{T}\, A)^4]}\sqrt{\mathrm{E}[(\mathrm{T}\, B)^4]}$, so we get

$$\begin{aligned} \mathrm{E}[(\mathrm{T}\, f)^4] &\le \mathrm{E}[(\mathrm{T}\, A)^4] + 2\,\mathrm{E}[(\mathrm{T}\, A)^2(\mathrm{T}\, B)^2] + \mathrm{E}[(\mathrm{T}\, B)^4] \\ &\le \mathrm{E}[(\mathrm{T}\, A)^4] + 2\sqrt{\mathrm{E}[(\mathrm{T}\, A)^4]}\sqrt{\mathrm{E}[(\mathrm{T}\, B)^4]} + \mathrm{E}[(\mathrm{T}\, B)^4] \\ &\le \mathrm{E}[A^2]^2 + 2\mathrm{E}[A^2]\mathrm{E}[B^2] + \mathrm{E}[B^2], \end{aligned}$$

where the last inequality follows by the induction hypothesis. The right-hand side is just $(\mathrm{E}[A^2] + \mathrm{E}[B^2])^2$. It is clear that

$$\mathrm{E}[f^2] = \mathrm{E}[x_n^2 A^2 + 2x_n AB + B^2] = \mathrm{E}[A^2] + \mathrm{E}[B^2],$$

so we obtain $\mathrm{E}[(\mathrm{T}\, f)^4] \le \mathrm{E}[f^2]^2$. ■

3.3.3 Proof of the Kahn-Kalai-Linial Theorem

The following theorem is a corollary of the Hypercontractive Inequality. We will just need a special case of this result, so we omit the proof of the general result.

Theorem 3.21

If $f : \{-1,1\}^n \to \{-1,0,1\}$, and $p = 1+\delta \in [1,2]$, then

$$\mathrm{Stab}_\rho[f] = \sum_S \delta^{|S|} \widehat{f}(S)^2 \le \Pr_x[f(x) \neq 0]^{2/p}. \tag{3.6}$$

The proof of this theorem is an immediate application of Theorem 3.18 with $\rho = \sqrt{\delta}$ and $q = 2$, together with Parseval's Theorem (Theorem 2.10). However, since we did not prove Theorem 3.18 in its full generality, and since we want this thesis to be as self-contained as possible, we will again be content just to prove a special case.

The intuition behind Theorem 3.21 is as follows: *Boolean functions with small support have much Fourier mass on higher degrees.* (Recall that the *support* of a function is the set of all elements in its domain which are mapped to a nonzero value.) The reason why this is true is as follows. The right-hand side of (3.6) is, by Parseval's Theorem, the sum of squares of all Fourier-coefficients, raised to some power greater than one; the left-hand side, $\sum_S \delta^{|S|}\widehat{f}(S)^2$, can be thought of as roughly equal to the sum of squares of all *low-degree* Fourier-coefficients (the higher-degree ones being attenuated very fast, and therefore extremely small). Hence, as the exponent at the right-hand side is greater than one, the sum of low-degree Fourier-coefficients must be significantly smaller than the sum of *all* Fourier-coefficients; therefore, the function must have much Fourier mass on higher degrees. Once more, we refer to [29] for a more detailed explanation and further intuitions.

The special case which will be sufficient for our purpose is when $\delta = 1/3$:

THEOREM 3.22

If $f : \{-1,1\}^n \to \{-1,0,1\}$, *then*

$$\mathrm{Stab}_{\frac{1}{3}}[f] = \sum_{S\subseteq[n]} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2 \le \Pr_x[f(x) \neq 0]^{\frac{3}{2}}.$$

Proof. From Theorem 3.14 and Parseval's Theorem we know that

$$\sum_{S\subseteq[n]} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2 = ||\,\mathrm{T}_{\sqrt{\frac{1}{3}}}\, f||_2^2.$$

The latter is, by Theorem 3.19, at most $||f||_{\frac{4}{3}}^2$. However, since $|f(x)| = |f(x)|^t$ for any t (because f can only attain the values $-1, 0$, and 1), we have

$$||f||_{\frac{4}{3}}^{\frac{4}{3}} = \mathrm{E}_x[|f(x)|^{\frac{4}{3}}] = \mathrm{E}_x[|f(x)|] = \Pr_x[f(x) \neq 0],$$

and therefore

$$\sum_{S\subseteq[n]} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2 \le ||f||_{\frac{4}{3}}^2 = \left(||f||_{\frac{4}{3}}^{\frac{4}{3}}\right)^{\frac{3}{2}} = \Pr_x[f(x) \neq 0]^{\frac{3}{2}}.$$

The first equality is just Theorem 3.15. ■

We use this theorem to prove the KKL Theorem. It is convenient to use the following notation. If $f : \{-1,1\}^n \to \mathbb{R}$, then

$$\mathrm{MaxInf}[f] \stackrel{\mathrm{def}}{=} \max\{\mathrm{Inf}_i[f] \mid i \in [n]\}.$$

That is, $\mathrm{MaxInf}[f]$ is just the maximum of the influences of all voters.

THEOREM 3.23

*(*KAHN-KALAI-LINIAL THEOREM*) Let* $f : \{-1,1\}^n \to \{-1,1\}$ *be an unbiased Boolean-valued function. Then,* $\mathrm{MaxInf}[f] \ge \Omega(\log(n)/n)$.

Proof. Theorem 3.4 implies $\mathrm{Inf}_i[f] = \sum_{S \ni i} \widehat{f}(S)^2$. Let $d = (\log n)/4$. There are two cases:

Case 1: $\sum_{S\,:\,|S|>d} \widehat{f}(S)^2 \geq 1/2$. In that case

$$\sum_{i=1}^{n} \mathrm{Inf}_i[f] = \sum_{i=1}^{n} \sum_{S \ni i} \widehat{f}(S)^2 = \sum_{S} |S|\, \widehat{f}(S)^2.$$

Clearly this is at least

$$\sum_{S\,:\,|S|>d} |S|\, \widehat{f}(S)^2 > d \sum_{S\,:\,|S|>d} \widehat{f}(S)^2 \geq d/2.$$

The average of the numbers $\{\mathrm{Inf}_i[f] \mid i \in [n]\}$ is hence at least $d/(2n)$, so there is an $i \in [n]$ for which $\mathrm{Inf}_i[f] \geq d/(2n) = \log(n)/(8n)$.

Case 2: $\sum_{S\,:\,|S|>d} \widehat{f}(S)^2 < 1/2$. By Parseval's Theorem the sum of all Fourier weights is 1, so

$$\sum_{S\,:\,|S|\leq d} \widehat{f}(S)^2 \geq 1/2.$$

For each $i \in [n]$, we define a function $f_i : \{-1,1\}^n \to \{-1,0,1\}$ by

$$f_i(x) \stackrel{\text{def}}{=} (f(x) - f(x^{\oplus i}))/2.$$

Note that $\mathrm{Inf}_i[f] = \Pr_x[f_i \neq 0]$. Then, applying Theorem 3.22 on the function f_i, we get

$$\mathrm{Inf}_i[f]^{3/2} = \Pr_x[f_i \neq 0]^{3/2} \geq \sum_{S} \left(\frac{1}{3}\right)^{|S|} \widehat{f_i}(S)^2 = \sum_{S \ni i} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2,$$

where the last step follows because from the proof of Theorem 3.4 we know that

$$f_i(x) = \frac{f(x) - f(x^{\oplus i})}{2} = \sum_{S \ni i} \widehat{f}(S) \chi_S(x).$$

By summing the above inequality over all i, we obtain

$$\begin{aligned} \sum_{i=1}^{n} \mathrm{Inf}_i[f]^{3/2} &\geq \sum_{i=1}^{n} \sum_{S \ni i} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2 \\ &= \sum_{S} |S| 3^{-|S|} \widehat{f}(S)^2 \\ &\geq \sum_{S\,:\,|S|\leq d} |S| 3^{-|S|} \widehat{f}(S)^2. \end{aligned}$$

In the last inequality we used that the quantity $(k3^{-k})_k$ is decreasing in k. From the above inequality we deduce that

$$\sum_{i=1}^{n} \mathrm{Inf}_i[f]^{3/2} \geq \sum_{S\,:\,|S|\leq d} d3^{-d} \widehat{f}(S)^2 = d3^{-d} \sum_{S\,:\,|S|\leq d} \widehat{f}(S)^2 \geq \frac{1}{2} d3^{-d}.$$

A simple calculation shows that $\frac{1}{2}d3^{-d}$ is equal to $\frac{1}{8}\log(n)3^{\left(\frac{-\log n}{4}\right)}$. Therefore, the average of the numbers $\{\mathrm{Inf}_i[f]^{3/2} \mid i \in [n]\}$ must be at least $\frac{1}{8}\frac{\log(n)}{n}3^{\left(\frac{-\log n}{4}\right)}$, so there is an $i \in [n]$ for which

$$\mathrm{Inf}_i[f]^{3/2} \geq \frac{1}{8}\frac{\log(n)}{n}3^{\left(\frac{-\log n}{4}\right)}.$$

Then we have

$$\mathrm{Inf}_i[f] \geq \left(\frac{1}{8}\right)^{\frac{2}{3}} \left(\frac{\log(n)}{n}\right)^{\frac{2}{3}} \left(3^{\log n}\right)^{-\frac{1}{6}}.$$

We want to show that the right-hand side is $\Omega(\log(n)/n)$. Clearly, $\log(n)^{\frac{2}{3}} \geq 1$, so it suffices to show that

$$\left(3^{\log n}\right)^{-\frac{1}{6}} > \frac{\log n}{n}n^{\frac{2}{3}} = n^{-\frac{1}{3}}\log n$$

for n big enough, or equivalently,

$$n > \sqrt{3^{\log n}}(\log n)^3.$$

But clearly this is true for n big enough. (This can be seen as follows: let $n = 2^t$, then it suffices to see that $2^t > \sqrt{3}^t t^3$, i.e., $(2/\sqrt{3})^t > t^3$. But $2 > \sqrt{3}$, and, of course, any exponential function in t with base > 1 will be bigger than any polynomial in t, for t big enough.)

In both case we have reached the same conclusion. Therefore, there is an $i \in [n]$ such that $\mathrm{Inf}_i[f] \geq \Omega(\log(n)/n)$. ■

Our proof is based on [29].

In fact a slightly more general result can be proved: using similar techniques one can show that for each Boolean-valued function $f : \{-1,1\}^n \to \{-1,1\}$ there exists an $i \in [n]$ such that $\mathrm{Inf}_i[f] \geq \mathrm{Var}[f]\,\Omega(\log(n)/n)$. Note that if f is unbiased, then $\mathrm{Var}[f] = 1$. We will use this somewhat more general version of the KKL Theorem below.

3.3.4 Corollary of the KKL Theorem: Bribing Works

To state the next corollary of the KKL Theorem, we need to define a new notion. Let $f : \{-1,1\}^n \to \mathbb{R}$. Let $J \subseteq [n]$, also called a *coalition.* Let $\overline{j} \in \{-1,1\}^{|J|}$ be a tuple of length $|J|$. Then we define the *restriction of f in which the voters in J vote according to $\overline{j}$* as the function $f_{J\mapsto\overline{j}}(x) \stackrel{\text{def}}{=} f(x,\overline{j})$, for each $x \in \{-1,1\}^{n-|J|}$. Here with $(x,\overline{j})$ we mean the n-tuple in which all voters in J vote according to $\overline{j}$, while all voters in $[n] \setminus J$ vote according to x. Note that $f_{J\mapsto\overline{j}}$ only depends on the voters in $[n] \setminus J$. This definition might seem a bit abstract, but it is actually something very uncomplicated. A simple example should make clear what is meant. Recall the example from Section 2.1 in which we deduced that the Fourier expansion of $f = \mathrm{Maj}_3$ is given by

$$f(x) = \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 - \frac{1}{2}x_1x_2x_3.$$

Let J be the coalition $\{1,3\}$. The voters in the coalition, namely voters 1 and 3, already decided for whom to vote: say, $\overline{j} = (1,-1)$, i.e., voter 1 will vote 1 while voter 3 will vote -1. Then $f_{J\mapsto\overline{j}}$ is simply the above function in which x_1 is consistently replaced by 1 while x_3 is replaced by -1, so that

$$f_{J\mapsto\overline{j}}(x) = \frac{1}{2}(+1) + \frac{1}{2}x_2 + \frac{1}{2}(-1) - \frac{1}{2}(+1)\,x_2\,(-1) = x_2.$$

This makes sense: if, given three voters that are using the majority election scheme, two voters disagree on their preferred candidate, in that case the third candidate's choice will be the outcome.

The following theorem, based on [97], shows that in an unbiased, monotone, election scheme for two candidates, *each* of the two candidates can "bribe" a coalition of reasonably small size, thus manipulating the outcome of the election in their favor, almost surely:

Theorem 3.24

Let $f:\{-1,1\}^n \to \{-1,1\}$ *be monotone. If* $\mathrm{E}[f] \geq -0.9999999$*, then there exists a subset* $J \subseteq [n]$ *with* $|J| \leq O(n/\log n)$ *such that*

$$\mathrm{E}[f_{J\mapsto(1,\ldots,1)}] \geq 0.9999999. \tag{3.7}$$

Similarly, if $\mathrm{E}[f] \leq 0.9999999$*, then there exists a subset* $J' \subseteq [n]$ *with* $|J'| \leq O(n/\log n)$ *such that*

$$\mathrm{E}[f_{J'\mapsto(-1,\ldots,-1)}] \geq 0.9999999.$$

Proof. We write δ for 0.9999999. Because of symmetry it suffices to show the first claim, in which candidate 1 wants to find a coalition he can bribe in order to win the election almost surely.

The idea of the proof is straightforward. It comes down to the following "bribing algorithm":

- First, bribe voter i_1 with the largest influence on $f_0 \stackrel{\text{def}}{=} f$.
- Second, bribe voter i_2 with the largest influence on $f_1 \stackrel{\text{def}}{=} f_{\{i_1\}\mapsto 1}$.
- Third, bribe voter i_3 with the largest influence on $f_2 \stackrel{\text{def}}{=} f_{\{i_1,i_2\}\mapsto(1,1)}$.
- Fourth, bribe voter i_4 with the largest influence on $f_3 \stackrel{\text{def}}{=} f_{\{i_1,i_2,i_3\}\mapsto(1,1,1)}$, etc.

We have to show that at some point in this process (3.7) will be satisfied.

We claim that for all $k \geq 0$ it holds that

$$\mathrm{E}[f_{k+1}] = \mathrm{E}[f_k] + \mathrm{MaxInf}[f_k]. \tag{3.8}$$

We just show this for $k = 0$, as the other cases follows in exactly the same way. For $k = 0$ we have to show that

$$\mathrm{E}[f_{\{i_1\}\mapsto 1}] = \mathrm{E}[f] + \mathrm{Inf}_{i_1}[f].$$

Let us think about this in terms of the Fourier expansion

$$f(x) = a_0 + a_1 x_1 + \ldots + a_{i_1} x_{i_1} + \ldots + a_n x_n + \text{(higher-order terms)}.$$

The Fourier expansion of $f_{\{i_1\}\mapsto 1}$ is consequently obtained by plugging in $x_{i_1} = 1$ into the above expression, so it is equal to

$$f_{\{i_1\}\mapsto 1}(x) = a_0 + a_{i_1} + \sum_{j\in[n]\setminus\{i_1\}} a_j x_j + \text{(higher-order terms)}.$$

Recall from Theorem 2.8 that the constant coefficient in the Fourier expansion is just the expectation of the function. Thus, $\mathrm{E}[f] = a_0$ and $\mathrm{E}[f_{\{i_1\}\mapsto 1}] = a_0 + a_{i_1}$. However, we have $\mathrm{Inf}_{i_1}[f] = \widehat{f}(i_1) = a_{i_1}$, using monotonicity of f and Theorem 3.5. This proves the claim.

Therefore,

$$\mathrm{E}[f_{k+1}] = \mathrm{E}[f_0] + \sum_{i\leq k} \mathrm{MaxInf}[f_i]. \tag{3.9}$$

Let k be any natural number. If after k bribes the candidate still has not yet achieved (3.7), then $\mathrm{E}[f_k] < \delta$ must be the case. Then, as by hypothesis $\mathrm{E}[f_k] \geq \mathrm{E}[f] \geq -\delta$,

$$\mathrm{Var}[f_k] = \mathrm{E}[f_k^2] - \mathrm{E}[f_k]^2 > 1 - \delta^2,$$

so the extended KKL Theorem (which was mentioned after the proof of Theorem 3.23) implies that

$$\mathrm{MaxInf}[f_k] \geq (1-\delta^2)\,\Omega\left(\frac{\log n}{n}\right).$$

Hence from (3.9) we obtain

$$\mathrm{E}[f_{k+1}] = \mathrm{E}[f_0] + \sum_{i\leq k} \mathrm{MaxInf}[f_i] \geq -\delta + k\,(1-\delta^2)\,\Omega\left(\frac{\log n}{n}\right). \tag{3.10}$$

However, for any k satisfying

$$k \geq \frac{2\delta}{(1-\delta^2)\,\Omega\left(\frac{\log n}{n}\right)}$$

the quantity on the right-hand side of (3.10) is at least δ, so then $\mathrm{E}[f_{k+1}] \geq \delta$. Note that

$$\frac{2\delta}{(1-\delta^2)\,\Omega\left(\frac{\log n}{n}\right)} = O\left(\frac{n}{\log n}\right).$$

This shows that the bribing algorithm will always work. Furthermore, from the above expression it follows that at most $O\left(\frac{n}{\log n}\right)$ voters need to be bribed. ■

In fact, for the above statement (both claims) to be true, *any* δ will do, as long as δ is not -1 or 1: then, $\mathrm{Var}[f] = 0$, and in that case the KKL Theorem does not help, so the above proof strategy falls apart. In those cases, f is constant 1 or constant -1; obviously then any bribing attempt will be of no avail for one of the candidates.

3.3.5 Friedgut's Theorem

Given a Boolean-valued function having very small total influence, Friedgut's Theorem, first proved by Friedgut in 1998 [49], entails that there must be a small coalition, a *junta*, of voters which determine the function with high probability. More specifically, suppose the election function has an influence which is significantly lower than $\log(n)$, where n is the number of voters. In that case, Friedgut's Theorem implies that f depends on a sublinear amount of variables.

Before stating the theorem, we need the following notions, based on [97].

Definition 3.25

Let $\mathcal{D}$ be a collection of subsets of $[n]$ and let $\varepsilon > 0$. Then we say that the Fourier spectrum of $f : \{-1,1\}^n \to \mathbb{R}$ (or just f itself) is *ε-concentrated on $\mathcal{D}$* if

$$\sum_{S\subseteq[n],\, S\notin\mathcal{D}} \widehat{f}(S)^2 \leq \varepsilon,$$

i.e., if the sum of all Fourier weights of subsets not in $\mathcal{D}$ is at most ε.

A function $f : \{-1,1\}^n \to \mathbb{R}$ being ε-concentrated on some collection $\mathcal{D}$ intuitively means that $\mathcal{D}$ are the "most important" Fourier coefficients of f present: the remaining ones are less important, as their total Fourier weight is at most ε. This will become apparent further on.

The following lemma relates ε-concentration with ε-closeness:

Lemma 3.26

Let $\mathcal{D}$ be a collection of subsets of $[n]$ and $\varepsilon > 0$.

(1) If $f : \{-1,1\}^n \to \mathbb{R}$ is ε-concentrated on $\mathcal{D}$, then

$$||f - g||_2^2 \leq \varepsilon$$

where $g \overset{def}{=} \sum_{S\in\mathcal{D}} \widehat{f}(S)\chi_S$.

(2) Let $f : \{-1,1\}^n \to \{-1,1\}$ and $g : \{-1,1\}^n \to \mathbb{R}$ be such that $||f-g||_2^2 \leq \varepsilon$. We define a function $h : \{-1,1\}^n \to \{-1,1\}$ by $h(x) \overset{def}{=} \operatorname{sgn}(g(x))$ (here $\operatorname{sgn}(0)$ is left unspecified: it is either -1 or 1, irrelevantly). In that case, f and h are ε-close.

(3) If $f : \{-1,1\}^n \to \{-1,1\}$ is ε-concentrated on $\mathcal{D}$, then f and $\operatorname{sgn}(\sum_{S\in\mathcal{D}} \widehat{f}(S)\chi_S)$ are ε-close.

Proof. (1) Writing f in its Fourier expansion, we have

$$f - g = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_X - \sum_{S\subseteq[n],\, S\in\mathcal{D}} \widehat{f}(S)\chi_S = \sum_{S\subseteq[n],\, S\notin\mathcal{D}} \widehat{f}(S)\chi_S.$$

Parseval's Theorem implies

$$||f-g||_2^2 = \sum_{S\subseteq[n],\, S\notin\mathcal{D}} \widehat{f}(S)^2 \leq \varepsilon.$$

(2) The essential observation is that for each $x \in \{-1,1\}^n$, when $f(x) \neq \operatorname{sgn}(g(x))$ then, as f is Boolean-valued, $|f(x)-g(x)| \geq 1$, so also $|f(x)-g(x)|^2 \geq 1$. Therefore,

$$\operatorname{dist}(f,h) = \Pr_x[f(x) \neq h(x)] = \mathrm{E}_x[1_{[f(x)\neq \operatorname{sgn}(g(x))]}] \leq \mathrm{E}_x[|f(x)-g(x)|^2] = ||f-g||_2^2,$$

so $\operatorname{dist}(f,h) \leq ||f-g||_2^2 \leq \varepsilon$.
(3) This item follows immediately from (1) and (2). ■

The following theorem, based on [96], is sufficiently strong to imply Friedgut's Theorem:

Theorem 3.27

Let $f : \{-1,1\}^n \to \{-1,1\}$ and $0 < \varepsilon \leq 1$. Then there is a set $J \subseteq [n]$ of size at most $\exp(O(\mathrm{I}[f]/\varepsilon))$ such that f's Fourier spectrum is ε-concentrated on

$$\{S \subseteq J \mid |S| \leq 2\,\mathrm{I}[f]/\varepsilon\}.$$

Proof. We write α for $2\,\mathrm{I}[f]/\varepsilon$. We define

$$J \stackrel{\text{def}}{=} \{i \in [n] \mid \mathrm{Inf}_i[f] \geq \exp(-3\alpha)\},$$

where exp is the exponential function with base e. Then J is a set of "sufficiently influential" voters. Clearly we have

$$\mathrm{I}[f] \geq \sum_{i\in J} \mathrm{Inf}_i[f] \geq |J| \exp(-3\alpha),$$

so $|J|$ is at most

$$\exp(3\alpha)\,\mathrm{I}[f] = \exp\left(6\frac{\mathrm{I}[f]}{\varepsilon}\right)\frac{\mathrm{I}[f]}{\varepsilon}\varepsilon \leq \exp\left(7\frac{\mathrm{I}[f]}{\varepsilon}\right) = \exp\left(O\left(\frac{\mathrm{I}[f]}{\varepsilon}\right)\right).$$

Therefore, it suffices to show that f's Fourier spectrum is ε-concentrated on

$$\mathcal{G} \stackrel{\text{def}}{=} \{S \subseteq J \mid |S| \leq \alpha\},$$

i.e., that

$$\sum_{S\subseteq[n],\, S\notin\mathcal{G}} \widehat{f}(S)^2 \leq \varepsilon.$$

Note that for any $S \subseteq [n]$,

$$S \notin \mathcal{G} \quad \Leftrightarrow \quad (|S| > \alpha \text{ or } (|S| \leq \alpha \text{ and } S \not\subseteq J)),$$

so we have to inspect two terms:

1. We show that
$$\sum_{S\subseteq[n],\,|S|>\alpha} \widehat{f}(S)^2 \le \frac{\varepsilon}{2}.$$
This is easy, for by Theorem 3.9 we have
$$\mathrm{I}[f] = \sum_{S\subseteq[n]} |S|\,\widehat{f}(S)^2 \ge \sum_{S\subseteq[n],\,|S|>\alpha} |S|\,\widehat{f}(S)^2 \ge \alpha \sum_{S\subseteq[n],\,|S|>\alpha} \widehat{f}(S)^2,$$
and $\mathrm{I}[f]/\alpha = \varepsilon/2$.

2. We show that
$$\sum_{S\not\subseteq J,\,|S|\le\alpha} \widehat{f}(S)^2 \le \frac{\varepsilon}{2}.$$
Similarly as we did in the proof of case 2 of the KKL Theorem (Theorem 3.23), applying Theorem 3.22 on the function $f_i : \{-1,1\}^n \to \{-1,0,1\}$ defined by $f_i(x) \stackrel{\text{def}}{=} (f(x)-f(x^{\oplus i}))/2$, by noting that $\mathrm{Inf}_i[f] = \Pr_x[f_i \neq 0]$, we get
$$\mathrm{Inf}_i[f]^{3/2} = \Pr_x[f_i \neq 0]^{3/2} \ge \sum_S \left(\frac{1}{3}\right)^{|S|} \widehat{f_i}(S)^2 = \sum_{S\ni i} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2$$
for each $i \in [n]$. Summing this inequality over all $i \notin J$, we obtain
$$\sum_{i\notin J} \mathrm{Inf}_i[f]^{3/2} \ge \sum_{i\notin J}\sum_{S\ni i} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2.$$
This inequality is the essential ingredient of the proof. We investigate the left-hand side and the right-hand side separately.

   - First, since for each $i \notin J$ it holds that $\mathrm{Inf}_i[f] < \exp(-3\alpha)$ (i.e., its influence is small), we have
   $$\sum_{i\notin J} \mathrm{Inf}_i[f]^{3/2} = \sum_{i\notin J} \mathrm{Inf}_i[f]^{1/2}\,\mathrm{Inf}_i[f] \le \exp(-3\alpha)^{\frac{1}{2}} \sum_{i\notin J} \mathrm{Inf}_i[f],$$
   and this is upperbounded by $\exp(-3\alpha)^{\frac{1}{2}}\,\mathrm{I}[f] = \exp\left(-3\frac{\mathrm{I}[f]}{\varepsilon}\right)\mathrm{I}[f]$.
   - Second, writing J^c for $[n]\setminus J$, we have
   $$\begin{aligned}
   \sum_{i\notin J}\sum_{S\ni i} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2 &= \sum_{S\,:\,S\cap J^c\neq\emptyset} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2\,|S\cap J^c| \\
   &\ge \sum_{S\not\subseteq J,\,|S|\le\alpha} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2 \\
   &\ge \sum_{S\not\subseteq J,\,|S|\le\alpha} \left(\frac{1}{3}\right)^{\alpha} \widehat{f}(S)^2 \\
   &= 3^{-\alpha} \sum_{S\not\subseteq I,\,|S|\le\alpha} \widehat{f}(S)^2.
   \end{aligned}$$

Putting the two together, we have

$$\exp\left(-3\frac{\mathrm{I}[f]}{\varepsilon}\right)\mathrm{I}[f] \geq 3^{-\alpha} \sum_{S \not\subseteq I,\, |S| \leq \alpha} \widehat{f}(S)^2,$$

so $\sum_{S \not\subseteq I,\, |S| \leq \alpha} \widehat{f}(S)^2$ is at most

$$3^{\alpha}\exp\left(-3\frac{\mathrm{I}[f]}{\varepsilon}\right)\mathrm{I}[f] = 3^{\alpha}\exp\left(-\frac{3}{2}\alpha\right)\alpha\frac{\varepsilon}{2} = ((3e^{-\frac{3}{2}})^{\alpha}\,\alpha)\frac{\varepsilon}{2}.$$

Hence it suffices to show that $(3e^{-\frac{3}{2}})^{\alpha}\,\alpha \leq 1$. We can calculate that $3e^{-\frac{3}{2}} \approx 0.67$. A plot of the graph of the real function $x \mapsto 0.67^x\,x$ shows that for all $x \geq 0$ it is strictly smaller than 1 (in fact, its maximal value is about 0.9). Therefore, $(3e^{-\frac{3}{2}})^{\alpha}\,\alpha \leq 1$.

■

Friedgut's Theorem is an immediate corollary:

Theorem 3.28

*(*Friedgut's Theorem*) Let* $f : \{-1,1\}^n \to \{-1,1\}$ *and* $0 < \varepsilon \leq 1$. *Then* f *is* ε*-close to an* $\exp(O(\mathrm{I}[f]/\varepsilon))$*-junta.*

Proof. From Theorem 3.27 we get a set $J \subseteq [n]$ of size at most $\exp(O(\mathrm{I}[f]/\varepsilon))$ with the property that the Fourier spectrum of f is ε-concentrated on the set $\mathcal{D} \stackrel{\text{def}}{=} \{S \subseteq J \mid |S| \leq 2\,\mathrm{I}[f]/\varepsilon\}$.

Then item (3) of Lemma 3.26 implies that f and $\operatorname{sgn}(\sum_{S \in \mathcal{D}} \widehat{f}(S)\chi_S)$ are ε-close. All of the latter function's variables are in J, and the size of J is at most $\exp(O(\mathrm{I}[f]/\varepsilon))$. Therefore, f is ε-close to an $\exp(O(\mathrm{I}[f]/\varepsilon))$-junta. ■

Note the similarity between the proofs of the KKL Theorem and Friedgut's Theorem; indeed, the proof techniques are very much the same. In essence, they come down to the same thing: Bonami's Lemma (Theorem 3.19), or the form that we used, Theorem 3.19. By now we hope to have convinced the reader of the power of Bonami's Lemma. This particular case of the Hypercontractive Inequality is strong enough to entail lots of interesting results.

Chapter 4

Arrow's Theorem

The goal of this chapter is to prove Arrow's Theorem using Fourier analysis on the Boolean cube. One advantage of this approach is that we can prove a stronger, robust, version of the theorem, a *quantitative* variant of Arrow's Theorem. This result says that the more we want to avoid Condorcet's paradox, the more the election scheme will look like a dictator, under ICA.

This definitions and results from this chapter are based mainly on [97, 126, 63, 95, 96, 28]. Unless specifically mentioned, all our results come from these sources.

4.1 Arrow's Theorem

4.1.1 Introduction and Formalization

We have so far seen that for a 2-candidate election the majority rule is the unique best choice, as e.g. May's Theorem (Theorem 2.15) and Theorem 3.8 show. But what if there are more than two candidates? This is the question we turn to now.

For now, let us focus on the case when there are three candidates, say a, b, c. In social choice theory one then assumes each voter has a *preference* of the candidates. E.g., a voter might like c most, and b least. We can write this as $c > a > b$. A *preference*, also called *ballot*, is formally an irreflexive linear order on the set of candidates. A *profile* is a vector $R = (R_1, \dots, R_n)$ of ballots, where R_i is voter i's preference. Let $\mathcal{L}$ be the set of all linear orders on $\{a, b, c\}$. A *social welfare function (SWF)* is a function $\mathcal{L}^n \to \mathcal{L}$: for any preference the voters might have, it returns a preference order which we call the *societal outcome*.

Given all those individual's preferences $R_1, \dots, R_n$ of the voters, how are we going to aggregate them into a societal outcome? Condorcet, an 18th century French philosopher, suggested to do the following [27]. First, we should break up the election into three sub-elections:

$$a \text{ versus } b, \quad b \text{ versus } c, \quad c \text{ versus } a.$$

This is a good idea, Condorcet argues, since we know how to deal with 2-candidate elections. Particularly we might just use the majority rule for each of these sub-elections, and afterwards aggregate the winners of these sub-elections into a societal outcome; the candidate topping this list is then called the *Condorcet winner*.

However, in general a problem may arise. For example, consider the case in which there are three voters, and they vote as follows:

$$a > b > c, \quad c > a > b, \quad b > c > a.$$

For each of the three sub-elections we use the majority rule, say. For the a versus b election we get two votes in favor of $a > b$ while we have one vote in favor of $b > a$, so the majority rule implies $a > b$ should be the societal outcome. Similarly we get $b > c$ and $c > a$ as for the society; thus, the societal outcome is $a > b > c > a$, which is not even a linear order! This is the *Condorcet paradox*: it might happen that on the societal level we do not get a linear order—equivalently stated, there might be no Condorcet winner. In that case we also say that the election is *undecided*, or *irrational*. Another way of saying this is to say that there is a *cycle*.

We first need to be able to translate a linear order on $\{a, b, c\}$ into our language of Boolean analysis, i.e., in terms of 1 and -1. This is easy: for each voter, let us call a now $+1$ while we call b instead -1, and similarly for b versus c and c versus a. If e.g. voter i prefers b to a, then we write this as $x_i = -1$. In this way we get the Boolean vector $x = (x_1, \ldots, x_n)$ of length n. We similarly get the vectors $y, z \in \{-1, 1\}^n$, which correspond to the b versus c and the c versus a election, respectively. So we might have the following scheme:

	voters						
	1	2	3	$\ldots$			
$a\ (+1)$ vs. $b\ (-1)$	$+1$	$+1$	-1	$\ldots$	$\stackrel{\text{def}}{=} x$	$\rightsquigarrow$	$f(x)$
$b\ (+1)$ vs. $c\ (-1)$	$+1$	-1	$+1$	$\ldots$	$\stackrel{\text{def}}{=} y$	$\rightsquigarrow$	$g(y)$
$c\ (+1)$ vs. $a\ (-1)$	-1	-1	$+1$	$\ldots$	$\stackrel{\text{def}}{=} x$	$\rightsquigarrow$	$h(z)$

which corresponds to the following preferences of the first three voters:

$$a > b > c, \quad a > c > b, \quad b > c > a.$$

Let us say each of the three sub-elections uses the majority rule, i.e., $f = g = h = \mathrm{Maj}_n$. In this case there is a Condorcet winner, and the societal outcome is the linear order $a > b > c$.

Definition 4.1

> Let $f, g, h : \{-1, 1\}^n \to \{-1, 1\}$. A 3-*candidate Condorcet election based on the election rules* (f, g, h) is the election scheme as described above. In case the obtained result from this procedure is a linear order we say the outcome is *rational*, or that *there exists a Condorcet winner*, or also that *there is no cycle*; otherwise we say the outcome is *irrational*, *there is no Condorcet winner*, or *there is a cycle*.

Let V_i be the vector corresponding to voter i's preference. For example, for the example given in the above table we have

$$V_1 = \begin{pmatrix} +1 \\ +1 \\ -1 \end{pmatrix}, \quad V_2 = \begin{pmatrix} +1 \\ -1 \\ -1 \end{pmatrix}, \quad V_3 = \begin{pmatrix} -1 \\ +1 \\ +1 \end{pmatrix}.$$

Now when is a preference irrational? Note that the vectors which correspond to irrational preferences are precisely

$$\begin{pmatrix} +1 \\ +1 \\ +1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}.$$

So the valid preferences are those vectors satisfying the "not-all-equal" predicate. This will be one of the crucial parts of the Fourier proof of Arrow's Theorem.

4.1.2 Proof of Arrow's Theorem Using Boolean Analysis

In the above we have seen that when using the majority rule for each of the three sub-elections, we might get an irrational outcome in some cases. We now ask ourselves the question: *are there choices for f, g and h such that we never get an irrational outcome?* If so, this would seem like a very appealing situation. Unfortunately, under the very benign assumption that f, g and h have the property of unanimity, Arrow's Theorem says we will always get an irrational outcome in some cases, *unless* there is a dictator (which using the below terminology means $f = g = h = \chi_i$ for some i). Of course, having a dictator is rather objectionable. Thus, Condorcet's idea fundamentally does not work—at least if we want to avoid irrational outcomes at all costs. We come back to this point later.

Kenneth Arrow (1921) won the Nobel Memorial Prize in Economics in 1972 together with John Hicks, for "pioneering contributions to general equilibrium theory and welfare theory". To date, he is the youngest person to have received this award, at age 51. Arrow's most famous result is probably the impossibility result that is named after him, and that is studied in this chapter. He showed this result in his 1951 Ph.D. dissertation called *Social Choice and Individual Values* [1].

The idea of the proof of Arrow's Theorem using Boolean analysis simply consists of computing the probability of having an irrational outcome, say under the impartial culture assumption.

THEOREM 4.2

(ARROW'S THEOREM FOR THREE CANDIDATES) *Let $f, g, h : \{-1,1\}^n \to \{-1,1\}$ be unanimous and such that, when doing a 3-candidate Condorcet election based on (f, g, h), the outcome is always rational. Then $f = g = h$, and they are equal to a dictatorship.*

Proof. We just show that $f = g = h$; the result then instantly follows from Theorem 4.3 below. Take any $x = (x_1, \ldots, x_n) \in \{-1,1\}^n$. We subsequently make the voters vote as follows: $V_i = (x_i, -x_i, 1)$ for each i. As a result, the societal outcome is the linear order corresponding to the tuple $(f(x), g(-x), h(\overrightarrow{1}))$. By unanimity of h, this is $(f(x), g(-x), 1)$. But the outcome of the Condorcet election is always rational, so not all coordinates of $(f(x), g(-x), 1)$ are equal. Thus, $f(x) = g(-x) = -1$, or $f(x) \neq g(-x)$.

By a similar reasoning (taking $V_i = (x_i, -x_i, -1)$ for each i) we get $(f(x), g(-x), -1)$ as the outcome of the Condorcet election; by rationality we again have $f(x) = g(-x) = 1$, or $f(x) \neq g(-x)$.

Using these results and some elementary logic, we conclude $f(x) \neq g(-x)$. Hence we obtain that $f(x) = -g(-x)$ for all x.

By symmetry of the situation we likewise have $f(x) = -h(-x)$ for all x, therefore $-g(-x) = f(x) = -h(-x)$ for all x. This implies $g = h$. Again by symmetry we obtain also $f = g$. Therefore, $f = g = h$. ■

THEOREM 4.3

Let $f : \{-1,1\}^n \to \{-1,1\}$ be unanimous. Suppose that f is such that, when doing a 3-candidate Condorcet election based on f, the outcome is always rational. Then f is a dictatorship.

Proof. The whole idea of the proof is to consider the quantity

$$\Pr_{\text{ICA}}[\text{rational outcome}].$$

Here ICA refers to the fact that we are computing the probability of a rational outcome under the impartial culture assumption. By assumption this probability equals 1. However we want to compute it alternatively using Fourier analysis on the Boolean cube.

By ICA the voters are voting independently and uniformly, so in the scheme

$$\begin{array}{ccccccccccccccccc}
 & & & 1 & & 2 & & 3 & & 4 & & 5 & & \dots & & & \\
\hline
x & = & (& x_1 & , & x_2 & , & x_3 & , & x_4 & , & x_5 & , & \dots &) & \rightsquigarrow & f(x) \\
y & = & (& y_1 & , & y_2 & , & y_3 & , & y_4 & , & y_5 & , & \dots &) & \rightsquigarrow & f(y) \\
z & = & (& z_1 & , & z_2 & , & z_3 & , & z_4 & , & z_5 & , & \dots &) & \rightsquigarrow & f(z)
\end{array}$$

the columns V_i ($i \in [n]$) are independent and uniformly distributed over the set of six possible preference rankings, i.e., NAE-vectors—namely, they are

$$\left\{ \begin{pmatrix} +1 \\ +1 \\ -1 \end{pmatrix}, \begin{pmatrix} +1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} +1 \\ -1 \\ +1 \end{pmatrix}, \begin{pmatrix} -1 \\ +1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ +1 \\ +1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ +1 \end{pmatrix} \right\}.$$

Now Fourier analysis comes in. Let $\text{NAE} : \{-1,1\}^n \to \{0,1\}$ be the indicator of the property "not all equal". It is not hard to verify that its Fourier expansion is

$$\text{NAE}(t_1, t_2, t_3) = \frac{3}{4} - \frac{1}{4}t_1t_2 - \frac{1}{4}t_2t_3 - \frac{1}{4}t_1t_3.$$

Therefore, letting x, y, z be distributed according to the above distribution, we have

$$\begin{aligned}
\Pr_{\text{ICA}}[\text{rational outcome}] &= \Pr_{x,y,z}[\mathrm{NAE}(f(x), f(y), f(z)) = 1] \\
&= \mathrm{E}_{x,y,z}[\mathrm{NAE}(f(x), f(y), f(z))] \\
&= \mathrm{E}_{x,y,z}\left[\frac{3}{4} - \frac{1}{4}f(x)f(y) - \frac{1}{4}f(y)f(z) - \frac{1}{4}f(x)f(z)\right] \\
&= \frac{3}{4} - \frac{1}{4}\,\mathrm{E}_{x,y}[f(x)f(y)] - \frac{1}{4}\,\mathrm{E}_{y,z}[f(y)f(z)] \\
&\quad -\frac{1}{4}\,\mathrm{E}_{x,z}[f(x)f(z)].
\end{aligned}$$

Now we claim that (x, y) is a $\left(-\frac{1}{3}\right)$-correlated pair. Notice that because of the comment following Definition 3.11, it suffices to check that all pairs (x_i, y_i) are independent, each of them is $\left(-\frac{1}{3}\right)$-correlated, and each x_i and each y_i is uniformly distributed. This is indeed the case:

(1) The pairs (x_i, y_i) are independent because under ICA it is assumed all the voters vote independently.

(2) We have $\Pr[x_i = 1] = 1/2$, as the voters are assumed to be voting randomly by ICA, so $x_i \sim \{-1, 1\}$; for the same reason we have $y_i \sim \{-1, 1\}$.

(4) Finally, $\mathrm{E}[x_i y_i] = (+1)\frac{1}{4} + (-1)\frac{2}{3} = -\frac{1}{3}$, since $x_i y_i$ can be either $+1$ (which happens when V_i is $(+1, +1, -1)^T$ or $(-1, -1, +1)^T$, so with probability equal to $2/6 = 1/3$), or -1, which happens with probability $2/3$.

By symmetry, the same is true for the other two pairs: (y, z) and (x, z) are $\left(-\frac{1}{3}\right)$-correlated pairs as well. Following Definition 3.12 we can ultimately conclude that

$$\Pr_{\text{ICA}}[\text{rational outcome}] = \frac{3}{4} - \frac{3}{4}\,\mathrm{Stab}_{-\frac{1}{3}}[f]. \tag{4.1}$$

Therefore we have

$$\begin{aligned}
\Pr_{\text{ICA}}[\text{rational outcome}] &= \frac{3}{4} - \frac{3}{4}\,\mathrm{Stab}_{-\frac{1}{3}}[f] \\
&= \frac{3}{4} - \frac{3}{4}\Big(\mathrm{W}^0[f] + \left(-\frac{1}{3}\right)\mathrm{W}^1[f] + \left(\frac{1}{9}\right)\mathrm{W}^2[f] + \\
&\quad \left(-\frac{1}{27}\right)\mathrm{W}^3[f] + \dots\Big) \\
&\leq \frac{3}{4} - \frac{3}{4}\left(-\frac{1}{3}\right)(\mathrm{W}^0[f] + \mathrm{W}^1[f] + \dots + \mathrm{W}^n[f]) \\
&= 1.
\end{aligned}$$

The inequality holds since by inspection $-1/3$ is the smallest coefficient present. In the second equality we used Theorem 3.15 and in the last step Equation (2.2).

Equality holds if and only if the inequality is an equality, that is, $\mathrm{W}^i[f] = 0$ for all $i \neq 1$. But this means that all Fourier weight is on degree 1, i.e., $\mathrm{W}^1[f] = 1$. By Lemma 2.14 this means that f is $\pm\chi_i$ for some i. Since f is unanimuous, f cannot be an anti-dictator. Thus f is a dictator, finishing the proof of the theorem. ■

The above proof strategy is based on Kalai's influential paper *A Fourier-theoretic Perspective on the Condorcet Paradox and Arrow's Theorem* from 2002 [68]. It was this paper that initiated the use of Fourier analysis for social choice. Our proof is an adapted version of Kalai's proof by O'Donnell [97].

4.1.3 Equivalence with Arrow's Original Statement

It might at first not be clear that Theorem 4.2 really is the same as the statement which has come to be known as *Arrow's Theorem* and which Kenneth Arrow proved while being a Ph.D. student at Columbia University. We briefly explain why they are in fact the same.

In what follows we go further using the notations introduced at the beginning of Subsection 4.1.1. Let $\mathcal{N} \stackrel{\text{def}}{=} [n]$ be the set of voters and $\mathcal{X}$ be the set of alternatives. For given alternatives $x, y \in \mathcal{X}$, we denote with $\mathcal{N}^R_{x>y}$ the set of all voters which rank alternative x above alternative y, under profile R. Let F be social welfare function. Recall that F returns a societal linear order for *all* given preferences of the n voters, i.e., the function F is total. In social choice the following are some of the properties being considered. We say F satisfies

- *the Pareto condition* if, whenever all voters rank x above y, then so does society, i.e., for all $x, y \in \mathcal{X}$ and each profile R, we have that $\mathcal{N}^R_{x>y} = \mathcal{N}$ implies $(x, y) \in F(R)$.
- *independence of irrelevant alternatives (IIA)* if the relative social ranking of two alternatives only depends on their relative individual rankings, i.e., for all $x, y \in \mathcal{X}$ and all profiles R, R', if $\mathcal{N}^R_{x>y} = \mathcal{N}^{R'}_{x>y}$ then
$$(x, y) \in F(R) \text{ if and only if } (x, y) \in F(R').$$
- *dictatoriality*, if there is an $i \in \mathcal{N}$ such that $F(R) = R_i$ for all profiles R.

Arrow's Theorem for three alternatives formulated in the classical way then reads as follows [38, 39, 132]):

Theorem 4.4

*(*Arrow's Theorem*) Any social welfare function for three alternatives that satisfies the Pareto condition and independence of irrelevant alternatives must be a dictatorship.*

First, the assumption that we are dealing with a social welfare function, which is a *total* function, corresponds in the terminology of Theorem 4.2 to the hypothesis that the outcome of the Condorcet election always be rational.

Second, the independence of irrelevant alternatives in the terminology of Theorem 4.2 precisely means that we are doing a Condorcet election using some (possible different) election functions f, g, h. Third and last, it is clear that the Pareto condition correlates with f, g and h being unanimous.

Finally, note that Arrow's Theorem for three candidates implies the general form of Arrow's Theorem, in which there is any finite number of candidates at least three. Indeed, this follows by an easy induction proof, the induction step of which is the following lemma (based on Lemma 1 in [131]):

Lemma 4.5

> *Let F be a social welfare function for $m+1$ candidates such that F satisfies the Pareto condition, IIA, and non-dictatoriality, where $m \geq 3$. Then there is a social welfare function for m candidates that satisfies these three properties.*

We omit the proof here; it can be found in [131].

Perhaps a historical anecdote is in place. Interestingly, Arrow did not actually know about Condorcet's work, let alone about the paradox named after him [82]. He rediscovered the result all by himself. At the time, Arrow was interested in understanding how firms make decisions about their production plans, keeping in mind the mesh of stakeholders. Even if we assume that all contributors share the common goal of maximizing profit of the firm, they might (and in reality indeed do) have different beliefs as to how to reach this goal. In this context, Arrow started looking at the majority rule, and soon stumbled upon the paradox. He subsequently referred to it as the "well-known paradox of voting". It was only after publication that he was notified by other researchers that the paradox already had a name. Still hopeful, Arrow was convinced there ought to be another voting mechanism that would avoid the Condorcet paradox, and that would be "reasonable" as well. After having unsuccessfully tried out a number of voting rules, he started to wonder whether there actually *is* a voting rule satisfying the aforementioned properties; in that way he started focusing on finding an *impossibility* result rather than a *possibility* result. Eric Maskin [82] calls it a good example of inductive science, as Arrow first collected evidence *against* possibility before he tried to show impossibility.

4.2 Quantifying Arrow's Theorem

4.2.1 Benefits of the Analytical Approach

We could ask ourselves what is the advantage of our Fourier analysis-based approach for proving Arrow's Theorem compared with classical, mostly combinatorial proofs of that theorem. After all, even though the proof of the essential part of the proof of Arrow's Theorem, Theorem 4.3, is fairly uncomplicated, it took us quite some preliminary work to get there. Moreover we proved a theorem for which plenty of proofs are known already.

Nevertheless, there are two main benefits of Kalai's analytic proof [68]. We briefly describe them now; in the next subsections we go into detail about each specifically.

1. We have achieved *more* than we had aspired: Equation (4.1) provides us with a concrete formula for the probability of having a rational outcome when doing a 3-candidate election. We can use this formula for example to calculate the probability of having a Condorcet winner when using the majority rule.
2. In constrast with earlier proofs [13, 38, 39, 132] of Arrow's Theorem which were outspokenly combinatorial and hence *qualitative* by nature, the given analytic proof is *quantitative* by nature. This makes us able to prove a *robust version* of Arrow's Theorem.

4.2.2 Probability of a Rational Outcome

Equation (4.1) says that $\Pr_{\mathrm{ICA}}[\text{rational outcome}] = \frac{3}{4} - \frac{3}{4}\operatorname{Stab}_{-\frac{1}{3}}[f]$. Recall from (3.2) that the noise stability of Maj_n at ρ converges to $\frac{2}{\pi}\arcsin(\rho) = 1 - \frac{2}{\pi}\arccos(\rho)$ as n goes to infinity. We have that the probability of having a rational outcome in a 3-candidate Condorcet election using Maj_n is equal to $\frac{3}{4}\frac{2}{\pi}\arccos(\rho) = \frac{3}{2\pi}\arccos(\rho)$. Therefore we can say that, when n is very big, we have

$$\Pr_{\mathrm{ICA}}[\text{3-candidate Condorcet election using } \mathrm{Maj}_n \text{ is rational}] \approx 91.2\%. \quad (4.2)$$

That means that if we assume the public votes independently and uniformly, the probability that Condorcet's paradox arises is circa 8.8%. This occurrence probability is not immense, but non-negligible nonetheless.

Still, for all we know there exists an adequate election scheme for which Condorcet's paradox crops up with negligible probability. Arrow's original theorem does not shed any light on this issue at all. Using a robust variant of Arrow's Theorem, we will answer this question in the next subsection. The following lemma, which is an easy corollary from the proof of Theorem 4.3, puts us on track.

LEMMA 4.6

> *Let $f : \{-1,1\}^n \to \{-1,1\}$. Then we have*
>
> $$\Pr_{ICA}[\textit{3-candidate Condorcet election using } f \textit{ is rational}] \leq \frac{7}{9} + \frac{2}{9}\mathrm{W}^1[f].$$

Proof. The idea is just to go back to the proof of Theorem 4.3. We got the upper bound on the probability of the outcome being rational by noticing that the coefficient $-1/3$ was the smallest one present; note however that the *next smallest* coefficient, $-1/27$, is significantly far from it. Therefore we obtain an upper bound on $\mathrm{W}^1[f]$ in the following way: we know that $\Pr_{\mathrm{ICA}}[\text{rational outcome}] = \frac{3}{4} - \frac{3}{4}\operatorname{Stab}_{-\frac{1}{3}}[f]$, and the latter is equal to

$$\begin{aligned} & \frac{3}{4} - \frac{3}{4}\left(\mathrm{W}^0[f] + \left(-\frac{1}{3}\right)\mathrm{W}^1[f] + \left(\frac{1}{9}\right)\mathrm{W}^2[f] + \left(-\frac{1}{27}\right)\mathrm{W}^3[f] + \dots\right) \\ = \; & \frac{3}{4} + \frac{1}{4}\mathrm{W}^1[f] - \frac{3}{4}\left(\mathrm{W}^0[f] + \left(\frac{1}{9}\right)\mathrm{W}^2[f] + \left(-\frac{1}{27}\right)\mathrm{W}^3[f] + \dots\right) \\ \leq \; & \frac{3}{4} + \frac{1}{4}\mathrm{W}^1[f] - \frac{3}{4}\left(-\frac{1}{27}\right)\left(\mathrm{W}^0[f] + \mathrm{W}^2[f] + \mathrm{W}^3[f] + \dots\right). \end{aligned}$$

This last equation equals $\frac{3}{4} + \frac{1}{4}\mathrm{W}^1[f] - \frac{3}{4}\left(-\frac{1}{27}\right)\left(1 - \mathrm{W}^1[f]\right) = \frac{7}{9} + \frac{2}{9}\mathrm{W}^1[f]$. Note that in the above calculation the inequality was true since $-1/27$ is the lowest (meaning most negative) coefficient present. ■

We want to examine if there exists a reasonable election scheme other than majority for which the probability of having a rational outcome in a 3-candidate Condorcet election (as always under the impartial culture assumption) is *higher* than the percentage of 91.2% by the majority rule in Equation (4.2). Of course, this depends on what is meant with "reasonable election scheme". One condition on $f : \{-1,1\}^n \to \{-1,1\}$ is demanding all its degree-one Fourier coefficients be equal; for example if f is symmetric this condition is satisfied. This condition is particularly interesting for us as we have an obvious upper bound on the degree-one weight of such a function:

Lemma 4.7

> *Let $f : \{-1,1\}^n \to \{-1,1\}$ be such that all of its degree one Fourier coefficients are equal. Then we have $\mathrm{W}^1[f] \leq \frac{2}{\pi} + O(n^{-1})$.*

Proof. By Theorem 3.8 and since all degree-one coefficients are equal, for each $j \in [n]$ we have

$$n\,\widehat{f}(j) = \sum_{i=1}^{n}\widehat{f}(i) \leq \sqrt{\frac{2}{\pi}}\sqrt{n} + O(n^{-\frac{1}{2}}),$$

so $\widehat{f}(j) \leq \sqrt{\frac{2}{\pi n}} + O(n^{-\frac{3}{2}})$. After squaring we obtain for each $j \in [n]$ that

$$\widehat{f}(j)^2 \leq \frac{2}{\pi n} + O(n^{-3}) + O(n^{-\frac{1}{2}}n^{-\frac{3}{2}}) = \frac{2}{\pi n} + O(n^{-2}).$$

Since all degree-one coefficients are equal, by summing this inequality we are done. ■

From the previous two lemmas we can conclude the following.

Theorem 4.8

Let $f : \{-1,1\}^n \to \{-1,1\}$ be such that all of its degree-one Fourier coefficients are equal. Then we have

$$\Pr_{ICA}[\text{3-candidate Condorcet election using } f \text{ is rational}] \leq \frac{7}{9}+\frac{4}{9\pi}+O(n^{-1}).$$

Proof. This follows at once from Lemma 4.6 and Lemma 4.7. ∎

For $f : \{-1,1\}^n \to \{-1,1\}$ such that all of its degree-one Fourier coefficients are equal, taking the limit for n going to infinity this means that the probability of obtaining a rational outcome in a 3-candidate Condorcet election using f is at most $\frac{7}{9} + \frac{4}{9\pi} \approx 91.9\%$. This is just slightly higher than the 91.2% we got by using the majority rule. Therefore it seems that, in order to avoid the Condorcet paradox, we cannot do significantly better than using the majority rule.

4.2.3 Robust Version of Arrow's Theorem

In the proof of Theorem 4.3 we argued as follows:

$$\Pr_{\text{ICA}}[\text{rational outcome}] = 1 \quad \text{if and only if} \quad \mathrm{W}^1[f] = 1.$$

However, going back to the calculation of the upper bound immediately after Equation (4.1), by examining the numeric values of the coefficients we see that in fact we have something more, namely:

$$\Pr_{\text{ICA}}[\text{rational outcome}] \approx 1 \quad \text{if and only if} \quad \mathrm{W}^1[f] \approx 1.$$

This can be made precise by using Lemma 4.6, as we will do below.

What does it mean for $f : \{-1,1\}^n \to \{-1,1\}$ to satisfy $\mathrm{W}^1[f] = 1 - \varepsilon$ with $\varepsilon > 0$ small? Does it then necessarily have to be "close" (in the sense of Definition 2.6) to a dictator function? Or is it the case that there are functions which are unlike a dictatorship but that still have degree-one Fourier weight very close to one? This is not an elementary question, but the answer is known: the Friedgut-Kalai-Naor (FKN) Theorem.

Theorem 4.9

*(*FKN Theorem*) Let $f : \{-1,1\}^n \to \{-1,1\}$ with $W^1[f] \geq 1 - \varepsilon$. Then there exists an $i \in [n]$ such that f is $O(\varepsilon)$-close to χ_i or $-\chi_i$.*

To make the statement more precise: given an f and some ε, "f is $O(\varepsilon)$-close to χ_i or $-\chi_i$" means that there is a universal constant $M > 0$ such that $\mathrm{dist}(f, \pm\chi_i) \leq M\varepsilon$.

Before we show the FKN Theorem using Theorem 3.19, it is worth noting that for any Boolean-valued function $h : \{-1,1\}^n \to \{-1,1\}$ we have

$$\text{for some } i \in [n],\ h \text{ is } O(\varepsilon)\text{-close to either } \chi_i \text{ or } -\chi_i$$

if and only if

$$|\widehat{h}(i)| \geq 1 - O(\varepsilon).$$

This is easy to see using Theorem 2.8 and Equation (2.1). We will use this observation in what follows.

Proof of Theorem 4.9. Without loss of generality we may assume $\mathrm{W}^1[f] = 1-\varepsilon$.

To make the below proof easier, we first show that without loss of generality we may assume that f does not have a constant coefficient in its Fourier expansion, i.e., $\mathrm{W}^0[f] = 0$. So suppose $f : \{-1,1\}^n \to \{-1,1\}$ satisfies $\mathrm{W}^1[f] = 1 - \varepsilon$. Consider the function $f' : \{-1,1\}^{n+1} \to \{-1,1\}$ defined by

$$f'(x_0, x_1, \ldots, x_n) \stackrel{\text{def}}{=} x_0\, f(x_0x_1, \ldots, x_0x_n).$$

If the Fourier expansion of f is

$$\widehat{f}(\emptyset) + \widehat{f}(\{1\})x_1 + \ldots + \widehat{f}(\{n\})x_n + \widehat{f}(\{1,2\})x_1x_2 + \ldots,$$

then by looking at the definition of f' we see that the Fourier expansion of f' is of the form

$$\widehat{f}(\emptyset)x_0 + \widehat{f}(\{1\})x_1 + \ldots + \widehat{f}(\{n\})x_n + \widehat{f}(\{1,2\})x_0x_1x_2 + \ldots.$$

Note that here we have used the fact that $x_0^2 = 1$ for any $x_0 \in \{-1,1\}$. We did not write the higher-order terms since they are not relevant for us. It is important to notice that f' does not have a degree-zero coefficient so, assuming the theorem has been proven for such functions, from

$$\mathrm{W}^1[f'] = \widehat{f}(\emptyset)^2 + \mathrm{W}^1[f] \geq \mathrm{W}^1[f] = 1 - \varepsilon$$

we deduce that f' is $O(\varepsilon)$-close to χ_i or $-\chi_i$, for some $i \in [n]$. By the comment preceding this proof, $|\widehat{f'}(i)| \geq 1 - O(\varepsilon)$ for some $i \in \{0\} \cup [n]$. However since $\widehat{f'}(0) = \widehat{f}(\emptyset)$ and $W^1[f] = 1 - \varepsilon$, it cannot be that $i = 0$. Thus, $i \in [n]$ and $|\widehat{f'}(i)| \geq 1 - O(\varepsilon)$; but from the Fourier expansions we directly read off that $\widehat{f'}(i) = \widehat{f}(i)$, so by again applying the comment preceding this theorem we are done.

The previous argument shows that we can assume without loss of generality that f does not have a constant coefficient in its Fourier expansion. Therefore f is of the form

$$f(x) = \sum_{i=1}^{n} \widehat{f}(i)x_i + \sum_{|S|>1} \widehat{f}(S)x_S.$$

By hypothesis, $\sum_{i=1}^{n} \widehat{f}(i)^2 = 1-\varepsilon$, so Parseval's Theorem implies $\sum_{|S|>1} \widehat{f}(S)^2 = \varepsilon$.

Let us define $g, h : \{-1,1\}^n \to \mathbb{R}$ by

$$\begin{aligned} g(x) &\stackrel{\text{def}}{=} \sum_{i=1}^{n} \widehat{f}(i)x_i, \\ h(x) &\stackrel{\text{def}}{=} \sum_{|S|>1} \widehat{f}(S)x_S, \end{aligned}$$

so that $f = g + h$.

Since the square of a Boolean-valued funtion is always 1, we have $f^2 = 1$, where 1 denotes the constant one function. Thus, $(g + h)^2 = 1$, so

$$\begin{aligned} g^2 + 2gh + h^2 &= 1, \\ g^2 + h(2g + h) &= 1. \end{aligned}$$

Since $f = g + h$ this can be rewritten as

$$g^2 + h(2f - h) = 1. \tag{4.3}$$

We now consider these two terms separately.

- First we analyze the term g^2. We have

  $$g(x)^2 = \sum_{i=1}^{n} \widehat{f}(i)^2 x_i^2 + \sum_{i \neq j} \widehat{f}(i)\widehat{f}(j)x_i x_j = \sum_{i=1}^{n} \widehat{f}(i)^2 + \sum_{i \neq j} \widehat{f}(i)\widehat{f}(j)x_i x_j,$$

  so

  $$g^2 = 1 - \varepsilon + q \tag{4.4}$$

  where $q : \{-1, 1\}^n \to \mathbb{R}$ is defined as $q(x) \stackrel{\text{def}}{=} \sum_{i \neq j} \widehat{f}(i)\widehat{f}(j)x_i x_j$.

- Second we analyze the term $h(2f - h)$. As h does not have a constant coefficient in its Fourier expansion, from of the comment after Definition 2.12 we get $\mathrm{E}_x[h(x)] = 0$; also, $\mathrm{E}_x[h(x)^2] = \varepsilon$ by Parseval's Theorem, so we have $\mathrm{Var}[h(x)] = \varepsilon$. Chebychev's inequality implies that

  $$\Pr[|h(x)| \geq 10\sqrt{\varepsilon}] \leq \frac{1}{10^2} = \frac{1}{100}.$$

  So with probability at least 99% we have

  $$|h(x)(2f(x) - h(x))| \leq |h(x)|\,(2|f(x)| + |h(x)|) < 10\sqrt{\varepsilon}(2 + 10\sqrt{\varepsilon}).$$

  Notice that in the term $10\sqrt{\varepsilon}(2+10\sqrt{\varepsilon}) = 20\sqrt{\varepsilon}+100\varepsilon$, the nonnegligible one is $20\sqrt{\varepsilon}$. It is therefore easy to see that for ε small enough we have that the this term is smaller than or equal to $21\sqrt{\varepsilon}$. To be very concrete: letting $\varepsilon' = \varepsilon/10000$, since $\varepsilon < 1$ this bound will hold.

From Equation (4.3) and Equation (4.4) we deduce $h(2f-h) = \varepsilon-q$. Therefore we obtain that $|\varepsilon - q(x)| = |h(x)(2f(x) - h(x))| \leq 21\sqrt{\varepsilon}$ with probability at least 99%. Since it holds that $\varepsilon + 21\sqrt{\varepsilon} \leq 22\sqrt{\varepsilon}$, using the triangle inequality we obtain

$$\Pr[|q(x)| \leq 22\sqrt{\varepsilon}] \geq \frac{99}{100}. \tag{4.5}$$

The idea is now to find an upper bound for $\mathrm{E}[q(x)^2]$ in the form of a (possible big) constant times ε. Since the random variable $q(x)$ has a particularly "innocent" form being a multilinear polynomial composed of random bits $x_1, \dots, x_n$, this will be possible; here is where the Hypercontractive Inequality will come into play.

Claim. Let $q(x_1, \dots, x_n)$ be a random variable which is of multilinear polynomial form, where $x_1, \dots, x_n$ are independent and uniformly distributed on $\{-1, 1\}$. Then if we have $\Pr[|q(x)| \leq 22\sqrt{\varepsilon}] \geq 99/100$, it follows that $\mathrm{E}[q(x)^2] \leq 5000\varepsilon$.

We prove this claim. First, it is clear we may assume without any loss of generality that $\Pr[|q(x)| \leq 22\sqrt{\varepsilon}] = \frac{99}{100}$. Let us say $\mathrm{E}[q(x)^2] = K\varepsilon$. Taking conditional expectations, we have

$$K\varepsilon = \mathrm{E}[q^2] \leq \frac{99}{100}(22\sqrt{\varepsilon})^2 + \frac{1}{100}\,\mathrm{E}[q^2 | q^2 > 484\varepsilon].$$

By the Hypercontractive Inequality (Theorem 3.19) we have $\mathrm{E}[q^4] \leq 9^2\,\mathrm{E}[q^2]^2 = 9^2K^2\varepsilon^2$. Therefore $9^2K^2\varepsilon^2$ is at least

$$\mathrm{E}[q^4] \geq \frac{1}{100}\,\mathrm{E}[q^4 | q^2 > 484\varepsilon] \geq \frac{1}{100}\,\mathrm{E}[q^2 | q^2 > 484\varepsilon]^2 \geq \frac{1}{100}(100K - 99\cdot 484)^2\varepsilon^2.$$

Analyzing the quadratic equation in K gives us a bound: we get $K \in \,]252, 4791[$. In particular, $K \leq 5000$. This ends the proof of the claim.

Now it is quite easy to finish the proof of the FKN Theorem. By Parseval's Theorem we get

$$\mathrm{E}[q(x)^2] = \sum_{i \neq j} \widehat{f}(i)^2\widehat{f}(j)^2 = \Big(\sum_{i=1}^{n} \widehat{f}(i)^2\Big)^2 - \sum_{i=1}^{n} \widehat{f}(i)^4 = (1-\varepsilon)^2 - \sum_{i=1}^{n} \widehat{f}(i)^4.$$

By the above claim we get

$$\sum_{i=1}^{n} \widehat{f}(i)^4 \geq (1-\varepsilon)^2 - 5000\varepsilon = 1 - (5002\varepsilon - \varepsilon^2) = 1 - O(\varepsilon).$$

Also, $\sum_{i=1}^{n} \widehat{f}(i)^4$ is at most $\max_j \widehat{f}(j)^2 \sum_{i=1}^{n} \widehat{f}(i)^2$. But this is just equal to $\max_j \widehat{f}(j)^2$, by Parseval's Theorem. Thus there is an $i \in [n]$ such that $\widehat{f}(i)^2 \geq 1 - O(\varepsilon)$.

There are two possibilities. If $\widehat{f}(i) \geq \sqrt{1 - O(\varepsilon)}$, then by Equation (2.1) we have

$$\mathrm{dist}(f, \chi_i) = \frac{1 - \langle f, \chi_i \rangle}{2} = \frac{1 - \widehat{f}(i)}{2} \leq \frac{1 - \sqrt{1 - O(\varepsilon)}}{2}.$$

By a Taylor expansion we know $\sqrt{1-\delta} = 1 - \frac{1}{2}\delta + O(\delta^2)$ for each $-1 < \delta < 1$, so we get $\mathrm{dist}(f, \chi_i) \leq \frac{1}{4}O(\varepsilon) = O(\varepsilon)$. If $\widehat{f}(i) \leq -\sqrt{1 - O(\varepsilon)}$, we can similarly deduce $\mathrm{dist}(f, -\chi_i) \leq O(\varepsilon)$. We conclude that f is $O(\varepsilon)$-close to either χ_i or $-\chi_i$. ■

Finally we get the robust version of Arrow's Theorem:

Theorem 4.10

> (Robust Version of Arrow's Theorem) *Let $f : \{-1,1\}^n \to \{-1,1\}$ be unanimous. Furthermore, suppose f is such that when doing a 3-candidate Condorcet election based on f, the probability that the outcome is rational under the impartial culture assumption is $1-\varepsilon$. Then there exists an $i \in [n]$ such that f is $O(\varepsilon)$-close to χ_i.*

Proof. From Lemma 4.6 we get $1-\varepsilon \leq \frac{7}{9} + \frac{2}{9}\,\mathrm{W}^1[f]$, i.e., $\mathrm{W}^1[f] \geq 1 - \frac{9}{2}\varepsilon$. The FKN Theorem then implies that there is an $i \in [n]$ such that f is $O(\varepsilon) = O(\frac{9}{2}\varepsilon)$-close to χ_i or $-\chi_i$. However, unanimity of f implies that the latter of these two possibilities is not viable. ■

Our proof is based on O'Donnell [97], which is in turn based on Kalai [68].

4.3 Arrow's Theorem in Practice: A Havoc?

4.3.1 No Need for Overpessimism, Says Maskin

When all is said and done, how do we cope with Arrow's Theorem in practice? Specifically, how "catastrophic" is it, really? In [82], Eric Maskin warns not to be overly pessimistic. One key aspect, he points out, is the (in our formalism so far implicit) assumption that voting rules satisfy the *universal domain axiom*. This axiom states that *all* $m!$ rankings are available for the voters (here m is the number of alternatives); there are no limitations to the voters' will. Roughly stated, Maskin argues that this is a *theoretical* consideration rather than a practical one. One fundamental question is: how do people actually vote in reality? In practice, it might frequently be the case that some preferences are not very likely. If a given voting rule fails to satisfy some of the desirable properties only for such unlikely, "degenerate", instances, then perhaps we should not worry about the paradox after all.

Maskin gives the example of the 2000 US presidential election. Bush and Gore had been neck and neck all along. It had become clear that everything was going to be decided in the Florida election: the candidate (be it Bush or Gore) winning the Florida election, would become the new president. However, the crux of the story is that there actually was another candidate, Nader, who played a decisive role. In fact, we can safely say that his presence in the election resulted in Bush being elected as the president rather than Gore. Remarkably, among about $6,000,000$ votes cast in Florida, the margin between the number of votes for Bush and Gore was only 600 votes. Nearly $100,000$ people voted for Nader.

From a number of trustable sources we know for a fact that a lot of people who in reality voted for Nader, would have voted for Gore in case Nader had not been candidate in the election. As a result, *almost surely Gore would have become president had Nader not run for president.* Nader was, in Maskin's words, a "spoiler". This is a prime example of an election in which the independence of irrelevant alternatives condition was heavily violated: the "irrelevant" alternative (Nader) was not irrelevant at all. On the contrary, to some extent he was more relevant for the choice than the "real" candidates (Bush and Gore) themselves. Quoting Maskin [82], "there is a sense in which that is a highly undemocratic thing to be, because, after all, Nader got less than 2% of the votes in Florida, and yet he ended up determining the entire outcome of the election".

Let us imagine that the voters in the Florida election had cast a complete ranking of the candidates Bush, Gore, Nader.[1] In that case, each voter's preference corresponds to one of the elements of the following set:

$$\left\{ \begin{pmatrix} G \\ B \\ N \end{pmatrix}, \begin{pmatrix} B \\ G \\ N \end{pmatrix}, \begin{pmatrix} N \\ G \\ B \end{pmatrix}, \begin{pmatrix} G \\ N \\ B \end{pmatrix}, \begin{pmatrix} N \\ B \\ G \end{pmatrix}, \begin{pmatrix} B \\ N \\ G \end{pmatrix} \right\}. \tag{4.6}$$

Research has shown that a tremendous number of voters opted for one of the first three rankings, and among those the first two significantly more than the last one. On the other hand, there were extremely few voters selecting one of the last two rankings. This is not surprising, as Bush and Nader were at extreme sides of the political spectrum: Bush was the most right-wing candidate, while Nader was among these three candidates clearly the most left-wing one. Therefore, the combinations "liking Bush most, and *secondly* Nader" as well as "liking Nader most, and *secondly* Bush" are implausible, assuming a mild degree of "ideological consistency". In other words, people who like Nader best, almost always like Bush the least; and people who like Bush best, almost surely like Nader least. Let us call these last two preferences the *ideologically inconsistent* preferences.

The central point of the argument comes now: it follows from a general theorem [121] that, ruling out the two ideologically inconsistent preferences, the majority rule (applied on pairs of candidates) actually *never* gives rise to the Condorcet paradox! This is an instance of what is called *single-peaked preferences* [11]: there exists a "left-to-right" ordering on the alternatives such that any voter prefers x to y if x is between y and her top alternative with respect to the ordering [39]. In many political elections, single-peakedness is a quite natural thing to assume. The majority rule does satisfy all of Arrow's conditions[2] in that case. In practice, besides single-peaked preferences other preference class restrictions which may cause the majority rule to avoid the Condorcet paradox can be found. An interesting example of this, given in [121], is the 2002 French presidential election. At the time, the three principal contenders were Jospin, Chirac, and Le Pen. According to research, Le Pen, from the extreme-right Front National, had strong polarizing effects: nearly 100% of the electorate ranked him either *first*, or *last*, among the three candidates. It can be shown that such restriction (namely, that there

[1] To be clear: this did not happen in reality. We are pursuing a "thought experiment".

[2] These are: the Pareto condition, IIA, non-dictatoriality, and *decisiveness*. We say that a voting rule is *decisive* if it manages to avoid the Condorcet paradox (equivalently, there is always a Condorcet winner, or the rule is rational).

is one candidate who is ranked second by no voter) causes the majority rule to be decisive. Having said that, some research indicates that most real-life electorates do not fully satisfy restrictions such as single-peakedness [42].

According to Maskin [82], the natural follow-up question regarding Arrow's impossibility theorem is: given that the theorem tells us that no voting rule satisfies all aforementioned desirable properties *all the time*, which voting rule satisfies them *as often as possible*? Metaphorically speaking, we know for sure that we cannot reach the top of the mountain, but which path leads us to the highest reachable point, giving the most enjoyable view?

Formalizing[3] this question somewhat, we consider *classes* of preferences. By definition, this is just a subset of the set of all preferences: e.g., the first four preferences from (4.6) form a class of preferences, call it $\mathcal{C}$. Given a fixed voting rule and a class of preferences, we say that the given voting rule *works well for that class of preferences* if the voting rule satisfies all desirable properties (the Pareto condition, IIA, non-dictatoriality, and decisiveness) from Arrow's Theorem provided all voters' preferences belong to that class of preferences. For example, the majority rule works well for the class $\mathcal{C}$. Our aim, then, is to look for voting rules that "work well" for as large as possible classes of preferences. Remarkably, this aim can in some sense indeed be achieved. The answer lies in the so-called *Domination Theorem* (Theorem 2 in *On the Robustness of Majority Rule* [121]), which, informally speaking, says:

Theorem 4.11

> (Domination Theorem) *Let f be a voting rule and let $\mathcal{C}$ be a class of preferences. If f works well for $\mathcal{C}$, then the following is satisfied:*
>
> 1. *The majority rule also works well for $\mathcal{C}$.*
> 2. *In addition, there is a class of preferences $\mathcal{C}'$ for which the majority rule works well, but f does not.*

In other words, whenever a voting rule works well, also the majority rule works well, but there are instances in which the majority rule works well but the given voting rule does not. In this way we can say that the majority rule *dominates* all other voting rules.

Quoting Maskin [82], "...we go back to this very old method, hundreds of years old: majority rule. I think that's in a sense a satisfying conclusion to draw. (...) He was led to consider all other possible voting rules as alternatives. But there's a sharp sense in which, in the end, we actually can't really do much better than majority rule after all".

[3] We will not be very formal, and refer to [121] for the mathematical details. Our aim is to convey the idea rather than to be rigorous. The "theorem" stated below is in this sense not really a mathematical result, as we refrain from introducing all required concepts formally.

4.3.2 Occurrence of the Condorcet Paradox: Some Empirical Findings

We review some data from real-life elections. To start, we ought to mention that in this context almost all empirical evidence suffers from the same affliction: nearly each piece of available data on elections excludes the voters' *pairwise* comparisons of all candidates. However, in the context of the occurrence of the Condorcet paradox, the latter happens to be the most salient piece of evidence. For example, in the Riker-Mackie discussion from Subsection 1.3.2, we noted that Riker speculated on how US voters *would have voted* had they been required to submit a complete ranking of the four candidates in the 1860 election. It is undoubtedly in part for this reason—lack of reliable data—that, as we will shortly see, researchers still have not reached a consensus regarding the actual rarity of the Condorcet paradox.

		m			
		3	4	5	6
	3	0.94444	0.88889	0.84000	0.79778
	5	0.93056	0.86111	0.80048	0.74865
	7	0.92498	0.84977	0.78467	0.72908
	9	0.92202	0.84405	0.77628	0.71873
	11	0.92019	0.84037	0.77108	0.71231
	13	0.91893	0.83786	0.76753	0.70794
	15	0.91802	0.83604	0.76496	0.70476
	17	0.91733	0.83466	0.76300	0.70235
	19	0.91678	0.83357	0.76146	0.70046
	21	0.91635	0.83269	0.76023	0.69895
	23	0.91599	0.83197	0.75921	0.69769
n	25	0.91568	0.83137	0.75835	0.69664
	27	0.91543	0.83085	0.75763	0.69575
	29	0.91521	0.83041	0.75700	0.69498
	31	0.91501	0.83003	0.75646	0.69431
	33	0.91484	0.82969	0.75598	0.69373
	35	0.91470	0.82939	0.75556	0.69321
	37	0.91456	0.82913	0.75519	0.69275
	39	0.91444	0.82889	0.75485	0.69234
	41	0.91434	0.82867	0.75455	0.69196
	43	0.91424	0.82848	0.75427	0.69162
	45	0.91415	0.82830	0.75402	0.69132
	47	0.91407	0.82814	0.75379	0.69104
	49	0.91399	0.82799	0.75358	0.69078
	⋮	⋮	⋮	⋮	⋮
	Limit	0.91226	0.82452	0.74869	0.68476

Table 4.1: Probability under ICA of having a Condorcet winner in a majority election for n voters and m alternatives. Taken from [56].

Source	# elections	# candidates	# voters	Paradox?
Flood [47]	1	16	21	No
Riker [109]	1	4	255	Yes
Riker [110]	1	3	426	Yes
Taylor [135]	1	3	16	Yes
Niemi [93]	22	3 – 6	81 – 463	Yes (1)
Blydenburgh [14]	2	3	386	Yes (1)
Fishburn [44]	1	5	175	No
Dyer & Miles [34]	1	36	10	No
Bjurulf & Niemi [10]	1	3	87	Yes
Norpoth [94]	5	3	818 – 1872	No
Dobra & Tullock [31]	1	37	4 – 6	No
Riker [111]	2	3 – 4	172	Yes (2)
Toda et al. [138]	1	6	5281	No
Dobra [32]	32	3 – 37	4 – 27	Yes (1)
Chamberlin et al. [20]	5	5	> 11000	No
Dietz & Goodman [30]	1	4	Large	No
Fishburn & Little [45]	3	3 – 5	> 1500	No
Rosen & Sexton [113]	1	4	31	No
Radcliff [105]	4	3	Large	No
Neufeld et al. [92]	1	3	70	Yes
Gaubatz [53]	1	4	Large	Yes
Browne & Hamm [18]	1	3	621	Yes
Vergunst [146]	1	3	150	Yes
Lagerspetz [73]	10	3 – 4	300	Yes (3)
Beck [5]	3	4 – 8	20	Yes (1)
Flanagan [46]	1	3	224	Yes
Morse [84]	1	4	52	Yes
Taylor [134]	1	3	Large	No
Hsieh et al. [64]	1	3	350	No
Taplin [133]	1	4	12	No
Regenwetter et al. [107]	7	3	Large	Yes (1)
Truchon [139]	24	5 – 9	5 – 23	No
Van Dam [142]	1	3	133	Yes
Van Deemen et al. [143]	4	9 – 13	1500	No
Stensholt [127]	1	3	165	Yes
Kurrild-Klitgaard [71]	1	20	Large	No
Tideman [137]	87	3 – 29	9 – 3500	No
Regenwetter et al. [108]	4	5	Large	No
Kurrild-Klitgaard [72]	8	9 – 11	> 1000	No
Smith [124]	1	4	Large	Yes
Bochsler [15]	1	3	Large	Yes

Table 4.2: Summary of the occurence of the Condorcet paradox in several elections. The first column names the sources, which are put in chronological order. The first three columns represent the number of elections, candidates, and voters, respectively. The last column indicates whether an instance of the Condorcet paradox was observed. If so, the number of paradoxes observed is written in parentheses in case the number of elections is greater than one. This table is taken from van Deemen [145].

Table 4.1 suggests that, under ICA, the probability of cycles increases as the number of alternatives increases and also as the number of individuals increases. Note that the table confirms the value of 91.2% for the probability of having a Condorcet winner in case $m = 3$, which we obtained in Equation (4.2). This table, together with the impossibility results that this thesis has been dealing with, have initially led some researchers (Riker is just one example) to adopt a pessimistic attitude regarding the feasibility of meaningful consensus formation [102]. Additionally, given a large number of voters, the probability rapidly increases with the number of alternatives. For example, for $m = 3$ the probability of encountering the Condorcet paradox under ICA is 8.77%, for $m = 5$ already 25.13%, further 45.45% for $m = 9$, and 51.87% for $m = 11$ [145]. In summary, under the impartial culture assumption the Condorcet paradox is ubiquitous, the more so when the number of voters or the number of alternatives is high.

Conspicuously, these results are out of line with a great part, though not all, of the empirical findings. An overview due to van Deemen [145] is listed in Table 4.2 on page 71. We immediately see that both Yes and No are omnipresent. All in all, in the 265 elections listed in Table 4.2 we find 25 empirical[4] occurrences of the Condorcet paradox. This comes down to about 9.4% of all[5] elections.

We give some remarks.

- Researchers have quite different viewpoints as to the occurrence of the Condorcet paradox. In Subsection 1.3.2 we have seen that Mackie argues that cycles are empirically improbable; in *Democracy Defended* [79], he contends that most of Riker's claims to the contrary are false. Researchers from the social sciences, most notably psychology, often seem to think so, too. Quoting the psychologist Regenwetter [106], "Theoretical and empirical work in psychological science takes a more optimistic perspective [than the mathematically sophisticated literature]. Psychologists have found little empirical evidence for voting paradoxes and, so far, little behavioral support for the famed incompatibility of social choice methods. These findings suggest that technical assumptions or theoretical problem formulations in social choice theory might bias our understanding".

- One can try to explain the fact that the paradox occurrence probabilities from Table 4.1 seem relatively high compared to the empirical data from Table 4.2. In [140], Regenwetter et al. showed that, for three candidates, the impartial culture assumption is the "worst-case scenario" among a

[4] The data from the table vary greatly in nature: German political elections during the 60s and 70s [94], a case treated in the Dutch Tweede Kamer [146], more Dutch politics [142], a Swiss referendum in Bern [15], and a U.S. federal government policy case [92], just to name a few. We refer to [145] for a more complete overview.

[5] Table 4.2 is a somewhat (but not entirely) complete overview of the empirical evidence from the literature that deals with the Condorcet paradox. (One interesting recent research not included in Table 4.2 is [83], in which Netflix data are used to generate several million elections; remarkably, the Condorcet paradox is almost completely absent ($< 0.4\%$).) In any case, the number 9.4% is rather artificial, so we should be careful not to overvalue it.

very broad range of possible voter preference distributions. More precisely, any deviation from the impartial culture over linear orders reduces the probability of majority cycles in infinite samples (unless, of course, the culture from which we sample is itself inherently intransitive). List [76] (Chapter 4) had already observed a similar result.

- Some research [74, 144] suggests that the Condorcet paradox may be less salient in case orderings are modeled as weak (i.e., a partial ordering). rather than total orderings. For example, in [74] shows that for three candidates, the paradox probability rapidly *decreases* for increasing number of voters. This might explain why in a significant number of cases the Condorcet paradox cannot be found. We refer to [145] for a more in-depth discussion.

In conclusion, van Deemen [145] argues that it is too soon to come to a final conclusion regarding the actual occurrence of the Condorcet paradox. The occurrence percentage of 9.4% from Table 4.2 that we got is too big to neglect; however, to use these data to contend that the paradox is omnipresent (such as, among others, Riker [111] argues), is not justifiable either. The fact of the matter is that the empirical evidence collected so far is, according to him, "casual" and mainly "ad hoc", and thus insufficient either to confirm or to refute the statement that the paradox is empirically relevant. Quoting van Deemen [145], "...we are still far away from being able to settle the question of the empirical relevance of Condorcet's paradox. Much research still has to be done".

Chapter 5

Gibbard-Satterthwaite Theorem

In this chapter we investigate the Gibbard-Satterthwaite Theorem. First, we give an introduction. After that, we state and prove the theorem in classical setting. Then we arrive at our main goal of this chapter: we show a quantitative version of the Gibbard-Satterthwaite Theorem, which says that a random manipulation by a single random voter will succeed with a non-negligible probability for any election rule among three alternatives that is “far” from being a dictatorship and from having only two alternatives in its range.

Further, we give an overview of other recent approaches and improvements. Some of these results have important consequences in computational social choice theory. Most notably we will see that certain advances imply that using computational complexity as a barrier against manipulation cannot solve the problem of manipulation completely; the reason is that manipulation is easy on average, unless the voting rule in question is very unsatisfying, namely, either very similar to a dictator or none but at most two candidates can get elected.

All the definitions and results from this chapter are based on [57, 39, 38, 50, 88]. Unless specifically mentioned, all our results come from these sources.

5.1 Introduction

The Gibbard-Satterthwaite Theorem is a result about deterministic voting systems that choose a single winner using the preferences of all voters. Here, each voter ranks all candidates in order of preference. We will represent this mathematically by a linear order; in the next section we will elaborate on the used mathematical notations.

The theorem deals with *strategic voting*. Strategic voting arises when voters purposely vote untruthfully in order to achieve a for them better outcome according to their “true” preferences. Indeed, sometimes it is in the voters' interests not to reveal their true preferences. Consider for example the following situation, taken from [38], which connects well with the discussion from Subsection 4.3.1:

$$\begin{array}{lccccc} 49\% : & \text{Bush} & > & \text{Gore} & > & \text{Nader} \\ 20\% : & \text{Gore} & > & \text{Nader} & > & \text{Bush} \\ 20\% : & \text{Gore} & > & \text{Bush} & > & \text{Nader} \\ 11\% : & \text{Nader} & > & \text{Gore} & > & \text{Bush} \end{array}$$

Suppose that the plurality rule is used, i.e., the candidate most often elected first is the winner. Let us moreover assume the above preferences are truthful. According to these preferences, Bush wins. However, the Nader supporters had better voted differently: if instead they had voted Gore as their top preference, then Gore would have won the contest instead of Bush—which they, the Nader fans, would have liked better.

The question arises if there is a reasonable voting rule for at least three candidates that manages to avoid this problem altogether. The Gibbard-Satterthwaite Theorem answers this question negatively. It says that for any voting rule for at least three candidates at least one of the following properties must hold:

1. The rule is *dictatorial*, i.e., the winner is chosen by one particular fixed voter.
2. There is a candidate who can never win under the rule.
3. The rule is susceptible to *strategic voting*, i.e., there are conditions under which a voter with full knowledge of how the other voters are to vote and of the rule being used would have an incentive to vote in a manner that does not reflect that voter's true preferences.

A rule not susceptible to strategic voting is called *strategy-proof*. Note the presence of the full-information assumption in the above: when studying strategy-proofness, we make the classical assumption that the manipulator has full information about the ballots of the other voters. Although this assumption is far from realistic, it is not bad to inoculate against manipulation even in such a "worst-case" model in which everyone has full knowledge about the entire situation.

5.2 Classical Formulation and Proof

In this section we formulate and prove the classical Gibbard-Satterthwaite Theorem. There are two main reasons for doing so. First, to get the reader acquainted with the notion of manipulability. Second, to show the reader the difference between the qualitative proofs and quantitative proofs. A reader who is already familiar with the classical Gibbard-Satterthwaite Theorem is advised to skip to Section 5.3.

5.2.1 Preliminaries

In this section we briefly introduce the formalism necessary to formulate the classical Gibbard-Satterthwaite Theorem. Most of the notations are based on [39, 38].

- $\mathcal{N} = \{1, 2, \ldots, n\}$ with $n \geq 2$ is the set of voters;
- $\mathcal{X}$ is the set of alternatives;

- $\mathcal{L}$ is the set of all irreflexive total linear orders on $\mathcal{X}$;
- A *preference*, also called *ballot*, is an element of $\mathcal{L}$;
- A *profile* is a vector $R = (R_1, \ldots, R_n) \in \mathcal{L}^{\mathcal{N}}$ of ballots, where R_i is voter i's preference;
- For given alternatives $x, y \in \mathcal{X}$, we denote with $\mathcal{N}^R_{x>y}$ the set of all voters which rank alternative x above alternative y under profile R;
- A *social choice function* is a function $\mathcal{L}^{\mathcal{N}} \to \mathcal{P}(\mathcal{X}) \setminus \{\emptyset\}$. It is called *resolute* if $|F(R)| = 1$ for all profiles R, i.e., if there is always a unique winner. In that case we write $F(R) = \{x\}$ shorthand as $F(R) = x$.

We say that a resolute social choice function $F : \mathcal{L}^{\mathcal{N}} \to \mathcal{X}$ is

- *Pareto*, if whenever all voters rank x above y, then y cannot win, i.e., for all $x, y \in \mathcal{X}$ and each profile R it is the case that $\mathcal{N}^R_{x>y} = \mathcal{N}$ implies $y \neq F(R)$.
- *surjective*, also called *nonimposed*, if for each $x \in \mathcal{X}$ there is a profile R such that $F(R) = x$.
- *S-dictatorial*, where $S \subseteq \mathcal{X}$ is nonempty, if there exists an $i \in \mathcal{N}$ such that $F(R) = \text{top}(R_i|_S)$ for all profiles $R = (R_1, \ldots, R_n)$. (Here $R_i|_S$ is the restriction of R_i to S and $\text{top}(R_i|_S)$ is the maximal element in the order $R_i|_S$.)
- *dictatorial*, if F is S-dictatorial for some nonempty $S \subseteq \mathcal{X}$.
- *range-dictatorial*, if F is range(F)-dictatorial.
- *strong-dictatorial*, if F is $\mathcal{X}$-dictatorial.
- *weakly-dictatorial*, if F is a 1-junta, i.e., if F is a function of only one coordinate.
- *independent*, if for all alternatives $x \neq y$ and all profiles R, R', if it holds that $F(R) = x$ and the relative rankings of x versus y do not change, then y will still not win under R'. That is,

  $$\forall x, y \in \mathcal{X}\ \forall R, R'\ [\ (x \neq y\ \&\ F(R) = x\ \&\ \mathcal{N}^R_{x>y} = \mathcal{N}^{R'}_{x>y}) \Rightarrow F(R') \neq y\ ].$$

- *weakly monotonic*, if for any alternative x and any profile R for which x is the winner, if R' is a profile in which, with respect to R, the alternative x has been moved up (or stayed the same) in each voter's preference while all other relative preferences have remained the same, then x is still the winner under R'. That is,

  $$\forall x \in \mathcal{X}\ \forall R, R'\ [\ \forall y, z \in \mathcal{X} \setminus \{x\}\ (\mathcal{N}^R_{x>y} \subseteq \mathcal{N}^{R'}_{x>y}\ \&\ \mathcal{N}^R_{y>z} = \mathcal{N}^{R'}_{y>z})$$
  $$\Rightarrow (x = F(R)\ \Rightarrow\ x = F(R'))\ ].$$

- *strongly monotonic*, if for any alternative x and any profile R for which x is the winner, if R' is a profile for which for each voter it holds that any alternative that was lower-ranked than x in R is still ranked lower than x in R', then x is still the winner under R'. That is,

$$\forall x \in \mathcal{X}\ \forall R, R'\ [\ \forall y \in \mathcal{X} \setminus \{x\}\ (\mathcal{N}^{R}_{x>y} \subseteq \mathcal{N}^{R'}_{x>y})$$
$$\Rightarrow (x = F(R)\ \Rightarrow\ x = F(R'))\ ].$$

  The difference between weak monotonicity and strong monotonicity is that in the latter it is allowed for the other relative rankings to change, whereas in the former it is not.

- *strategy-proof*, also called *immune to manipulation*, if for no individual $i \in \mathcal{N}$ there exists a profile R (including the "truthful preference" R_i of voter i) and a linear order R'_i (representing the "untruthful" ballot of i) such that $F(R_{-i}, R'_i)$ is ranked above $F(R)$ according to R_i. (Following notation from game theory, given a profile R, with (R_{-i}, R'_i) we mean the profile obtained when in R we replace R_i by R'_i.)

Note the relations between the several notions of dictatorships:

$$\text{strong dictator} \ \Rightarrow\ \text{range dictator} \ \Rightarrow\ \text{dictator} \ \Rightarrow\ \text{weak dictator}. \qquad (5.1)$$

5.2.2 The Muller-Satterthwaite Theorem

Intuitively this theorem says that surjectivity and strong monotonicity are jointly too strong for a voting rule in order to still be "reasonable". The proof essentially consists of a reduction to Arrow's Theorem for social choice functions, as will become apparent in the proof.[1]

One way to formulate the theorem is as follows.

Theorem 5.1

> (Muller-Satterthwaite Theorem) *Any resolute social choice function for at least three alternatives that is surjective and strongly monotonic must be a strong dictatorship.*

Proof. As said before, the proof is essentially a reduction to Arrow's Theorem. The hardest part is proving Arrow's Theorem; the remainder of the proof is not at all complicated.

Let F be a social choice function that is surjective and strongly monotonic. The proof plan is as follows:

(1) We show that strong monotonicity implies independence.

[1]Note that Arrow's Theorem is originally about social *welfare* functions, i.e., about functions which take profiles and return *a linear order* (rather than a unique winner, as is the case with resolute social choice functions). However, with the same effort we can prove Arrow's Theorem also when voting rules are instead formalized as social choice functions. We will do so in Subsection 5.2.4.

(2) Using (1), we show that F is Pareto.

(3) Finally, we prove a version of Arrow's Theorem for social choice functions.

(1) Let $x, y \in \mathcal{X}$ and let R, R' be profiles such that $x \neq y$, $F(R) = x$, and the relative rankings of x and y in R and R' are the same. We have to show that $F(R') \neq y$.

Let R'' be any profile satisfying the following conditions:

- x and y are ranked at the top-two positions for every voter;
- the relative rankings of x and y are as in R.

Note that R'' is a profile for which for each voter it holds that any alternative that was lower-ranked than x in R is still ranked lower than x in R''. Thus strong monotonicity implies that $F(R'') = x$. If, by contradiction, we would have $F(R') = y$, then because of a similar reasoning by strong monotonicity we would obtain $F(R'') = y$, implying the contradiction $x = y$. Therefore we have $F(R') \neq y$.

(2) Let $x \neq y \in \mathcal{X}$. Let R be a profile such that $\mathcal{N}^R_{x>y} = \mathcal{N}$. We have to show that $F(R) \neq y$.

By surjectivity, let R' be such that $F(R') = x$; by strong monotonicity (in fact, weak monotonicity suffices) if necessary we can move x to the top of each voter's ranking, so without loss of generality we may suppose x tops each voter's ranking in R'. Now note that, as $\mathcal{N}^R_{x>y} = \mathcal{N} = \mathcal{N}^{R'}_{x>y}$, the relative rankings of x and y are the same in R' and R. Therefore independence implies $F(R) \neq y$.

(3) Arrow's Theorem for social choice functions, which says that any resolute social choice function for at least three alternatives that is independent and Pareto must be a dictator, is proved in Subsection 5.2.4 below.

Now (1), (2), and (3) together imply that F is a dictator. ■

This proof is based on [39].

We can strengthen the Muller-Satterthwaite Theorem a little by replacing the assumption that F be surjective and defined for at least three alternatives by the slightly weaker assumption that F be a voting rule the range of which has at least three elements. To do that, we use the notion of range-dictatoriality. The reason for doing so is that a strong dictator function is always surjective, whereas a range-dictator function need not necessarily.

Then we have the following stronger version of the Muller-Satterthwaite Theorem:

THEOREM 5.2

*(*STRONG MULLER-SATTERTHWAITE THEOREM*) A resolute social choice function whose range contains at least three alternatives and which satisfies strong monotonicity must be a range dictatorship.*

Proof. Suppose F is a social choice function that has at least three alternatives in its range and that satisfies strong monotonicity. If F is surjective then we are done by the previous theorem, so we may assume F is not surjective.

The idea is a reduction to the weaker version of the theorem, Theorem 5.1. For notational convenience we assume that

$$\mathcal{X} = \{x, y, z, t\}, \quad \mathrm{range}(F) = \{x, y, z\}, \quad \mathcal{X} \setminus \mathrm{range}(F) = \{t\}.$$

Below it will become apparent that this assumption is without loss of generality.

We now define, for all fixed positions of the alternative t in the voters' rankings, a "reduction function" which is a social choice function for the alternatives $\{x, y, z\}$. To do this, let $(p_1, p_2, \ldots, p_n)$ be *a list of "positions"*, i.e., $p_i \in \{1, 2, 3, 4\}$ for each $i \in [n]$. The idea is that p_i indicates the position that the alternative t has in voter i's preference (given that voter i already made up his mind about the ranking of the alternatives x, y, z). Concretely, if we let the alternatives x, y, and z be symbolized by dots, then in the below picture the circles represent the four respective positions where t may be placed:

$$\bigcirc > \cdot > \bigcirc > \cdot > \bigcirc > \cdot > \bigcirc$$

From left to right the first circle we call the first position, the second circle is the second position, etc. For example if voter i votes according to $z > x > t > y$ then we have $p_i = 3$, and if voter j votes according to $t > z > y > x$ then we have $p_j = 1$.

For such a fixed list of positions $p = (p_1, p_2, \ldots, p_n)$, we define a social choice function F_p for the three alternatives x, y, z as follows. For any given $\{x, y, z\}$-profile R, let R^p be the $\{x, y, z, t\}$-profile which is exactly like R but with alternative t placed at spot p_i for each voter $i \in [n]$. Then we define $F_p(R) \stackrel{\text{def}}{=} F(R^p)$.

Of course, F_p is still strongly monotonic, and by construction it is surjective as well. Therefore the Muller-Satterthwaite Theorem that we already proved implies that F_p is a dictator function; say the i-th voter is the dictator.

We now claim that for any choice of list of positions we make, we always get *the same* dictator:

Claim. *Let $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$ be two lists of positions. Then we have that $F_p = F_q$ and furthermore this function is dictatorial.*

Proof of claim. Because of the above reasoning both F_p and F_q are dictator functions. By contradiction, suppose they are different; without loss of generality we may assume that the first voter is the dictator F_p and the second voter is the dictator F_q.

Let R be any $\{x, y, z\}$-profile in which the voters vote as follows:

- Voter 1: $x > y > z$;
- Voter 2: $y > x > z$.

Then we have $F(R^p) = F_p(R) = x$ and $F(R^q) = F_q(R) = y$. However, from the proof of part (1) of Theorem 5.1 we know that F satisfies independence.

But clearly the x-versus-y rankings are the same in R^p and R^q, so the fact that x wins in R^p implies that y cannot win in R^q, i.e., $F(R^q) \neq y$. This contradiction finishes the proof of the claim. □

Because of the claim we can say that F_p is the i-th dictator function, for *every* list of positions $p = (p_1, p_2, \ldots, p_n)$. But this immediately implies that F is a range-dictator function, namely the i-th voter is the range-dictator. ■

Because of (5.1) we have that Theorem 5.2 also holds when the word range dictatorship is replaced by dictatorship, or by weak dictatorship.

5.2.3 Classical Gibbard-Satterthwaite Theorem: Statement and Proof

The precise formulation of the theorem is as follows:

THEOREM 5.3

If any resolute social choice function for at least three alternatives is both surjective and strategy-proof, then it must be a dictatorship.

Proof. Because of Lemma 5.4 below, we have that the given social choice function is strongly monotonic as well; the result then follows immediately by an application of the Muller-Satterthwaite Theorem. ■

LEMMA 5.4

For a resolute social choice function, strategy-proofness implies strong monotonicity.

Proof. We use contraposition. So let us assume F is not strongly monotonic; we then show that F is not strategy-proof.

Since F is not strongly monotonic, there exist $x, x' \in \mathcal{X}$ such that $x \neq x'$ and profiles R, R' such that

- $\mathcal{N}^R_{x>y} \subseteq \mathcal{N}^{R'}_{x>y}$ for all alternatives $y \in \mathcal{X} \setminus \{x\}$; $(*)$
- $F(R) = x$ and $F(R') = x'$.

There must be a *first* voter affecting the outcome. We can go from R to R' by changing one coordinate at a time. Hence there will exist two profiles (which for simplicity we call again R and R') differing on one exactly one coordinate, say i, which satisfy the above conditions. Now there are two cases:

Case 1: $i \in \mathcal{N}^{R'}_{x>x'}$. That is, $x >_i x'$ in R'. Say voter i's "true preferences" are as in R'. Then voter i can benefit from voting instead as in R. Thus F is not strategy-proof.

Case 2: $i \notin \mathcal{N}^{R'}_{x>x'}$. By $(*)$ we get $i \notin \mathcal{N}^{R}_{x>x'}$. That is, $x' >_i x$ in R. Say voter i's "true preferences" are as in R. Then voter i can benefit from voting instead as in R'. In consequence, once again F is not strategy-proof. ■

Because of the stronger version of the Muller-Satterthwaite Theorem, namely Theorem 5.2, we have the following:

Theorem 5.5

> *Any resolute social choice function which has at least three alternatives in its range and which is strategy-proof must be a range dictatorship.*

Because of Theorem 5.2, Lemma 5.4, and (5.1), we have the following:

Theorem 5.6

> *(*Gibbard-Satterthwaite Theorem*) If any resolute social choice function takes at least three values and is not a weak dictator (i.e., not depending on one voter only), then it is manipulable.*

Usually it is this version of the theorem which is called the Gibbard-Satterthwaite Theorem.

5.2.4 Arrow's Theorem for Social Choice Functions

We will prove Arrow's Theorem for social choice functions, based on [39]. The reason for having deferred this proof is that one proof strategy to show Arrow's original result, which deals with social welfare functions, applies equally well also when showing the result for social choice functions; in fact the proof can essentially be copied. Exactly the same proof approach works in both formalisms: the *decisive coalition method*. This proof approach is demonstrated below.

Theorem 5.7

> *(*Arrow's Theorem for resolute social choice functions*) Any resolute social choice function for at least three alternatives that is independent and satisfies the Pareto condition must be a dictatorship.*

Proof. Let F be a resolute social choice function for at least three alternatives that is independent and Pareto. The proof plan is as follows:

(1) We show the *Coalition Lemma* (to be explained below).

(2) Using (1), we prove the *Reduction Lemma* (to be explained below).

(3) We finish by applying the Reduction Lemma multiple times.

Before we start, we first need some more terminology. A *coalition* is just a nonempty subset of $\mathcal{N}$.

Given distinct alternatives $x, y \in \mathcal{X}$, we say that a coalition $G \subseteq \mathcal{N}$ is *decisive on* (x, y) if and only if

$$\forall R \, (G \subseteq \mathcal{N}^R_{x>y} \Rightarrow F(R) \neq y).$$

Given distinct alternatives $x, y \in \mathcal{X}$, we say that a coalition $G \subseteq \mathcal{N}$ is *weakly decisive on* (x, y) if and only if

$$\forall R\ (G = \mathcal{N}^R_{x>y} \Rightarrow F(R) \neq y).$$

The idea behind the proof is the so-called *decisive coalition strategy.*

(1) **Coalition Lemma.** *For any distinct alternatives $x, y \in \mathcal{X}$ and $G \subseteq \mathcal{N}$, if G is weakly decisive on (x, y), then G is decisive on (x', y') for all distinct $x', y' \in \mathcal{X}$.*

Proof. Take distinct $x', y' \in \mathcal{X}$. We assume x, y, x', y' are all distinct, as all other cases are proved analogously.

Let R be any profile satisfying the following conditions:

- G-members: $x' > x > y > y' >$ rest;
- non-G-members: $x' > x, \quad y > y', \quad y > x, \quad x, y, x', y' >$ rest.

Note that here the x'-versus-y' rankings were *not* specified for the non-G-members; this will in fact be crucial further on in the proof.

Since G is weakly decisive on (x, y) and the G-members are precisely the ones ranking $x > y$, we have $F(R) \neq y$. All voters rank $x' > x$, so because of the Pareto condition we get $F(R) \neq x$; similarly we get $F(R) \neq y'$ as well. The Pareto condition implies also that no alternative in $\mathcal{X} \setminus \{x, x', y, y'\}$ can win. Therefore, $F(R) = x'$.

We now show that G' is decisive on (x', y'). Let R' be such that $G \subseteq \mathcal{N}^{R'}_{x'>y'}$. Then let R be any profile satisfying the above two conditions but which furthermore is such that the set of non-G-members which satisfy $x' > y'$ in R is exactly equal to $\mathcal{N}^{R'}_{x'>y'} \setminus G$. (We can do this precisely because earlier we had left the x'-versus-y' rankings for non-G-members undetermined.) Then, of course, we still have $F(R) = x'$. However the relative rankings of x'-versus-y' are, by construction, the same in R and R'. Hence independence of F implies that $F(R') \neq y'$, concluding the proof of the Coalition Lemma. $\square$

(2) **Reduction Lemma.** *Suppose $G \subseteq \mathcal{N}$ is decisive on all pairs of alternatives (x, y), and $|G| \geq 2$. Then there exists a smaller coalition $G' \subset G$ that is decisive on all pairs as well.*

Proof. Let G_1 and G_2 be disjoint nonempty subsets of G such that $G = G_1 \cup G_2$; these can be found for G has at least two elements. Let x, y, z be three distinct alternatives, which exist by assumption.

Let R be any profile satisfying the following conditions:

- G_1-members: $x > y > z >$ rest;
- G_2-members: $y > z > x >$ rest;
- all other voters: $z > x > y >$ rest.

Since $G = G_1 \cup G_2$ is decisive on (y, z), we have $F(R) \neq z$. Because F is Pareto, $F(R) \notin \mathcal{X} \setminus \{x, y, z\}$, and hence we have $F(R) \in \{x, y\}$. Consequently there are two cases:

Case 1: $F(R) = x$. Then by independence we get that G_1 is weakly decisive on (x, z). By the Coalition Lemma we obtain that G_1 is decisive on all pairs.

Case 2: $F(R) = y$. Similarly as above, by independence we obtain that G_2 is weakly decisive on (y, x). By the Coalition Lemma we obtain that G_2 is decisive on all pairs. □

(3) Since F is Pareto we have that $\mathcal{N}$ is decisive on *all* pairs. Now we can apply the Reduction Lemma over and over again; finally we obtain a coalition of size one, say $\{i\}$, that is decisive on all pairs. Then the i-th voter is a dictator, finishing the proof of Arrow's Theorem. ■

5.3 Quantifying the Gibbard-Satterthwaite Theorem

The classical Gibbard-Satterthwaite Theorem says roughly speaking that for any "reasonable" voting rule manipulation is possible. It does, however, say nothing about *how frequently* manipulation may occur. In particular, it is possible that manipulation is conceivable only a very small, perhaps negligible, fraction of time. Quantitative approaches to the Gibbard-Satterthwaite Theorem intend to answer such questions. It will turn out that manipulation is easy *on average*, giving a negative answer to the above question.

We will give an overview of the developments regarding the quantification of the Gibbard-Satterthwaite Theorem. The first, from 2011, is due to Friedgut, Keller, Kalai, and Nisan [50]. It builds directly on the quantitative version of Arrow's Theorem from Kalai [68] that we proved in Subsection 4.2.3.

This publication prompted a great deal of follow-up work, all attempting to generalize or strengthen (by weakening the assumptions) previously obtained results. We will focus on only one of these papers, due to Mossel and Rácz [88], which contains the strongest quantitative version of the Gibbard-Satterthwaite Theorem to date. In the last subsection we discuss the consequences of the achieved results, in particular regarding the idea of using computational hardness against manipulation.

In this section, all voting rules are assumed to be *resolute*, meaning that for any input profile there is exactly one winner.

5.3.1 The Friedgut-Keller-Kalai-Nisan Approach

In this subsection we sketch the approach given in [50].

We use the same terminology as introduced in Subsection 5.2.1. As always, n is the number of voters, and in what follows we let m be the number of alternatives. We now write profiles as $x = (x_1, \ldots, x_n) \in \mathcal{L}$. A *profitable manipulation by voter* i *at the profile* $(x_1, \ldots, x_n)$ is a preference $x_i' \in \mathcal{L}$ such that $F(x_i', x_{-i})$ is preferred by voter i over $F(x_i, x_{-i})$. Note that this definition makes sense because all voting rules are assumed to be resolute. A profile is called *manipulable* if there exists a profitable manipulation for some voter at that profile.

Definition 5.8

> Let F be a social choice function. The *manipulation power of voter i on F*, denoted as $\mathrm{M}_i(F)$, is the probability that x_i' is a profitable manipulation of F by voter i at profile $(x_1, \dots, x_n)$, where $x_1, \dots, x_n, x_i'$ are chosen uniformly at random among all elements of $\mathcal{L}$. The *total manipulation power of F* is $\sum_{i=1}^n \mathrm{M}_i(F)$.

Once again, note the presence of the impartial culture assumption in this definition; all quantitative results in this thesis will typically require it. Also, notice the similarity between this notion and the notion of *influence* from Definition 3.1.

It might at first seem a bit strange that in this definition the x_i' is taken at random. To get some intuition, what is $\mathrm{M}_i(\mathrm{Maj}_n)$ equal to? The answer lies in Equation (3.1): remember that $\mathrm{Inf}_i[\mathrm{Maj}_n] = \sqrt{2/\pi}/\sqrt{n} + O(n^{-\frac{3}{2}}) = O(1/\sqrt{n})$. Therefore, only a $O(1/\sqrt{n})$ fraction of profiles can be manipulated at all by any given voter.

The main result from [50] can be formulated as follows:

Theorem 5.9

> (Robust Version of the Gibbard-Satterthwaite Theorem) *There exists a universal constant $C > 0$ such that for every $\varepsilon > 0$ and any n the following holds: if F is a neutral social choice function for n voters and three alternatives such that the distance of F from any dictatorship and from having only two alternatives in its range is at least ε, then*
> $$\sum_{i=1}^n \mathrm{M}_i(F) \geq C\varepsilon^2.$$

The *distance* between two social choice functions F, G is defined as the fraction of inputs on which they differ, i.e., it is $\mathrm{Pr}_x[F(x) \neq G(x)]$, which is the probability that they are different under the impartial culture assumption. The distance between a social choice function F and a finite set of social choice functions is simply the minimum of the distances between F and any member of the set.

What Theorem 5.9 says is: if the total manipulation power of a voting rule on three alternatives is small, then it is either close to being a dictatorship or close to having only two alternatives in its range. Of course, the significance of this theorem also depends on how big the universal constant C is. The result could become insignificant for too small C; ideally we would like C to be as big as possible.

Theorem 5.6 implies that any non-weak-dictatorship (i.e., not depending on one coordinate only) that has at least two alternatives in its range, is manipulable. Dictatorships are obviously non-manipulable. But what is the set of all non-manipulable SCFs equal to? Given a fixed number of voters and number of alternatives, define the set

$$\text{NONMANIP} \stackrel{\text{def}}{=} \{F \text{ is an SCF depending on one coordinate only}\}$$
$$\cup \{F \text{ is monotone and takes on exactly two values}\}.$$

Here, an SCF is called *non-monotone* if for some profile a voter can change the outcome from, say, alternative a to alternative b by moving a ahead of b in her preference. Such an SCF is clearly manipulable. An SCF is *monotone* if it is not non-monotone. Note that this notion of monotonicity is weaker than the ones defined in Subsection 5.2.1. By Theorem 5.6 it is then easy to see that NONMANIP indeed equals the set of all non-manipulable SCFs. From now on, if we say "F is a non-manipulable SCF", we mean that $F \in$ NONMANIP.

We will need the following notion:

Definition 5.10

Let n be the number of voters and m the number of alternatives. A *generalized social welfare function (GSWF)* is a function

$$G : \mathcal{L}^n \to \{0,1\}^{\binom{m}{2}},$$

where $\mathcal{L}$ is the set of total orders on the set of all alternatives. That is, given the preference orders of the voters (i.e., a profile), G outputs the preferences of the society amongst each pair of alternatives.

This means that G consists of a function $G^{a,b} : \mathcal{L}^n \to \{0,1\}$, for each pair of distinct alternatives a, b. For example, $G^{a,b}(x_1, \ldots, x_n) = 1$ means that under profile $(x_1, \ldots, x_n)$ on the societal level a is preferred to b.

Of course, it is possible that a GSWF will fail to output a total order: the ordering it induces could be inconsistent.[2]

The IIA condition, neutrality, dictatoriality, etc., can be readily defined as before. For completeness, we give the formal definitions. Let G be a GSWF for three alternatives a, b, c.

- We say that G *satisfies the independence of irrelevant alternatives (IIA) condition* if for all alternatives a, b the function $G^{a,b}$ depends only on $x^{a,b} \in \{0,1\}^n$. Here $x^{a,b}$ is the vector whose i-th coordinate is 1 if candidate i prefers $a > b$, and 0 if he prefers $b > a$. In other words, a GSWF satisfying IIA is given by $\binom{m}{2}$ functions Boolean-valued Boolean functions $G^{a,b} : \{0,1\}^n \to \{0,1\}$ (one for each pair of alternatives a, b) where a 1 in the i-th bit of the input corresponds to candidate i preferring a to b and a 0 to candidate i preferring b to a.

[2] In the terminology of Subsection 4.1.1, we would say that its output is irrational or has a cycle, or also that there is no Condorcet winner.

Note that the condition that G satisfy IIA in the terminology of Subsection 4.1.1 precisely means that we are doing a Condorcet election, say based on (f, g, h) (here, f is $G^{a,b}$, g is $G^{b,c}$, and h is $G^{c,a}$).

- A GSWF is called *neutral* if all alternatives are treated symmetrically. A neutral GSWF for three alternatives that satisfies IIA corresponds exactly with the condition that we are doing a Condorcet election based on (f, g, h), say, for which we have $f = g = h$ (see Definition 4.1).

- G is called a *dictator* if there is a voter i such that the output of G is completely determined by x_i, for any input profile $(x_1, \ldots, x_n)$; G is called an *anti-dictator* if there is a voter i such that the output of G is completely determined by $-x_i$ (which is the reverse order of x_i), for any input profile $(x_1, \ldots, x_n)$.

Definition 5.11

Let G be a GSWF on three alternatives. We put

$$\mathrm{NT}(G) \stackrel{\text{def}}{=} \Pr_{\mathrm{ICA}}[G \text{ gives a non-transitive outcome}].$$

Recall from Subsection 4.1.1 that "a non-transitive outcome", "a non-rational outcome", "a cycle", "no Condorcet winner", are all different expressions to mean the same thing.

The strengthened version of Arrow's Theorem, Theorem 4.10, can then be formulated in terms of generalized social welfare functions as follows:

Theorem 5.12

*(*Robust Version of Arrow's Theorem for GSWFs*) There is an absolute constant C such that the following holds. If G is a neutral GSWF for three alternatives satisfying IIA, then in case G is at least ε-far from dictatorships and anti-dictatorships, we have* $\mathrm{NT}(G) \geq C\varepsilon$.

Proof. As explained before, G being a GSWF that satisfies IIA and neutrality means that we are doing a 3-candidate Condorcet election based on the Boolean functions f, g, h, say (see Definition 4.1). Neutrality implies that $f = g = h$. The proof is thus in essence the same as the one from Theorem 4.10. ■

We will now sketch the proof of Theorem 5.9. Although the idea behind the proof is quite natural and elegant, some parts of the proof are a bit tedious. Hence we will skip some details along the way. It will be our primary aim to convey the *idea* of the proof to the reader.

On the highest level, the proof in essence consists of a reduction to the quantitative version of Arrow's Theorem (Theorem 5.12). The question, of course, is how to achieve this reduction. For one thing, the quantitative version of Arrow's Theorem uses the formalism of social *welfare* functions (i.e., the outcome of the voting mechanism is a full ranking of all alternatives), whereas the Gibbard-Satterthwaite Theorem deals with social *choice* functions (i.e., the outcome is one alternative).

On the second-highest level, this goes in three steps. Given is an SCF F with low total manipulation power. Our aim is to show that F is either close to a dictatorship or close to having only two alternatives in its range.

STEP 1. We show that an SCF F with low total manipulation power has *weak dependence on irrelevant alternatives.* Below we will make this notion more precise, but intuitively the idea is: given any two alternatives a, b, how much does the irrelevant alternative c affect the question whether a or b is elected by F? If this is small, we say that F has weak dependence of irrelevant alternatives.

STEP 2. Given is an SCF F with weak dependence on irrelevant alternatives, obtained from the first step. It is then shown how F can be used to construct a GSWF G which satisfies IIA and is "almost transitive". (This construction will require that F be far from having only two alternatives in its range. Clearly this assumption is without loss of generality, since otherwise we are done immediately.) Additionally, this "translation" of F into G preserves the relevant distances, in the sense that the distance of G from any dictatorship, any anti-dictatorship and from always ranking one alternative at the top/bottom is about the same as the distance of F from any dictatorship, any anti-dictatorship and from having only two alternatives in its range, respectively.

STEP 3. This is the easy step, given that we already established the quantitative version of Arrow's Theorem: indeed, in the previous step we obtained a GSWF G which satisfies IIA and is "almost transitive", so by Theorem 5.12 G has to be close to a dictator or an anti-dictator. Thus, by the stipulation from Step 2, F too has to be close to a dictator or an anti-dictator. The latter possibility, however, cannot actually happen: an SCF close to an anti-dictator is clearly highly manipulable by the anti-dictator, so its total manipulation power is big. We conclude that F is close to a dictator, finishing the sketch of the proof.

Next, we go through all three steps in more detail. A reader who is not so interested in the technical details can consider skipping to Subsection 5.3.2.

Step 1. We show that an SCF F with low total manipulation power has weak dependence of irrelevant alternatives. Throughout, F is an SCF on three alternatives.

If $x \in \mathcal{L}^n$ is a profile and a, b are alternatives, then $x^{a,b} \in \{0,1\}^n$ is defined by $x_i = 1$ if $a > b$, and $x_i = 0$ if $b > a$.

Definition 5.13

Let F be an SCF on three alternatives, and let a, b be two alternatives. The *dependence of the choice between a and b on the third (irrelevant) alternative c* is

$$\mathrm{D}^{a,b}(F) \stackrel{\text{def}}{=} \Pr[F(x) = a, F(x') = b],$$

where $x, x' \in \mathcal{L}^n$ are chosen uniformly at random but subject to the restriction $x^{a,b} = (x')^{a,b}$.

Kalai et al. [50] write $\mathrm{M}^{a,b}(F)$ instead of $\mathrm{D}^{a,b}(F)$. Imagine $\mathrm{D}^{a,b}(F)$ is high. This means that there is a significant dependence of the election rule F on where c is put in the voters' preferences. Hence $\mathrm{D}^{a,b}(F)$ measures how much the irrelevant alternative c affects the question whether F elects a or b.

The following lemma is crucial.

Lemma 5.14

Let F be an SCF on three alternatives. Then for any alternatives a, b it holds that

$$\mathrm{D}^{a,b}(F) \leq 6 \sum_{i=1}^{n} \mathrm{M}_i(F).$$

Crucially, this lemma implies that if the total manipulation power is small, then there is weak dependence of irrelevant alternatives, which is the essence of Step 1.

As it turns out, proving this lemma is the most tedious part of the whole proof. Furthermore, in some ways it can also be seen as the most crucial part of the entire argument. For example, the other parts of the proof are relatively straightforward (although we will not show it) generalizable to the case in which there are more than three voters. At the time, Kalai et al. [50] were, however, not able to generalize this first lemma to the case of more than three voters.

How does one go about proving Lemma 5.14? We will not go into all the details. At the risk of being vague, we briefly sketch the main ideas and tools that are used. The lemma relates $\mathrm{D}^{a,b}(F)$ to $\sum_{i=1}^{n} \mathrm{M}_i(F)$. The idea is to search for a combinatorial structure that can be connected to both.

Let $z^{a,b} \in \{0,1\}^n$, where a, b are any alternatives. (This is just notation; z need not necessarily be a profile here.) We think of $z^{a,b}$ as fixing the positions of all a-versus-b preferences, for all voters. In order to completely specify a profile x for which $x^{a,b} = z^{a,b}$, it suffices to determine for each voter whether the third alternative c is ranked "before", "in between", or "after" a and b. These three options are symbolically represented as $0, 1, 2$, respectively. Thus, for any fixed $z^{a,b} \in \{0,1\}^n$, the sets

$$A(z^{a,b}) \stackrel{\text{def}}{=} \{x \mid x^{a,b} = z^{a,b}, F(x) = a\}, \quad B(z^{a,b}) \stackrel{\text{def}}{=} \{x \mid x^{a,b} = z^{a,b}, F(x) = b\}$$

can be considered subsets of $\{0,1,2\}^n$.

Since "x, x' are random profiles subject to $x^{a,b} = (x')^{a,b}$" means precisely "first, let p be a random profile, subsequently let x, x' be random subject to $x^{a,b} = p^{a,b} = (x')^{a,b}$", we have

$$\mathrm{D}^{a,b}(F) = \Pr[F(x) = a, F(x') = b] = \mathrm{E}_{p\in\mathcal{L}^n}\left[\frac{|A(p^{a,b})|}{3^n}\frac{|B(p^{a,b})|}{3^n}\right]. \tag{5.2}$$

Next, we want to relate $\mathrm{M}_i(F)$ to $A(x^{a,b})$ and $B(x^{a,b})$. To do that, Kalai et al. use the notion of *upper edge border* of any given $S \subseteq \{0,1,2\}^n$. Intuitively, $\delta_i S$ is the set of all points which lie on the border of S and for which one "step" in direction i is still possible. That is,

$$\delta_i S \stackrel{\text{def}}{=} \{(v_{-i}, v_i, v_i') \mid (v_{-i}, v_i) \in S, (v_{-i}, v_i') \notin S, v_i < v_i'\} \subseteq \{0,1,2\}^{n+1}.$$

We also put $\delta S \stackrel{\text{def}}{=} \bigcup_{i=1}^n \delta_i S$.

There is an obvious link with manipulability. For example, for x a profile and a, b alternatives, what is $\delta_i(A(x^{a,b}))$? Suppose $(v_{-i}, v_i, v_i') \in \delta_i(A(x^{a,b}))$. Let x_i' be the profile which is the same as x_i on a-versus-b, and whose preferences regarding c are as given by v_i'. Let x' be the same as x but with x_i replaced with x_i'.

We claim that either x_i' is a manipulation of x or x_i is a manipulation of x'. Moving from x_i to x_i' caused the output of F to change from a to $t \in \{b, c\}$; notice that x_i and x_i' are the same, except c is ranked strictly lower. If voter i has $t > a$ in x_i, then voting instead according to x_i' is a profitable manipulation for him. If, on the other hand, voter i has $a > t$ in x_i, then also $a > t$ in x_i' (since t is b or c, c moved to the right, and a-versus-b stayed the same); thus, x_i is a profitable manipulation for voter i in the profile x'.

From this we conclude that each point in $\delta_i(A(x^{a,b}))$ leads to profitable manipulation for voter i. Analyzing this probabilistically results in the following lemma:

Lemma 5.15

Let F be an SCF on three alternatives, and $1 \le i \le n$. Then for any alternatives a, b it holds that

$$\mathrm{M}_i(F) \ge \frac{1}{6}\,3^{-n}\ \mathrm{E}_{x\in\mathcal{L}^n}\left[|\delta_i(A(x^{a,b}))| + |\delta_i(B(x^{a,b}))|\right].$$

Proof. Fix a profile x. Because of the above reasoning, we must have

$$\Pr_{x_i'}\left[x_i' \text{ is a profitable manipulation of } x\right] \ge \frac{|\delta_i(A(x^{a,b}))|}{|\{0,1,2\}^{n+1}|} = \frac{|\delta_i(A(x^{a,b}))|}{3^{n+1}}.$$

Therefore, $\mathrm{M}_i(F) \ge \frac{1}{3}\,3^{-n}\ \mathrm{E}_x[|\delta_i(A(x^{a,b}))|]$. But the same holds of the part involving B, so by summing the conclusion follows. ■

Taking the sum, we obtain

$$\sum_{i=1}^{n} \mathrm{M}_i(F) \geq \frac{1}{6}\, 3^{-n}\ \mathrm{E}_{x\in\mathcal{L}^n}\left[|\delta(A(x^{a,b}))| + |\delta(B(x^{a,b}))|\right].$$

Keeping in mind Equation (5.2), Lemma 5.14 then follows by the next lemma, the proof of which we will not give:

Lemma 5.16

Let $A, B \subseteq \{0,1,2\}^n$ be disjoint. Then we have

$$|\delta(A)| + |\delta(B)| \geq 3^{-n}|A|\,|B|.$$

We refer to Proposition 3.6 in [50] for a proof, although it is not particularly hard. The proof is a consequence of a famous correlation inequality between monotone functions on the discrete cube, the *Harris-Kleitman Lemma* [62], or its generalization, the *FKG Inequality* [48]; see also [69].

This ends Step 1. We managed to show that an SCF F with low total manipulation power has weak dependence of irrelevant alternatives.

Step 2. Given an SCF F with weak dependence on irrelevant alternatives, we show how F can be used to construct a GSWF G which satisfies IIA and is "almost transitive".

We let

$$\mathrm{TR}_3 \stackrel{\text{def}}{=} \{G \text{ a GSWF} \mid G \text{ satisfies IIA and } G \text{ is always transitive}\},$$

where the 3 refers to the fact that we are only considering GSWFs on three alternatives. What is this set equal to? Mossel [85] gave a full characterization: besides the dictators and anti-dictators, it contains only GSWFs which always output a fixed alternative at the top, or those that always output a fixed alternative at the bottom. In any case, it is clear that any GSWF in TR_3 is objectionable from the point of view of social choice: we want any reasonable voting rule to be as far as possible from any member of TR_3.

If G is a GSWF, the *distance* between G and a finite set S of GSWFs is

$$\mathrm{dist}(G,S) \stackrel{\text{def}}{=} \min\{\mathrm{dist}(G,G') \mid G' \in S\},$$

i.e., it is the minimal number of output values of G that should be changed in order to make G a member of S.

The following lemma quantifies the intuitions we described in the informal description of Step 2:

LEMMA 5.17

Let $\varepsilon_1, \varepsilon_2 > 0$. Suppose F is an SCF for three alternatives which satisfies the following three conditions:

(1) $\mathrm{D}^{a,b}(F) \leq \varepsilon_1$ *for all pairs a, b.*

(2) F is at least ε_2-far from dictators and anti-dictators.

(3) $\mathrm{Pr}_{x \in \mathcal{L}^n}[F(x) = a] \geq \varepsilon_2$, *for each alternative a.*

Then there exists a GSWF G on three alternatives for which:

(1') G satisfies IIA.

(2') $\mathrm{dist}(G, \mathrm{TR}_3) \geq \varepsilon_2 - 3\sqrt{\varepsilon_1}$.

(3') $\mathrm{NT}(G) \leq 3\sqrt{\varepsilon_1}$.

The construction of G itself is straightforward: given F, define a GSWF G by

$$G^{a,b}(x) = \begin{cases} 1 & \text{if } \mathrm{Pr}_{x'}[F(x') = a \mid (x')^{a,b} = x^{a,b}] > \mathrm{Pr}_{x'}[F(x') = b \mid (x')^{a,b} = x^{a,b}] \\ 0 & \text{if } \mathrm{Pr}_{x'}[F(x') = a \mid (x')^{a,b} = x^{a,b}] < \mathrm{Pr}_{x'}[F(x') = b \mid (x')^{a,b} = x^{a,b}] \end{cases},$$

for each profile $x \in \mathcal{L}^3$ and alternatives a, b. If the probabilities happen to be equal, then some arbitrary but fixed voter takes the decision.

By the very construction, G satisfies IIA, so (1') is fulfilled.

We now show that (3') is satisfied. We need the following notion. A profile $x \in \mathcal{L}^n$ is called a *minority preference on the alternative* a, b if $F(x) = a$ while $G^{a,b}(x) = 0$, or if $F(x) = b$ while $G^{a,b}(x) = 1$. Intuitively, this means that x belongs to a minority, since, e.g., $G^{a,b}(x) = 0$ by definition means that on *most* profiles with the same a-versus-b rankings, F elects b. We call a profile x a *minority preference* if there are alternatives (a, b) for which x is a minority preference. For fixed alternatives a, b, we put

$$\mathrm{N}^{a,b}(F) \stackrel{\text{def}}{=} \Pr_{x \in \mathcal{L}^n}[x \text{ is a minority preference on } a, b].$$

If there are lots of minority preferences for a, b, then deciding on whether a or b is elected under F depends significantly on the irrelevant third alternative c:

LEMMA 5.18

Let F be an SCF. For any pair of alternatives a, b, we have

$$\mathrm{D}^{a,b}(F) \geq (\mathrm{N}^{a,b}(F))^2.$$

Proof. For any $t \in \{0, 1\}^n$ (which we imagine to be determining the a-versus-b rankings), we define

$$p_a(t) \stackrel{\text{def}}{=} \mathrm{Pr}[F(z) = a], \quad p_b(t) \stackrel{\text{def}}{=} \mathrm{Pr}[F(z) = b],$$

where $z \in \mathcal{L}_3$ is randomly distributed but subject to $z^{a,b} = t$.

Since "x, x' are random profiles subject to $x^{a,b} = (x')^{a,b}$" means precisely "first, let p be a random profile, subsequently let x, x' be random subject to $x^{a,b} = p^{a,b} = (x')^{a,b}$", it is clear that $\mathrm{D}^{a,b}(F) = \mathrm{E}_{t\in\{0,1\}^n}[p_a(t)\, p_b(t)]$. In a similar fasion, $\mathrm{N}^{a,b}(F) = \mathrm{E}_{t\in\{0,1\}^n}[\min\{p_a(t), p_b(t)\}]$. Then, we get

$$\begin{aligned}\mathrm{D}^{a,b}(F) &= \mathrm{E}_t[p_a(t)\, p_b(t)] \geq \mathrm{E}_t[(\min\{p_a(t), p_b(t)\})^2] \\ &\geq \mathrm{E}_t[\min\{p_a(t), p_b(t)\}]^2 = \mathrm{N}^{a,b}(F)^2,\end{aligned}$$

where we used the Cauchy-Schwarz inequality in the last inequality. ■

Given a GSWF G and profile x, we call an alternative a a *generalized Condorcet winner (GCW) at profile x* if for any alternative $b \neq a$, we have $G^{a,b}(x) = 1$. Clearly "G has no GCW at x" is equivalent with "G gives a non-transitive outcome on x".

Note that if G does not have a GCW at profile x, then x must be a minority preference of F. Indeed, were x not a minority preference of F, then for all a, b, either $G^{a,b}(x) = 1$ and $F(x) = a$, or $G^{a,b}(x) = 0$ and $F(x) = b$; letting $a = F(x)$, we obtain that $G^{a,b}(x) = 1$ for all alternatives b, so a is a GCW. Therefore, we have

$$\begin{aligned}\mathrm{NT}(G) &= \Pr_x[G \text{ gives a non-transitive outcome on } x] \\ &= \Pr_x[G \text{ does not have a GCW at } x] \\ &\leq \Pr_x[x \text{ is a minority preference of } F].\end{aligned}$$

Notice that by Lemma 5.18, this is at most

$$\sum_{a,b} \mathrm{N}^{a,b}(F) \leq \sum_{a,b} \sqrt{\mathrm{D}^{a,b}(F)} \leq 3\sqrt{\varepsilon_1}.$$

Thus, (3') is met.

Finally, we show that condition (2') is satisfied. Put $\varepsilon = \mathrm{dist}(G, \mathrm{TR}_3)$. We need to show that $\varepsilon \geq \varepsilon_2 - 3\sqrt{\varepsilon_1}$. Let $H \in \mathrm{TR}_3$ achieve the minimum, i.e., G can be converted into H by changing just an ε-fraction of its values.

Case 1: H ranks one fixed alternative, say a, at the top. In that case, as G and H are ε-close, we must have $\Pr[a \text{ is a GCW}] \geq 1 - \varepsilon$. If x is not a minority preference and a is a GCW of G at x, then evidently $F(x) = a$. Therefore,

$$\Pr[F(x) = a] \geq (1 - \varepsilon) - \Pr[x \text{ is a minority preference}] \geq 1 - \varepsilon - 3\sqrt{\varepsilon_1}.$$

However, (3) implies in particular that $\Pr[F(x) = a] \leq 1 - \Pr[F(x) = b] \leq 1 - \varepsilon_2$, from which $\varepsilon \geq \varepsilon_2 - 3\sqrt{\varepsilon_1}$ follows.

Case 2: H ranks one fixed alternative, say a, at the bottom. This follows similarly as the previous case, instead using the concept of a generalized Condorcet *loser*.

Case 3: H is dictatorship and i is the dictator. If x is a profile, $x_{\text{top}}(i)$ is the top alternative in the preference order of voter i. Then

$$\Pr[x_{\text{top}}(i) \text{ is a GCW at } x] \geq 1 - \text{dist}(G, H) = 1 - \varepsilon,$$

so similarly as before we have

$$\Pr[F(x) = x_{\text{top}}(i)] \geq (1 - \varepsilon) - \Pr[x \text{ is a minority preference}] \geq 1 - \varepsilon - 3\sqrt{\varepsilon_1}.$$

Were $\varepsilon_2 > \varepsilon + 3\sqrt{\varepsilon_1}$, then F would be ε_2-close to a dictatorship, which contradicts condition (2). Thus, $\varepsilon_2 \leq \varepsilon + 3\sqrt{\varepsilon_1}$.

Case 4: H is an anti-dictatorship and i is the anti-dictator. This case is similar to the previous one.

This finishes the proof of Lemma 5.17, and thus also Step 2: we have shown that, given an SCF with weak dependence on irrelevant alternatives, it can be used to construct a GSWF which satisfies IIA and is "almost transitive".

Step 3. Finally, we prove Theorem 5.9. By contradiction, suppose that for each universal constant $C > 0$ there is an $\varepsilon > 0$, an n, a neutral social choice function F for n voters and three alternatives, such that the distance of F from a dictatorship and from having only two alternatives in its range is at least ε, and $\sum_{i=1}^{n} \mathrm{M}_i(F) < C\varepsilon^2$.

Concretely, suppose $\sum_{i=1}^{n} \mathrm{M}_i(F) < C''\varepsilon^2$ where C'' is some as of yet undetermined absolute constant. From Step 1 (Lemma 5.14) we get $\mathrm{D}^{a,b}(F) < 6C''\varepsilon^2$. Letting $\varepsilon_1 = 6C''\varepsilon^2$ and $\varepsilon_2 = \varepsilon$, from Step 2 (Lemma 5.17) we obtain a GSWF G such that

(1') G satisfies IIA.

(2') $\text{dist}(G, \mathrm{TR}_3) \geq \varepsilon_2 - 3\sqrt{\varepsilon_1} = \varepsilon - 3\varepsilon\sqrt{6C''}$.

(3') $\mathrm{NT}(G) \leq 3\sqrt{\varepsilon_1} = 3\varepsilon\sqrt{6C''}$.

As F is neutral, by the construction of G in Step 2, G is neutral as well; this together with (1') implies that we may apply the quantitative version of Arrow's Theorem (Theorem 5.12). However, by taking C'' to be small enough it is clear that we can get $\text{dist}(G, \mathrm{TR}_3)$ to be at least $\approx \varepsilon$. By contrast, $\mathrm{NT}(G)$, the probability under ICA that the Condorcet paradox arises, will be arbitrarily close to 0 for C'' small enough. This contradicts Theorem 5.12: the quantitative version of Arrow's Theorem says that being ε-far from dictators and anti-dictators means that the Condorcet paradox will appear at the rate of at least $\Omega(\varepsilon)$. The proof of Theorem 5.9 is finished. ■

This finishes the proof of Step 3. The argument is complete.

5.3.2 The State of the Art: the Mossel-Rácz Approach

Another way of stating Theorem 5.9 is as follows: there is a universal constant C such that, given a neutral social choice function for n voters and three alternatives which is at least ε-far from any member of NONMANIP (in particular, from any dictatorship and from any SCF having only two alternatives in its range), then

$$\Pr_{\text{ICA},i}[\text{a random manipulation by voter } i \text{ is profitable}] \geq C\frac{\varepsilon^2}{n},$$

where i is uniformly random among all voters. That is, a *random manipulation by a random voter i* will succeed with nonnegligible[3] probability.

Two apparent limitations regarding Kalai et al.'s version of the quantified Gibbard-Satterthwaite Theorem immediately come to mind: first, the result (Theorem 5.9) holds only for three number of voters, and second, it requires the assumption of neutrality. It is for many reasons important to be able to drop neutrality; below we will briefly discuss some of those reasons. Supplementary research was therefore needed.

Driven by their result, Kalai et al. [50] also conjectured the following:

CONJECTURE 5.19

> *If a social choice function for n voters and m alternatives is ε-far from the family of non-manipulable functions* NONMANIP*, then the probability of a profile being manipulable is bounded from below by a polynomial in $\frac{1}{n}$, $\frac{1}{m}$, and ε. Furthermore, a random manipulation by a random voter will succeed with nonnegligible probability.*

The above statement is at the current time not a conjecture anymore, as we will see in the following overview.

Overview of Developments. Further research by Isaksson, Kindler, and Mossel [65] generalized the quantitative Gibbard-Satterthwaite Theorem to any number of alternatives, while still assuming neutrality. Additionally, they showed that a random manipulation which replaces four adjacent alternatives in the preference order of the manipulating voter by a random permutation of them succeeds with nonnegligible probability. Their paper is titled *The Geometry of Manipulation - A Quantitative Proof of the Gibbard-Satterthwaite Theorem*. As the title suggest, besides combinatorics the employed proof techniques are mostly geometrical in nature. In particular, they do not involve discrete harmonic analysis (Fourier analysis), which is the reason why we will not go into the details of their proof. We just mention that a variant of the canonical path method to prove isoperimetric bounds is applied.

[3] In computer science, a function $f : \mathbb{N} \to \mathbb{R}$ is called *negligible* [99] if for every positive polynomial poly$(\cdot)$ there exists an integer $N > 0$ such that for all $x > N$, it holds that $|f(x)| < \frac{1}{\text{poly}(x)}$, and *nonnegligible* if it is not negligible. The reason why the $1/\text{poly}(\cdot)$ form is used is similar to the reason why in computational complexity computational tractability is defined in terms of *polynomial* running time [99]. In the context of cryptography, for example, an attacker might have a success probability of hacking some cryptographic scheme with negligible probability; however, even if the attack is repeated a polynomial number of times, the success probability of the overall attack still continues to be negligible.

Finally, in 2012 Mossel and Rácz [88] managed to remove the assumption of neutrality, and thus turned Conjecture 5.19 in a theorem. In fact, they proved the following:

THEOREM 5.20

Let $n \geq 1$ and $m \geq 3$. Given a social choice function F for n voters and m alternatives that is at least ε-far from any member of NONMANIP*, the probability under ICA that a randomly picked voter obtains a profitable manipulation by randomly permuting four randomly picked adjacent alternatives in her own preference, is at least*

$$p\left(\varepsilon, \frac{1}{n}, \frac{1}{m}\right) = \frac{\varepsilon^{15}}{10^{41} n^{68} m^{167}}.$$

The proof combines ideas and techniques from both Kalai et al.'s approach as well as the paper by Isaksson, Kindler, and Mossel [65]. We will not go into the proof, as it is long and tedious. Interestingly, one of the crucial new ingredients in their proof is a *reverse* Hypercontractive Inequality leading to a new isoperimetric inequality on the discrete cube. This result is due to Mossel et al. [87]; we refer to their Theorem 3.2 for the precise formulation.

Importance of Omitting Neutrality. Neutrality might at first seem a rather innocent condition. This need not be the case. We give two motives.

- First, Moulin [89] has shown that there is a conflict between neutrality and anonymity. More specifically, he proved that there exists an SCF on n voters and m alternatives that is both anonymous and neutral if and only if m cannot be written as the sum of non-trivial divisors of n. This problem arises due to tie-breaking. Clearly, anonymity is a highly desirable property, the more so in real-life elections. Many of the common voting rules (Borda, plurality, STV, etc.) are anonymous; therefore, by Moulin's result, for some values of n and m those rules cannot be neutral too.

- Second, for computer science applications neutrality is often unnatural. Imagine, for example, a so-called "meta-search engine" (see [33] and [40]). This is a search engine based on other search engines (like Google, Yahoo, etc.): given a user request, it queries the other search engines, all of which give a ranking of the relevant web pages, and subsequently it aggregates those into a final ranking of web pages. Various restrictions are possible, making such a rule non-neutral. Think for example about language restrictions: if the user only wants to be suggested pages in Dutch, say, then clearly the voting rule cannot be neutral: all non-Dutch web pages are excluded in advance.

5.3.3 Computational Hardness vs. Manipulation: Hopeless?

Already in 1989, Bartholdi, Tovey and Trick [4] suggested to use computational complexity as a barrier against manipulation. The rationale is: if it is hard for a voter to *compute* a manipulation for a given voting rule, then finding a manipulation is practically infeasible and thus we should not worry about it. For example, the manipulation problem for the well-known voting rule called single transferable vote (STV) was proven to be NP-hard [3].

A fundamental problem with this idea is that computational complexity is a *worst-case* approach. Concretely, in the case of STV for example, in the worst case it is infeasible to find a manipulation. It does not imply anything about the more relevant question: how hard is it to manipulate *on average*? Ideally we would like manipulation to be hard for *most* instances.

The establishment of Conjecture 5.19 is pivotal in this regard: under ICA, for any social choice function which is far from the non-manipulable voting rules, a random manipulation by a single randomly chosen voter will succeed with non-negligible probability. In particular this means that a voter with black-box access to the social choice function will be able to find a manipulation efficiently (if it exists). The conclusion is:

> *Under the impartial culture assumption, manipulation is easy on average.*

Do these results end the project of the use of computational hardness against manipulation? Not necessarily. A few remarks:

- The result from Theorem 5.20 is mostly of importance from a *theoretical* perspective. After all, the number
$$\frac{\varepsilon^{15}}{10^{41}n^{68}m^{167}}$$
is very small. In practice, for example if we think about a real-life political election (say, in Belgium, which has roughly $10,000,000 \approx 10^7$ inhabitants), m might be about 10 while n might be about 10^7. In that case, the above lower bound on the probability is about $\frac{\varepsilon^{15}}{10^{684}}$, a devastatingly small number. From a practical viewpoint it is just too small.

- Are we willing to accept the impartial culture assumption? If so, then indeed computational hardness cannot prevent manipulation. However, as we have seen before in this thesis, the impartial culture assumption, while useful, is in some (many?) cases not realistic. Furthermore, in Subsection 4.3.2 we have seen empirical findings suggesting that in practice the Condorcet paradox does not arise frequently. These observations can, in some sense, be seen as a *reductio ad absurdum* of the statement "the ICA is a realistic assumption" (at least for those empirical findings): were ICA realistic, then the Condorcet paradox would arise significantly more frequently than has been observed.

- It can be expected that voting profiles have some structure. What would be interesting, is to go through empirical data to find out more about voter tendencies, and then to put them into a feasible mathematical model. This appears to be complicated. If we could prove a result suggesting that even for such distributions manipulation is easy on average, the inevitable conclusion would be that manipulation is inescapable. Given the mathematical prowess shown in [88] that was needed just to establish that manipulation on average is easy *even assuming ICA*, such a result seems out of reach for the moment.

- As Kalai et al. [50] point out, one has to interpret the above result (Theorem 5.20) carefully. It says that a random manipulation by a randomly chosen voter succeeds with non-negligible probability. This statement does not, however, preclude the possibility that for *most* of the voters manipulation cannot be found efficiently (and for a polynomially small portion of the voters a manipulation can be found efficiently). That is, it is still possible that only a few voters can manipulate efficiently whereas most voters cannot.

The conclusion is: although the above results appear to indicate the existence of a burden for the hardness-against-manipulation agenda, not all hope is lost. In the words of Faliszewski and Procaccia [43]: "...the final word regarding the (non-)existence of voting rules that are usually hard to manipulate is yet to be said".

Chapter 6

A Simple Model for Biases in Sequential Binary Voting

Inspired by Kahneman's *Thinking, Fast and Slow* [67, 35], in this short last chapter we come up with a simple, new, model to simulate the various biases that show up in small meetings in which people have to vote, one after the other, "yes" or "no" for a given proposal.

6.1 Introduction

Meetings are prevalent in the working life of plenty of individuals. To give an example, a recent anthropological study at a US university [150] shows that their professors spend as much as 17 percent of their workweek days in meetings. Remarkably, a significant amount of those who regularly attend meetings consider them to be costly, unproductive, and dissatisfying [112]. Many salient decisions, such as in courtrooms or at corporate executive boards, are taken during assemblies. Although it is clear that our society depends profoundly on decisions that resulted from meetings, work by Kahneman and Tversky [67] has shown that decisions taken by committees are liable to pervading biases.

We noted in Subsection 1.2.2 that people as a whole can make a better choice than individuals separately, a phenomenon named "the wisdom of the crowd" [129]. Condorcet's Jury Theorem formalizes this idea. The following example is given in [67]. Suppose a large number of individuals have to estimate the number of pennies in a glass jar. Some people will overestimate the number, and some will underestimate it. Assuming the people's errors are independent, in the long run the errors tend to cancel each other out, thus providing a good estimate by taking the average of all individuals' estimates. The requirement of independence is, however, a *conditio sine qua non*: as Kahneman [67] puts it, the "magic" of error reduction works well only when the observations are independent and their errors uncorrelated. The motto: decorrelate error!

The essential point is hence: by the wisdom of the crowds, meetings *could*, in principle, give rise to well-conducted decisions, but because of several obstinate biases, in practice it often turns out differently.

What are these biases, exactly? We give a brief overview of the most pertinent ones. Most of the items we mention below come from [67].

- *Exaggerated Emotional Coherence (Halo Effect).* Human beings tend to like (dislike) everything about a person, idea, or agument—including features that have not been observed—whenever just one aspect of it attracts (repels) them. Favorable first impressions therefore influence later judgments. For example, a charming speaker might "automatically" be perceived by an audience as competent too.

- *Anchoring Effect.* This bias connects well with the Halo effect; it says that, when making decisions, individuals have the tendency to rely too much on the first piece of information put forward, the so-called "anchor". Subsequent judgments are made by comparing with the anchor, even if the anchor was completely irrelevant. For example, when negotiating the price of a new car, it is advantageous for a buyer to "start low", i.e., first to offer a ridiculously low amount, just to set the anchor; afterwards, all prices will be compared with respect to that very low price. The following example from [67], in which a *random* anchor is used, illustrates the anchoring effect more blatantly. In Germany, experienced judges first read a description of a woman who had been caught shoplifting. Then, two biased dice (fixed in such a way that they give either a 3 or a 9 on each throw) were rolled; directly after that the judges were asked whether they would sentence the woman to a term in prison greater or lesser than the number, in months, shown on their die. Lastly, the judges were asked how many months imprisonment they would impose on the shoplifter. The results were astonishing: on average, those who had rolled a 9 stated they would sentence her to 8 months, whereas those who rolled a 3 stated they would sentence her to 5 months.

- *Social Conformity and the Bandwagon Effect.* The bandwagon effect, strongly related to *groupthink* and *herding*, is a form of convergent social behaviour that can be broadly defined as the alignment of the thoughts or behaviours of individuals in a group ("herd") through local interaction and without centralized coordination [104]. In other words, in case many people come to believe in something, others also join in and "jump on the bandwagon". An interesting example is the phenomenon of "likes" on Facebook and similar social media, studied by Aral et al. [90]. Briefly summarized, "likes" (i.e., upvotes) of online articles or webpages make other people "like" that article as well; on the other hand, negative reactions[1] (i.e., downvotes) do *not* prompt others to dislike the article. That is, a positive nudge can institute a bandwagon of approval, but a negative gesture fails to achieve a negative effect. These findings are particularly bad news for users of websites such as TripAdvisor, Amazon, or Yelp: in order to receive good service they would have to benefit from the "wisdom of the crowd". Instead, there seems to be a persistent bias.

[1] Note that Facebook actually does not have a "dislike" button, but other similar websites such as Reddit and YouTube do. Mark Zuckerberg commented "that's [a dislike button] not something that we think is good for the world" [98], but perhaps (part of) the real reason has something to do with the bandwagon effect. At the least, the bandwagon effect shows once more why, for Zuckerberg, having a "like" button is positive: it attracts more people, resulting in more data extrapolation and thus an increase in revenue.

6.2 A Simple Model

In this section we will study the effects and biases that occur when people vote sequentially instead of in parallel. The model seeks to capture the effect that individuals who come later will be influenced by the speakers who have cast their vote already, and the strength of this "influence" is expressed by a *sway*. Below we will explain this notion in more detail.

6.2.1 Description

Suppose we have a committee consisting of n individuals, marked by $1, 2, \ldots, n$. There is some proposal. They have to come to an agreement together, i.e., at the end of the meeting the group as a whole has to output a 1 ("proposal accepted") or a 0 ("proposal rejected"). Each individual, one after the other, has an opportunity to expand on their viewpoints. All individuals vote publicly, and they do so in order of their index, so 1 goes first, then 2, etc.. When all individuals have cast their vote, the majority rule[2] is applied: thus, if the majority of individuals accepted the proposal, then the output is 1, else the output is 0.

We can imagine that among these individuals some have more "weight" than others. For example, in a court where there are junior and senior judges, if only because of their maturity and skill, among the judges the senior ones may be regarded as carrying more influence or weight; by their very seniority, they "carry a halo", metaphorically speaking. Therefore, we assign a real number $w_i \in [0,1]$ to each individual $i \in [n]$, and call it individual i's *sway*,[3] with the additional proviso that $w_1 + w_2 + \ldots + w_{n-1} \leq 1$. We let $x_i \in \{0,1\}$ be individual i's vote. We still need to stipulate the probabilistic model we use, which we do now:

- Individual 1 votes uniformly at random.
- For any $i \in \{2, 3, \ldots, n\}$, individual i votes according to

$$x_i \stackrel{\text{def}}{=} \begin{cases} x_j & \text{with probability } w_j \ (j < i) \\ \{0,1\} \text{ uniformly at random} & \text{with probability } 1 - \sum_{j=1}^{i-1} w_j \end{cases}.$$

  That is, voter i mimics ("follows") voter j with probability w_j, for each $j < i$, and votes uniformly at random otherwise.

We write the joint probability distribution of x_1, x_2, x_3 briefly as x. When we write $\Pr_x[\cdot]$ in this context[4] it means that we take the probability over x_1, x_2, x_3 as defined above.

[2] The reason why we opt for the majority rule is clear: for a 2-candidate election the majority rule is the best choice, as, e.g., May's Theorem (Theorem 2.15) shows.

[3] We do not want to use the word *influence*, as we reserved that name for another concept; see Definition 3.1 in Subsection 3.1.1.

[4] Note that in all previous chapters of the thesis this actually meant "$x = (x_1, \ldots, x_n)$ is uniformly random".

Our model intends to capture *small* group deliberation; indeed, for bigger groups such sequential voting seems infeasible due to practical constraints. Therefore, we should focus mostly on small values of n, say $n \leq 10$.

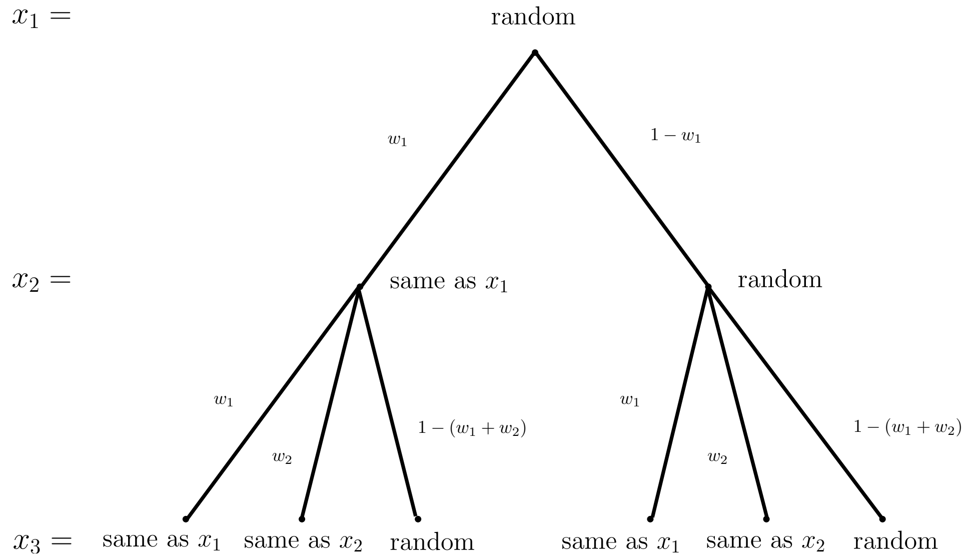

Figure 6.1: The probabilistic model for $n = 3$. The edges of the graph denote the probabilities, whereas the nodes at depth i indicate the possible values of x_i.

The tree in Figure 6.1 will be useful in order to calculate probabilities.

We are particularly interested in the probability that individual i's choice will end up winning the election:

Definition 6.1

The *force* of individual i, denoted by F_i, is the probability that the election outcome is equal to individual i's choice. That is,

$$F_i \stackrel{\text{def}}{=} \Pr_x[\text{Maj}_n(x_1, \ldots, x_n) = x_i].$$

In general, for a given n the force F_i will be a function of $w_1, \ldots, w_{n-1}$. What is the relationship between force and influence? It is not hard to see that, when $w_1 = w_2 = \ldots = w_{n-1} = 0$, then we have

$$F_i = \frac{1}{2} + \frac{\text{Inf}_i[f]}{2}.$$

In the next subsection we analyze the smallest interesting case: $n = 3$.

6.2.2 Analysis of the Case $n = 3$

We will use the following lemma repeatedly:

LEMMA 6.2

(1) $\Pr_x[x_1 = x_2] = \frac{1}{2}w_1 + \frac{1}{2}$.

(2) $\Pr_x[x_2 = x_3] = \frac{1}{2}w_1^2 + \frac{1}{2}w_2 + \frac{1}{2}$.

(3) $\Pr_x[x_1 = x_3] = \frac{1}{2}w_1 w_2 + \frac{1}{2}w_1 + \frac{1}{2}$.

(4) $\Pr_x[x_1 = x_2 = x_3] = \frac{1}{4}(w_1 + 1)(w_1 + w_2 + 1)$.

Proof. (1) We have $\Pr_x[x_2 = x_1] = w_1 1 + \frac{1}{2}(1 - w_1) = \frac{1}{2}w_1 + \frac{1}{2}$.
(2) Using part (1), in a similar fashion it holds that

$$\Pr_x[x_3 = x_2] = w_1 \Pr_x[x_1 = x_2] + w_2 1 + \frac{1}{2}(1 - (w_1 + w_2)) = \frac{1}{2}w_1^2 + \frac{1}{2}w_2 + \frac{1}{2}.$$

(3) In the same way, we get

$$\Pr_x[x_3 = x_1] = w_1 1 + w_2 \Pr_x[x_2 = x_1] + \frac{1}{2}(1 - (w_1 + w_2)) = \frac{1}{2}w_1 w_2 + \frac{1}{2}w_1 + \frac{1}{2}.$$

(4) Looking at Figure 6.1, we see that

$$\begin{aligned}\Pr_x[x_1 = x_2 = x_3] &= w_1 w_1 + w_1 w_2 + \frac{1}{2}w_1(1 - (w_1 + w_2)) \\ &\quad + \frac{1}{2}\left((1 - w_1)w_1 + (1 - w_1)w_2 + \frac{1}{2}(1 - w_1)(1 - (w_1 + w_2))\right).\end{aligned}$$

A calculation shows that this expression equals $\frac{1}{4}(w_1 + 1)(w_1 + w_2 + 1)$. ■

The introduction of the sways w_1, w_2 clearly induce correlations into the model: x_2 and x_3 are not independent anymore from x_1, but instead match the first individual's vote some percentage of the time (depending on the first individual's sway w_1).

A Tendency to Agree. In addition to the biases mentioned in Section 6.1, here is another, related, bias [35]. Even leaving aside a bias such as the halo effect, some members of the meeting might be more anxious about their outward appearance (for example, to make sure one looks competent, well-liked by the others, and with one's reputation intact) than about making correct decisions. Such a bias differs from, e.g., the Halo effect in that here the *aim* is changed; in a sense, for a person affected by this bias, the situation at hand is not about voting anymore, but more about self-preservation. This can lead participants of a meeting to avoid disagreement.[5] However, when members of a committee have incentives to agree with one another, they tend to overweight public information and discard private insights; this can, in turn,

[5]Note that although this effect might seem somewhat similar to the bandwagon effect, they are actually quite different.

generate a *status quo bias* [58].[6] Also, a bias towards "the obvious" is possible. For example, consider the scenario in which some professors are evaluating candidates for a PhD position. The professors might then only look at the self-evident, obvious, features that an apt candidate should possess (the ones about which all professors have common knowledge that they are objective, important, traits), whereas there could in fact be important issues which are more subjective, less self-evident, but nevertheless of great importance.

In any case, the probability that all three individuals agree in this model, $\Pr_x[x_1 = x_2 = x_3]$, is of concern. What could be interesting, is to compare this quantity (which involves sways and thus correlations) with the scenario in which the three voters take decisions completely independently: the impartial culture assumption.

Therefore, we define an *agreement bias function* B by

$$B(w_1, w_2) \stackrel{\text{def}}{=} \left| \Pr_{\text{ICA}}[x_1 = x_2 = x_3] - \Pr_x[x_1 = x_2 = x_3] \right|.$$

Of course, $\Pr_{\text{ICA}}[x_1 = x_2 = x_3] = \frac{1}{4}$, so using item (4) from Lemma 6.2 we find

$$\begin{aligned} B(w_1, w_2) &= \frac{1}{4}(w_1 + 1)(w_1 + w_2 + 1) - \frac{1}{4} \\ &= \frac{1}{4}(w_1^2 + w_1 w_2 + 2w_1 + w_2). \end{aligned}$$

This function seems somewhat complicated at first, although it is actually quite simple. The idea is to look at the contour lines. A drawing clarifies things. From the graph we deduce the following: in order to keep the bias small, it is foremost important that w_1 be small. Clearly, values for which $w_1 \leq w_2$ (corresponding to the left triangle of the dashed area) are generally better—in the sense that they keep the bias low—than values for which $w_1 \geq w_2$ (corresponding to the right triangle of the dashed area).

Calculating the Forces. Using Lemma 6.2, we can calculate all forces:

Theorem 6.3

We have

(1) $F_1 = \frac{1}{4}w_1 w_2 - \frac{1}{4}w_1^2 + \frac{1}{2}w_1 - \frac{1}{4}w_2 + \frac{3}{4}$;

(2) $F_2 = -\frac{1}{4}w_1 w_2 + \frac{1}{4}w_1^2 + \frac{1}{4}w_2 + \frac{3}{4}$;

(3) $F_3 = \frac{1}{4}w_1 w_2 + \frac{1}{4}w_1^2 + \frac{1}{4}w_2 + \frac{3}{4}$.

Proof. (1) Note that

$$F_1 = \Pr_x[\text{Maj}_3(x_1, x_2, x_3) = x_1] = 1 - \Pr_x[\text{Maj}_3(x_1, x_2, x_3) \neq x_1]$$

[6] The *status quo bias* indicates a preference for the current state of affairs. The status quo is taken as a reference point and any deviation from that baseline is perceived as a loss [37].

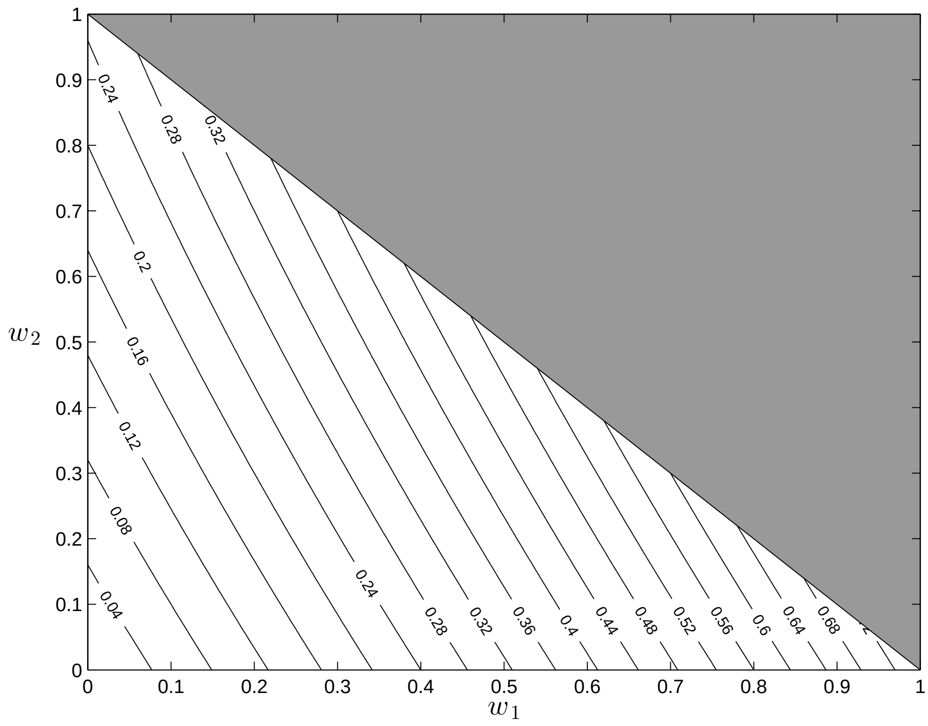

Figure 6.2: Contour lines of the function B. The gray area consists of points for which $w_1 + w_2 > 1$ and is thus irrelevant. Made using MATLAB.

$$= 1 - \Pr_x[x_2 = x_3 \;\&\; x_1 \neq x_2].$$

Now, $\Pr_x[x_2 = x_3 \;\&\; x_1 \neq x_2] = \Pr_x[x_2 = x_3] - \Pr_x[x_1 = x_2 = x_3]$, so Lemma 6.2 implies that

$$F_1 = 1 - \left(\frac{1}{2}w_1^2 + \frac{1}{2}w_2 + \frac{1}{2} - \frac{1}{4}(w_1+1)(w_1+w_2+1)\right).$$

A calculation shows that this equals $\frac{1}{4}w_1w_2 - \frac{1}{4}w_1^2 + \frac{1}{2}w_1 - \frac{1}{4}w_2 + \frac{3}{4}$.
The proofs of (2) and (3) are very similar, again using Lemma 6.2. ■

Below, figures depicting the contour lines of the functions F_1, F_2, and F_3 are included.

On Figure 6.3 we see that, whenever $w_1 \geq 0.55$, the probability that the meeting outcome will be in correspondence with the first individual's choice, is *at least* 90%. Of course, we have to keep in mind here: *relative to what are we comparing?* It seems most natural to compare with the impartial culture assumption. In that culture, the probability that any fixed individual's choice corresponds with the election outcome, is exactly 75%. This means that, given that the first person's sway is somewhat significant, the probability that his choice will correspond with the meeting outcome will be big.

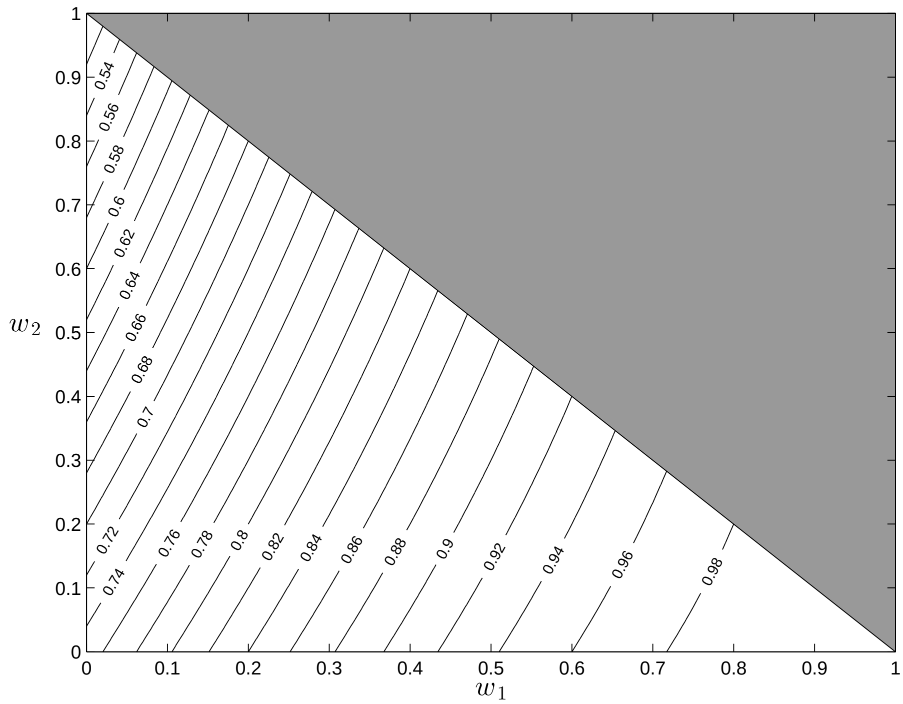

Figure 6.3: Contour lines of the function F_1. Made using MATLAB.

What is essential to the biases that we have been describing in this chapter, is that each of them in some way or the other involves individuals *adapting to*, or *correlating with*, other individuals' opinions. In this way we can think of biases as "deviating from the norm".[7] Similarly as in the previous paragraph, it makes sense to model the "norm" as *independence*; that is, we use the impartial culture assumption again. Again, the reason why this is justifiable is that the ideal scenario (i.e., "norm") we have in mind, is that all individuals act independently of one another. Indeed, in such a scenario the wisdom of the crowd works well.

Similarly as we did in the previous paragraph, we want to define a bias function. In this case, we define

$$B(w_1, w_2) \stackrel{\text{def}}{=} \max_{i \in \{1,2,3\}} \left| \Pr_x[\mathrm{Maj}(x_1, x_2, x_3) = x_i] - \Pr_{\mathrm{ICA}}[\mathrm{Maj}(x_1, x_2, x_3) = x_i] \right|.$$

The idea is that the deviation from the impartial culture should not be too big: indeed, by the above reasoning, deviating greatly from impartiality (and thus independence) indicates a bias.[8] Indeed, if $B(w_1, w_2)$ is significant, then it means that there is a voter i for which $\Pr_x[\mathrm{Maj}(x_1, x_2, x_3) = x_i]$ and $\Pr_{\mathrm{ICA}}[\mathrm{Maj}(x_1, x_2, x_3) = x_i]$ differ considerably. Intuitively, this signifies

[7] Perhaps better said: we *model* (define) biases as such.

[8] In fact, one could even say that his holds true *by the very definition* that we have given to the notion of *bias* in our model.

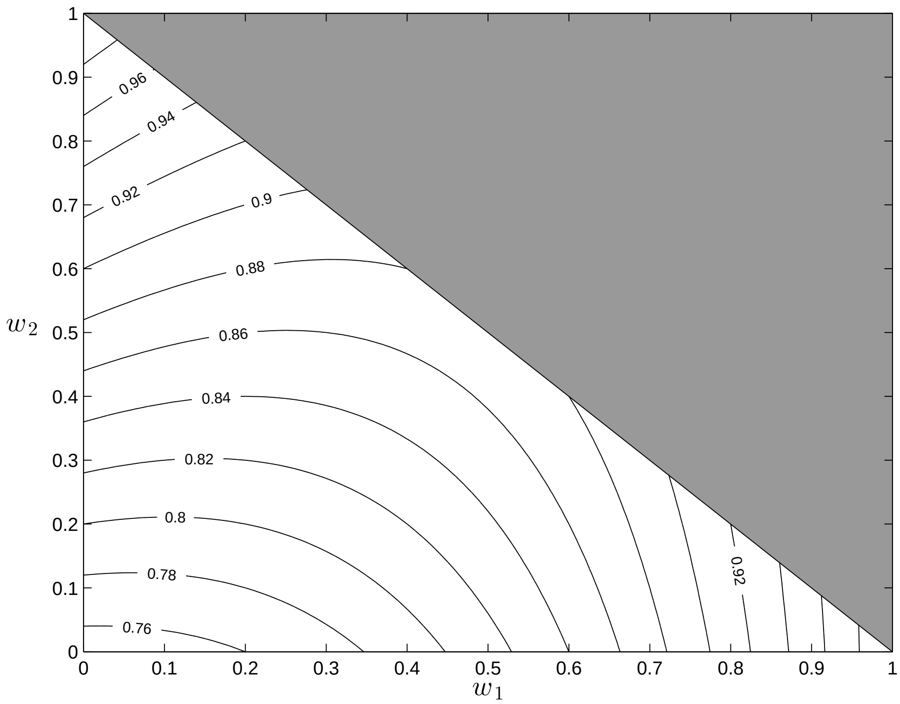

Figure 6.4: Contour lines of the function F_2. Made using MATLAB.

that there is a substantial correlation, and hence the presence of a bias. Note that

$$B(w_1, w_2) = \max_{i \in \{1,2,3\}} \left(\Pr_x[\mathrm{Maj}(x_1, x_2, x_3) = x_i] - \frac{3}{4} \right).$$

The bias being small means, say, that $B(w_1, w_2) \leq \varepsilon$, where $\varepsilon > 0$ is some small but fixed number. Equivalently,

$$F_i = \Pr_x[\mathrm{Maj}(x_1, x_2, x_3) = x_i] \leq \frac{3}{4} + \varepsilon$$

for each $i \in \{1, 2, 3\}$, i.e., *none of the forces should be much greater than* 75%.

What does this concretely mean? Let us pick $\varepsilon = 0.05$, so that $\frac{3}{4} + \varepsilon = 0.8$. Looking at the contour lines in Figures 6.4 and 6.5, we deduce that w_2 should be small,[9] say at most 0.2, while w_1 should definitely be smaller than 0.4. In fact, the area in the (w_1, w_2)-plane in which the values of w_1, w_2 lie that are "allowed" (meaning that they have force at most 0.8), is approximately given by the area encompassed by the triangle $\{(0, 0), (0, 0.2), (0.4, 0)\}$. However, looking back at Figure 6.3, we see that most of the values in that triangle that satisfy $F_1(w_1, w_2) \leq 0.8$, satisfy $w_1 \leq w_2$.[10] Furthermore, the values of

[9] Of course, if we would want to be completely formal way, we should pick these values depending on the given ε. We will not be so precise, and just speak about "small" and "big".

[10] A rough estimation using triangles suggests that the fraction is about $5 : 3$.

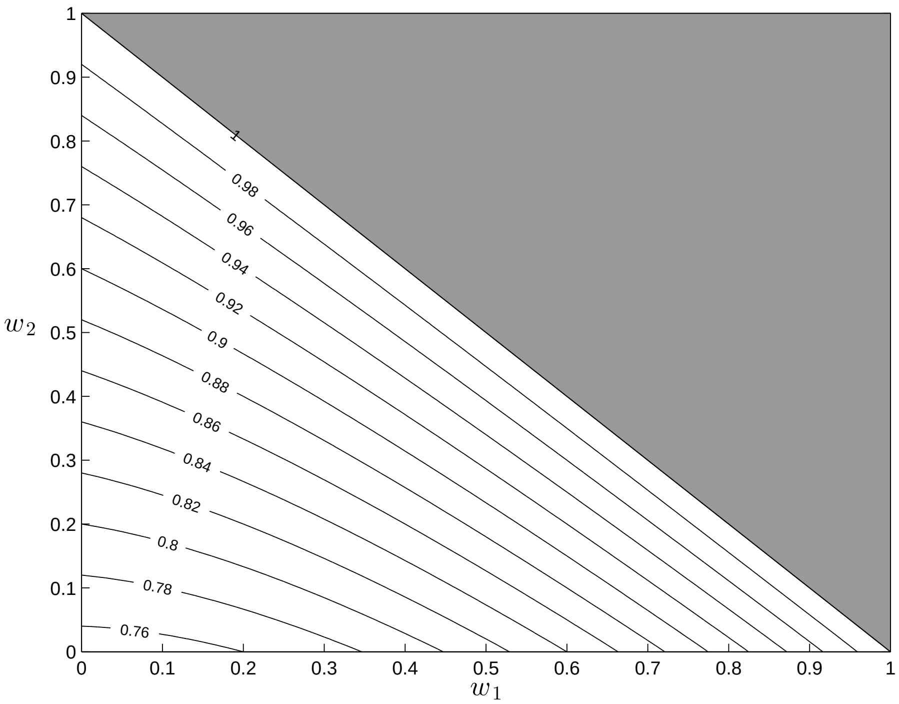

Figure 6.5: Contour lines of the function F_3. Made using MATLAB.

(w_1, w_2) that *"minimize"*[11] the bias B all satisfy $w_1 \leq w_2$.

We conclude that, in order to keep keep the bias low, we have to do two things:

- First, both w_1 and w_2 should not be too big.
- Second, given such "reasonably small" w_1, w_2, most pairs (w_1, w_2) which achieve a small bias satisfy w_1 being *smaller* than w_2. Moreover, all the pairs achieving the minimum (in the sense as explained in the above footnote) satisfy this inequality.

6.2.3 Case $n \geq 4$ and Concluding Remarks

Unfortunately, we did not find time to further investigate the case $n \geq 4$. Besides time constraints, there are a couple of technical issues. First, to calculate the probabilities *precisely* is, in principle, still possible using a tree-argument as we did earlier, but tedious; note that the number of branches of

[11] Of course, there is a unique minimum value of the bias B, namely 0. Obviously, we are not interested in that case, so we dismiss it. What we actually mean is: the set of *nonzero* values of the bias function B. Because the contour lines of Figure 6.3 lie much more closely together than the contour lines in Figures 6.4 and 6.5, the values of (w_1, w_2) that have a nonzero but "minimal" bias $B(w_1, w_2)$ can actually be found at the beginning (starting from the origin) of the 0.75-contour line in Figure 6.3. Observe that such couples (w_1, w_2) indeed satisfy $w_1 \leq w_2$.

the tree from Figure 6.1 increases exponentially in n. Also, we could still get numerical estimates by using a computer, but to make visualizations would be harder, as in this case all functions depends on at least *three* variables.

We finish by connecting back to the discussion about biases. What are the conclusions? As we have seen, what is certainly clear is the ultimate goal: to make the conditions such that individuals are capable of expressing their thoughts without inhibitions; that way, the group derives as much benefit from the diversity in the group as possible.

The question, then, is how that goal can be achieved. Kahneman [67, 35] advises the following concrete steps:

1. Before the meeting starts, all members secretly write down on paper a summary of their opinion.

2. The first person to speak is either picked uniformly at random (to avoid the same dominant personalities dominating the discussions time and again), or individuals are required to speak in *reverse* order of "dominance". In our terminology, that is: in reverse order of sway. In practice, this could, for example, be the level of seniority. The Supreme Court of the United States applies this method [22].

3. Disagreement should be supported and even rewarded.

A possible future challenge could be to try to incorporate rewards in our simple model for sequential voting.

Acknowledgments

First, I am grateful to Prof. Dr. Ronald de Wolf for his valuable advice throughout the thesis project. His comments, firm but always constructive, have benefited the present work very much. In retrospect, I can only praise myself lucky that he proposed this subject to me. It was great fun thinking about the topics of this thesis.

Next, I thank the other members of my thesis committee as well: Dr. Alexandru Baltag, Prof. Dr. Jan van Eijck, Dr. Ulle Endriss, and Prof. Dr. Hans Peters.

Thanks to Tanja for helping out at any moment. Her efficiency in dealing with issues of all sorts made the administrative side of life a lot easier.

Further, I would like to thank all ILLC members and Master of Logic students for providing a pleasant atmosphere to reason about this thing called logic. In particular, I am grateful to my friends Francesco and Thomas for all the (home)work we did together, and the nice times we shared. Much obliged to all my other friends as well for providing me with invaluable distractions; I remember each of the many enthralling discussions I had.

Finally, in exchanging thoughts with my three brothers, many stressful moments were eased. The same is true of my parents. Endless thanks to them.

Genoeg nu! Je moet ook van ophouden weten. Denk aan het woord van Goethe: 'In der Beschränkung zeigt sich erst der Meister'.

Harry Mulisch, *De ontdekking van de hemel*

Bibliography

[1] Kenneth J. Arrow. Social Choice and Individual Values. *New York*, 1951.

[2] John F. Banzhaf III. Weighted Voting Doesn't Work: A Mathematical Analysis. *Rutgers L. Rev.*, 19:317, 1964.

[3] John J. Bartholdi III and James B. Orlin. Single Transferable Vote Resists Strategic Voting. *Social Choice and Welfare*, 8(4):341–354, 1991.

[4] John J. Bartholdi III, Craig A. Tovey, and Michael A. Trick. The Computational Difficulty of Manipulating an Election. *Social Choice and Welfare*, 6(3):227–241, 1989.

[5] John H. Beck. Voting Cycles in Business Curriculum Reform, a Note. *The American Economist*, pages 83–88, 1997.

[6] William Beckner. Inequalities in Fourier Analysis. *Annals of Mathematics*, pages 159–182, 1975.

[7] Michael Ben-Or and Nathan Linial. Collective Coin Flipping. *Randomness and Computation*, 5:91–115, 1990.

[8] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise Sensitivity of Boolean Functions and Applications to Percolation. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 90(1):5–43, 1999.

[9] Eleanor Birrell and Rafael Pass. Approximately Strategy-proof Voting. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume One*, pages 67–72. AAAI Press, 2011.

[10] Bo H. Bjurulf and Richard G. Niemi. Strategic Voting in Scandinavian Parliaments. *Scandinavian Political Studies*, 1(1):5–22, 1978.

[11] Duncan Black. On the Rationale of Group Decision-making. *The Journal of Political Economy*, pages 23–34, 1948.

[12] Julian H. Blau. The Existence of Social Welfare Functions. *Econometrica: Journal of the Econometric Society*, pages 302–313, 1957.

[13] Julian H. Blau. A Direct Proof of Arrow's Theorem. *Econometrica: Journal of the Econometric Society*, pages 61–67, 1972.

[14] John C. Blydenburgh. The Closed Rule and the Paradox of Voting. *The Journal of Politics*, 33(01):57–71, 1971.

[15] Daniel Bochsler. The Marquis de Condorcet Goes to Bern. *Public Choice*, 144(1-2):119–131, 2010.

[16] Aline Bonami. Étude des coefficients de fourier des fonctions de $l^p(g)$. In *Annales de l'institut Fourier*, volume 20, pages 335–402. Institut Fourier, 1970.

[17] Felix Brandt, Vincent Conitzer, and Ulle Endriss. Computational Social Choice. *Multiagent systems*, pages 213–283, 2012.

[18] Eric C. Browne and Keith E. Hamm. Legislative Politics and the Paradox of Voting: Electoral Reform in Fourth Republic France. *British Journal of Political Science*, 26(02):165–198, 1996.

[19] Ioannis Caragiannis, Jason A. Covey, Michal Feldman, Christopher M. Homan, Christos Kaklamanis, Nikos Karanikolas, Ariel D. Procaccia, and Jeffrey S. Rosenschein. On the Approximability of Dodgson and Young Elections. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1058–1067. Society for Industrial and Applied Mathematics, 2009.

[20] John R. Chamberlin, Jerry L. Cohen, and Clyde H. Coombs. Social Choice Observed: Five Presidential Elections of the American Psychological Association. *The Journal of Politics*, 46(2):479–502, 1984.

[21] Yann Chevaleyre, Ulle Endriss, Jérôme Lang, and Nicolas Maudet. *A Short Introduction to Computational Social Choice*. Springer, 2007.

[22] Tom C. Clark. Internal Operation of the United States Supreme Court. *J. Am. Jud. Soc.*, 43:45, 1959.

[23] James S. Coleman. Control of Collectivities and the Power of a Collectivity to Act. *Social choice*, pages 269–300, 1971.

[24] Jules Coleman and John Ferejohn. Democracy and Social Choice. *Ethics*, pages 6–25, 1986.

[25] Robert Collins. CSE/EE486 Computer Vision I, Introduction to Computer Vision, Lecture 4: Smoothing. Pennsylvania State University Lecture Notes, 2007. available at http://www.cse.psu.edu/~rcollins/CSE486/.

[26] Fred Dallmayr. *The Promise of Democracy: Political Agency and Transformation*. SUNY Press, 2010.

[27] Nicolas de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Cambridge University Press, 2014.

[28] Ronald de Wolf. A Brief Introduction to Fourier Analysis on the Boolean Cube. *Theory of Computing, Graduate Surveys*, 1:1–20, 2008.

[29] Ronald de Wolf. Hypercontractivity in Theoretical Computer Science. Video Lecture, Monday February 23, 2015. available at http://www.birs.ca/events/2015/5-day-workshops/15w5098/videos.

[30] Henry A. Dietz and Michael J. Goodman. An Empirical Analysis of Preferences in the 1983 Multicandidate Peruvian Mayoral Election. *American Journal of Political Science*, pages 281–295, 1987.

[31] John Dobra and Gordon Tullock. An Approach to Empirical Measures of Voting Paradoxes. *Public Choice*, 36(1):193–194, 1981.

[32] John L. Dobra. An Approach to Empirical Studies of Voting Paradoxes: An Update and Extension. *Public Choice*, 41(2):241–250, 1983.

[33] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank Aggregation Methods for the Web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622. ACM, 2001.

[34] James S. Dyer and Ralph F. Miles Jr. An Actual Application of Collective Choice Theory to the Selection of Trajectories for the Mariner Jupiter/Saturn 1977 Project. *Operations Research*, 24(2):220–244, 1976.

[35] The Economist. Meeting up. Research hints at ways of making meetings more effective. Online Article, April 4, 2015. available at http://www.economist.com/news/finance-and-economics/21647680-new-research-hints-ways-making-meetings-more-effective-meeting-up.

[36] Ömer Eğecioğlu and Ayça E. Giritligil. The Impartial, Anonymous, and Neutral Culture Model: A Probability Model for Sampling Public Preference Structures. *The Journal of Mathematical Sociology*, 37(4):203–222, 2013.

[37] Scott Eidelman and Christian S. Crandall. The Intuitive Traditionalist: How Biases for Existence and Longevity Promote the Status Quo. *Advances in Experimental Social Psychology*, 50:53–104, 2014.

[38] Ulle Endriss. Logic and Social Choice Theory. *Logic and Philosophy Today*, 2:333–377, 2011.

[39] Ulle Endriss. Computational Social Choice. Lecture Slides, 2012. available at http://www.illc.uva.nl/~ulle/teaching/comsoc/2012/.

[40] Ronald Fagin, Ravi Kumar, and Dandapani Sivakumar. Efficient Similarity Search and Classification via Rank Aggregation. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 301–312. ACM, 2003.

[41] Dvir Falik and Alex Samorodnitsky. Edge-isoperimetric Inequalities and Influences. *Combinatorics, Probability and Computing*, 16(05):693–712, 2007.

[42] Piotr Faliszewski, Edith Hemaspaandra, Lane A. Hemaspaandra, and Jörg Rothe. The Shield that Never Was: Societies with Single-peaked Preferences are More Open to Manipulation and Control. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 118–127. ACM, 2009.

[43] Piotr Faliszewski and Ariel D. Procaccia. Ai's War on Manipulation: Are We Winning? *AI Magazine*, 31(4):53–64, 2010.

[44] Peter C. Fishburn. Voter Concordance, Simple Majorities, and Group Decision Methods. *Behavioral Science*, 18(5):364–376, 1973.

[45] Peter C. Fishburn and John D.C. Little. An Experiment in Approval Voting. *Management Science*, 34(5):555–568, 1988.

[46] Thomas Flanagan. The Staying Power of the Legislative Status Quo: Collective Choice in Canada's Parliament after Morgentaler. *Canadian Journal of Political Science*, 30(01):31–53, 1997.

[47] Merrill M. Flood. A Group Preference Experiment. *Mathematical Models of Human Behavior*, pages 130–148, 1955.

[48] Cees M. Fortuin, Pieter W. Kasteleyn, and Jean Ginibre. Correlation Inequalities on Some Partially Ordered Sets. *Communications in Mathematical Physics*, 22(2):89–103, 1971.

[49] Ehud Friedgut. Boolean Functions with Low Average Sensitivity Depend on Few Coordinates. *Combinatorica*, 18(1):27–35, 1998.

[50] Ehud Friedgut, Gil Kalai, Nathan Keller, and Noam Nisan. A Quantitative Version of the Gibbard-Satterthwaite Theorem for Three Alternatives. *SIAM Journal on Computing*, 40(3):934–952, 2011.

[51] Ehud Friedgut, Gil Kalai, and Noam Nisan. *Elections Can Be Manipulated Often*. Hebrew University, Center for the study of Rationality, 2008.

[52] Mark B. Garman and Morton I. Kamien. The Paradox of Voting: Probability Calculations. *Behavioral Science*, 13(4):306–316, 1968.

[53] Kurt Taylor Gaubatz. Intervention and Intransitivity: Public Opinion, Social Choice, and the Use of Military Force Abroad. *World Politics*, 47(4):534–554, 1995.

[54] William V. Gehrlein. The Expected Probability of Condorcet's Paradox. *Economics Letters*, 7(1):33–37, 1981.

[55] William V. Gehrlein. The Probability of a Condorcet Winner with a Small Number of Voters. *Economics Letters*, 59(3):317–321, 1998.

[56] William V. Gehrlein and Peter C. Fishburn. The Probability of the Paradox of Voting: A Computable Solution. *Journal of Economic Theory*, 13(1):14–25, 1976.

[57] Allan Gibbard. Manipulation of Voting Schemes: A General Result. *Econometrica: Journal of the Econometric Society*, pages 587–601, 1973.

[58] Tom Gole and Simon Quinn. Committees and Status Quo Bias: Structural Evidence from a Randomized Field Experiment. 2014. University of Oxford Department of Economics Discussion Paper No.733.

[59] Solomon Golomb. On the Classification of Boolean Functions. *Circuit Theory, IRE Transactions on*, 6(5):176–186, 1959.

[60] Leonard Gross. Logarithmic Sobolev Inequalities. *American Journal of Mathematics*, pages 1061–1083, 1975.

[61] Georges-Théodule Guilbaud. *Les théories de l'intérêt général et le problème logique de l'agrégation*, volume 63. Presses de Sciences Po, 2012.

[62] Theodore E. Harris. A Lower Bound for the Critical Probability in a Certain Percolation Process. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 56, pages 13–20. Cambridge Univ Press, 1960.

[63] Hamed Hatami. COMP 760 (Winter 2014): Harmonic Analysis of Boolean Functions. McGill University Lecture Notes, 2014. available at http://cs.mcgill.ca/~hatami/comp760-2014/.

[64] John Fuh-Sheng Hsieh, Niou Emerson, and Philip Paolino. Strategic Voting in the 1994 Taipei City Mayoral Election. *Electoral Studies*, 16(2):153–163, 1997.

[65] Marcus Isaksson, Guy Kindler, and Elchanan Mossel. The Geometry of Manipulation A Quantitative Proof of the Gibbard-Satterthwaite Theorem. *Combinatorica*, 32(2):221–250, 2012.

[66] Jeff Kahn, Gil Kalai, and Nathan Linial. The Influence of Variables on Boolean Functions. In *Foundations of Computer Science, 1988, 29th Annual Symposium on*, pages 68–80. IEEE, 1988.

[67] Daniel Kahneman. *Thinking, Fast and Slow*. Macmillan, 2011.

[68] Gil Kalai. A Fourier-theoretic Perspective on the Condorcet Paradox and Arrow's Theorem. *Advances in Applied Mathematics*, 29(3):412–426, 2002.

[69] Nathan Keller. Improved FKG Inequality for Product Measures on the Discrete Cube. 2008. available at www.math.huji.ac.il/~nkeller/Talagrand-Product5.pdf.

[70] Kiyoshi Kuga and Hiroaki Nagatani. Voter Antagonism and the Paradox of Voting. *Econometrica: Journal of the Econometric Society*, pages 1045–1067, 1974.

[71] Peter Kurrild-Klitgaard. An Empirical Example of the Condorcet Paradox of Voting in a Large Electorate. *Public Choice*, 107(1-2):135–145, 2001.

[72] Peter Kurrild-Klitgaard. Voting Paradoxes under Proportional Representation: Evidence from Eight Danish Elections. *Scandinavian Political Studies*, 31(3):242–267, 2008.

[73] Eerik Lagerspetz. Social Choice in the Real World. *Scandinavian Political Studies*, 16(1):1–22, 1993.

[74] Dominique Lepelley and Mathieu Martin. Condorcet's Paradox for Weak Preference Orderings. *European Journal of Political Economy*, 17(1):163–177, 2001.

[75] Saul Levmore. Public Choice Defended (reviewing Gerry Mackie, Democracy Defended (2003)), 2005.

[76] Christian List. *Mission Impossible?: The Problem of Democratic Aggregation in the Face of Arrow's Theorem.* PhD thesis, University of Oxford, 2001.

[77] Christian List. Social Choice Theory, 2013. available at http://personal.lse.ac.uk/list/pdf-files/sct-sep.pdf.

[78] Gerry Mackie. Is Democracy Impossible?: Riker's Mistaken Accounts of Antebellum Politics. 2001.

[79] Gerry Mackie. *Democracy Defended.* Cambridge University Press, 2003.

[80] Gerry Mackie. Deliberation, But Voting Too. *Approaching Deliberative Democracy: Theory and Practice*, pages 75–104, 2011.

[81] Yishay Mansour. Learning Boolean Functions via the Fourier Transform. In *Theoretical Advances in Neural Computation and Learning*, pages 391–424. Springer, 1994.

[82] Eric Maskin. Second Annual Arrow Lecture, Social Choice and Individual Values. Video Lecture, 2009. available at https://www.youtube.com/watch?v=ltTnw188950.

[83] Nicholas Mattei, James Forshee, and Judy Goldsmith. An Empirical Study of Voting Rules and Manipulation with Large Datasets. *Proc. of COMSOC-12*, 2012.

[84] Jennifer Roback Morse. Constitutional Rules, Political Accidents, and the Course of History. *Independent Review*, 2(2):173–200, 1997.

[85] Elchanan Mossel. A Quantitative Arrow Theorem. *Probability Theory and Related Fields*, 154(1-2):49–88, 2012.

[86] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise Stability of Functions with Low Influences: Invariance and Optimality. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 21–30. IEEE, 2005.

[87] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E. Steif, and Benny Sudakov. Non-interactive Correlation Distillation, Inhomogeneous Markov Chains, and the Reverse Bonami-Beckner Inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006.

[88] Elchanan Mossel and Miklós Z. Rácz. A Quantitative Gibbard-Satterthwaite Theorem Without Neutrality. In *Proceedings of the forty-fourth annual ACM symposium on Theory of Computing*, pages 1041–1060. ACM, 2012.

[89] Hervé Moulin. *The Strategy of Social Choice*, volume 18. Elsevier, 2014.

[90] Lev Muchnik, Sinan Aral, and Sean J. Taylor. Social Influence Bias: A Randomized Experiment. *Science*, 341(6146):647–651, 2013.

[91] Akira Nakashima. Theory of Relay Circuit Composition. *Nippon El. Com. Eng*, (3):197–206, 1936.

[92] John L. Neufeld, William J. Hausman, and Ronald B. Rapoport. A Paradox of Voting: Cyclical Majorities and the Case of Muscle Shoals. *Political Research Quarterly*, 47(2):423–438, 1994.

[93] Richard G. Niemi. The Occurrence of the Paradox of Voting in University Elections. *Public Choice*, 8(1):91–100, 1970.

[94] Helmut Norpoth. The Parties Come to Order! Dimensions of Preferential Choice in the West German Electorate, 1961–1976. *American Political Science Review*, 73(03):724–736, 1979.

[95] Ryan O'Donnell. CMU 18-859S (Spring 2007) Analysis of Boolean Functions. Carnegie Mellon Lecture Notes, 2007.

[96] Ryan O'Donnell. 15-859s / 21-801a: Analysis of Boolean Functions 2012, lectures 1, 3, 4. Video Lecture, 2012. available at https://www.cs.cmu.edu/~odonnell/aobf12/.

[97] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[98] Will Oremus. You Can't Dislike This Article. Facebook's inability to handle criticism is bad for democracy. Online Article, December 15, 2014. available at http://www.slate.com/articles/technology/future_tense/2014/12/facebook_dislike_button_why_mark_zuckerberg_won_t_allow_it.html.

[99] Christos H. Papadimitriou. *Computational Complexity*. John Wiley and Sons Ltd., 2003.

[100] Gert K. Pedersen. *Analysis Now*, volume 118. Springer New York, 1989.

[101] Lionel S. Penrose. The Elementary Statistics of Majority Voting. *Journal of the Royal Statistical Society*, pages 53–57, 1946.

[102] Sergey V. Popov, Anna Popova, and Michel Regenwetter. Consensus in Organizations: Hunting for the Social Choice Conundrum in APA Elections. *Decision*, 1(2):123, 2014.

[103] Ariel D. Procaccia. Computational Social Choice: The First Four Centuries. *XRDS*, 18(2):31–34, December 2011.

[104] Ramsey M. Raafat, Nick Chater, and Chris Frith. Herding in Humans. *Trends in Cognitive Sciences*, 13(10):420–428, 2009.

[105] Benjamin Radcliff. Collective Preferences in Presidential Elections. *Electoral Studies*, 13(1):50–57, 1994.

[106] Michel Regenwetter. Perspectives on Preference Aggregation. *Perspectives on Psychological Science*, 4(4):403–407, 2009.

[107] Michel Regenwetter and Bernard Grofman. Approval Voting, Borda Winners, and Condorcet Winners: Evidence from Seven Elections. *Management Science*, 44(4):520–533, 1998.

[108] Michel Regenwetter, Aeri Kim, Arthur Kantor, and Moon-Ho R. Ho. The Unexpected Empirical Consensus Among Consensus Methods. *Psychological Science*, 18(7):629–635, 2007.

[109] William H. Riker. The Paradox of Voting and Congressional Rules for Voting on Amendments. *American Political Science Review*, 52(02):349–366, 1958.

[110] William H. Riker. Arrow's Theorem and Some Examples of the Paradox of Voting. *Mathematical Applications in Political Science*, 1:41–60, 1965.

[111] William H. Riker. *Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice*. Freeman San Francisco, 1982.

[112] Nicholas C. Romano and Jay F. Nunamaker Jr. Meeting Analysis: Findings from Research and Practice. In *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, pages 13–pp. IEEE, 2001.

[113] Michael D. Rosen and Richard J. Sexton. Irrigation Districts and Water Markets: An Application of Cooperative Decision-making Theory. *Land Economics*, pages 39–53, 1993.

[114] Raphaël Rossignol. Threshold for Monotone Symmetric Properties Through a Logarithmic Sobolev Inequality. *The Annals of Probability*, pages 1707–1725, 2006.

[115] Jean-Jacques Rousseau. *Du contrat social*. Alcan, 1896.

[116] Mark Allen Satterthwaite. Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

[117] Amartya Sen. Liberty and Social Choice. *The Journal of Philosophy*, pages 5–28, 1983.

[118] Amartya Sen. Rationality and Social Choice. *American Economic Review*, 85(1):1–24, 1995.

[119] Amartya Sen. The Possibility of Social Choice. *American Economic Review*, pages 349–378, 1999.

[120] Amartya Sen. Second Annual Arrow Lecture, Social Choice and Individual Values. Video Lecture, 2009. available at https://www.youtube.com/watch?v=L0rF4AC9IVY.

[121] Amartya Sen and Eric Maskin. *The Arrow Impossibility Theorem*. Columbia University Press, 2014.

[122] Amartya Kumar Sen. *Choice, Welfare and Measurement*. Harvard University Press, 1997.

[123] Claude Elwood Shannon. A Symbolic Analysis of Relay and Switching Circuits. *American Institute of Electrical Engineers, Transactions of the*, 57(12):713–723, 1938.

[124] Warren D. Smith. The Romanian 2009 Presidential Election Featured One or More High Condorcet Cycles. *RangeVoting.org., Center for Range Voting*, 13, 2009.

[125] Ernst Snapper. The Three Crises in Mathematics: Logicism, Intuitionism and Formalism. *Mathematics Magazine*, pages 207–216, 1979.

[126] Daniel Štefankovič. Fourier Transform in Computer Science, 2000. Master's thesis, University of Chicago.

[127] Eivind Stensholt. Voteringens kvaler: flyplassaken i Stortinget 8. Oktober 1992. *Sosialøkonomen*, 4:28–40, 1999.

[128] Patrick Suppes. The Pre-history of Kenneth Arrow's Social Choice and Individual Values. *Social Choice and Welfare*, 25(2-3):319–326, 2005.

[129] James Surowiecki. *The Wisdom of Crowds.* Anchor, 2005.

[130] Richard Szeliski. *Computer Vision: Algorithms and Applications.* Springer Science & Business Media, 2010.

[131] Pingzhong Tang and Fangzhen Lin. Computer-aided Proofs of Arrow's and Other Impossibility Theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

[132] Terence Tao. Arrow's Theorem. 2012. available at http://www3.nd.edu/~pweithma/justice_seminar/Arrow%27s%20Theorem/Tao/Tao%20%28on%20Arrow%29.pdf.

[133] Ross H. Taplin. The Statistical Analysis of Preference Data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(4):493–512, 1997.

[134] Alan D. Taylor. A Glimpse of Impossibility. *Perspectives on Political Science*, 26(1):23–26, 1997.

[135] Michael Taylor. Graph-theoretic Approaches to the Theory of Social Choice. *Public Choice*, 4(1):35–47, 1968.

[136] William Thomson. International meeting of the society for social choice and welfare. Condorcet Lecture, 2000.

[137] Nicolaus Tideman. *Collective Decisions and Voting.* Ashgate Burlington, 2006.

[138] Mitsuhiko Toda, Kozo Sugiyama, and Shojiro Tagawa. A Method for Aggregating Ordinal Assessments by a Majority Decision Rule. *Mathematical Social Sciences*, 3(3):227–242, 1982.

[139] M. Truchon. Rating Skating and the Theory of Social Choice. Université Laval, 1998. Unpublished Manuscript, available at http://www.ecn.ulaval.ca/w3/recherche/cahiers/1998/9814.pdf.

[140] Ilia Tsetlin, Michel Regenwetter, and Bernard Grofman. The Impartial Culture Maximizes the Probability of Majority Cycles. *Social Choice and Welfare*, 21(3):387–398, 2003.

[141] Raquel Urtasun. Computer Vision: Filtering. Lecture Slides, 2013. available at http://www.cs.toronto.edu/~urtasun/courses/CV/lecture02.pdf.

[142] M. Van Dam. A Purple Paradox: Decision-making about the Modernization of the University Administration (MUB). *Acta Politica*, 33:77–85, 1998.

[143] Ad M.A. van Deemen and Noël P. Vergunst. Empirical Evidence of Paradoxes of Voting in Dutch Elections. In *Empirical Studies in Comparative Politics*, pages 257–272. Springer, 1998.

[144] Adrian van Deemen. The Probability of the Paradox of Voting for Weak Preference Orderings. *Social Choice and Welfare*, 16(2):171–182, 1999.

[145] Adrian van Deemen. On the Empirical Relevance of Condorcets Paradox. *Public Choice*, 158(3-4):311–330, 2014.

[146] N. Vergunst. Besluitvorming over kerncentrale Borssele: Een analyse van de stemparadox in de Nederlandse politiek. *Acta Politica*, 31:209–228, 1996.

[147] Joseph L. Walsh. A Closed Set of Normal Orthogonal Functions. *American Journal of Mathematics*, pages 5–24, 1923.

[148] Alan Weir. Formalism in the Philosophy of Mathematics. Online Article, 2015. available at http://plato.stanford.edu/archives/spr2015/entries/formalism-mathematics/.

[149] Shaun White. William Riker's *Liberalism Against Populism*. CMSS Talk, 2013. available at http://cmss.auckland.ac.nz/files/2013/10/CMSStalkslides15Oct.pdf.

[150] John Ziker. The Long, Lonely Job of Homo Academicus. Focusing the research lens on the professor's own schedule. Online Article, March 31, 2014. available at https://theblueview.org/faculty-time-allocation/.