LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Descriptive Assumptions and Normative Justifications in Social Choice Theory: Ambiguity, Strategic Voting and Measurement

Bele Wollesen

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

at the

Department of Philosophy, Logic and Scientific Method

March, 2025

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work. The copyright of this thesis rests with the author. In accordance with the Regulations, I deposited an electronic copy of it in LSE Theses Online held by the British Library of Political and Economic Science and will grant permission for my thesis to be made available for public reference. Otherwise, this thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 40958 words.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors, Liam Kofi Bright and Johanna Thoma. I am immensely grateful to have had such supportive and kind supervisors. This thesis was only possible because of their encouragement, critical feedback, and their bravery in working through my first drafts, which have now become the chapters of this thesis. Both have been invaluable in my academic journey, guiding me to grow as a philosopher and helping me navigate the challenges of graduate studies.

Special thanks should also be given to three others at LSE. Kate Vredenburgh mentored and encouraged me in my teaching and professional development and provided significant feedback on what is now Chapter 2 of the thesis. Richard Bradley has been invaluable in offering encouragement and feedback on what is now Chapter 1, as has Anna Mahtani. I am thankful to all three for enriching my experience at LSE by discussing ideas with me and providing deep insights into their respective fields.

I have been very fortunate to be surrounded by brilliant peers at LSE and beyond. My sincere thanks go to those who generously offered written feedback or engaged in discussions with me on earlier drafts of the material in this thesis: Eva Schmidt, Sebastian Zezulka, Shira Ahissar, eva Read, James Michelson, Matteo Michelini, Sam Kang, and most of all, Lukas Beck—my peer and partner for almost a decade. Thank you for tirelessly reading drafts and exchanging ideas over walks and dinners.

I would also like to thank my family, especially my sister Gyde, to whom the first chapter is dedicated—without her, I wouldn't have been able to work on the chapter during lockdown. I am also thankful for my grandparents, Oma Lisa and Opa Nike, whose steadfast belief in my ability to succeed, means the world to me. I would also like to thank the newest additions in my family, Janne, Luca, and Oscar, and their parents. Spending time with them was the very best distraction from writing this thesis.

This thesis is dedicated to the memory of my mother, Gabriele. As a single mother of three working night shifts, she decided it was a good idea to also become the first in our family to attend university. From her, I learned what a privilege and joy it is to gain a deeper understanding of what you are passionate about.

Abstract

Social choice theory provides fundamental tools for many sciences. Moreover, its models are applied and implemented in a remarkable range of diverse contexts to guide collective decision procedures. Understanding these models of collective decision-making and their normative verdicts poses an important challenge to philosophers of science, epistemologists, political philosophers, and philosophers of technology. This thesis is a collection of four papers that question the assumptions in social choice theory models at the intersection of individual decision-making and the design of collective decision mechanisms. The chapters, in particular, target the normative verdicts of these models by drawing on perspectives from the philosophy of science, political philosophy, and psychology. The central claims defended in the thesis are twofold. Firstly, the core claim is that abstracting away from individual decision-making does not always lead to greater generality but often makes the verdicts of social choice models inapplicable to the intended target systems. This observation draws on the fact that most target systems of social choice models involve individuals. Individuals are significant both in how they interact within a collective decision mechanism and as the object of normative consideration. Particular assumptions about individual decision-making, although often absent in a given social choice theory model itself, are needed for the model's normative verdict to transfer to many target systems. Secondly, many models developed by social choice theory not only require scrutiny of their descriptive assumptions but also need an explication of the normative commitments underpinning them. Normative assumptions in social choice theory typically take the form of axioms or desiderata that are prima facie difficult to reject. However, the values underlying the acceptance of a given axiom can vary greatly, which, in turn, can significantly affect whether the model's normative verdict successfully transfers to the intended target system.

Contents

0

Introduction

Social choice theory is not simply a descriptive tool but a modelling framework that shapes decision mechanisms in the world. It isolates and embeds desirable normative properties in decision-making systems. Given the growing application of social choice theory in fields such as computer science, it is time to renew philosophical engagement with these models.

This thesis aims to address some of the gaps in the philosophical literature on social choice theory, a field that has surprisingly received limited attention from philosophers of science. That is despite the fact that social choice theory provides fundamental tools for many sciences and is very broad in its application (Sen, 1977). In economics, social choice theory provides essential tools for analysing welfare economics, public choice, and resource allocation. In political science, it helps investigate the distribution of voting power in different voting mechanisms. Meanwhile, in computer science, particularly in algorithmic decision-making and artificial intelligence, social choice theory is applied to recommender systems and multi-agent systems. The lack of attention is all the more surprising, given that decision and game theory have been extensively debated within philosophy, especially in the philosophy of economics.[1] The four chapters of the thesis each offer a novel contribution, addressing both descriptive and normative assumptions in social choice theory models, particularly at the intersection of individual decision-making and the design of collective decision mechanisms. Social choice theory strives for general models and thereby often abstracts away from modelling details about individuals (with the exception of preferences). In this regard, the first core claim of this thesis is that abstracting away from individual decision making does not always mean more generality but often makes the verdicts of social choice models inapplicable to the intended target systems. The second core claim of the thesis is that many of the models developed by social choice theory do not only need scrutinising of their descriptive assumptions, but are in need of an explication of the normative commitments that underpin them.

Social choice theory consists of mathematical models (List, 2022) that analyse collective decision-making, particularly the aggregation of individual preferences into a collective outcome. Mathematical models rely on various forms of idealisation and abstraction (Mäki et al., 1992). They can allow us to formalise hypotheses about the world in a precise and testable way, enabling the prediction, explanation, and understanding of complex phenomena (Pincock, 2007). This abstraction may support a form of "surrogate reasoning", where we manipulate symbols according to mathematical rules rather than directly interacting with the physical world (Swoyer, 1991). Such models are not merely descriptive but can also be explanatory, offering insights into why systems behave in certain ways under specific conditions (Cartwright, 1983).[2]

[1]Discussions in formal social epistemology, particularly on the social organization of science and epistemic democracy, have engaged with related themes. However, these reflections are typically embedded within broader applications of social choice models rather than focusing explicitly on social choice theory itself.

[2]This thesis engages with the normative status of social choice theory but does not directly address broader debates in the philosophy of science on the status of idealized models. A more extensive discussion of these debates lies beyond its scope; however,

A key research focus in social choice theory, and the central object of this thesis, is the investigation of social choice rules—mechanisms that translate individual preferences into collective decisions. These rules isolate essential properties of collective decision-making processes, such as voting, often in contexts where obtaining direct empirical observation is challenging (Chamberlin, 1985). This is similar to models in other disciplines. For example in macroeconomics we can seldomly conduct randomised controlled trials to investigate the effects of tax policies. Empirical investigations are often infeasible. Thus, we build mathematical models to represent the behaviour of economies at a large scale. The model is then manipulated and studied and used to generate conclusions that aim to be transferable to the real world system that is modelled. Similarly, as in macroeconomics, empirical investigation of voting systems is hard. For instance, we can rarely reconstruct the complete preference rankings of voters from their ballot choices. Even when such reconstructions are possible, many of the voting rules we wish to investigate are either not currently in practice or are employed too infrequently to offer a more comprehensive understanding (Chamberlin, 1985).

Social choice theory is neither a descriptive theory—concerned with how individuals make decisions—nor a predictive theory[3], which seeks to forecast choices and abstracts from the psychological processes involved in decision-making.[4] Social choice theory predominantly works with mathematical models that aim to provide normative guidance. And yet there is rather little discussion within social choice theory or in philosophy of science on social choice theory as a modelling practice. And even less so on how to infer norma-

insights from this literature remain relevant to understanding the role of social choice models in normative inquiry.

[3]Unlike public choice theory, which also study collective decision-making but is more likely than social choice theory to publish positive and non-mathematical work (Mueller, 2015), social choice theory primarily addresses normative questions (Mueller, 2015).

[4]See Briggs (2023) for a similar description of expected utility as a normative theory.

tive guidance from social choice models. Generally, when inferring about the target system—the real-world phenomenon being modelled—it is essential to account for the ways in which idealisations have shaped those conclusions. If we have normative models[5] we face the additional challenge of determining how to infer from a normative result generated by an idealised model to a normative conclusion about the relevant aspect in the real world (Roussos, 2022). Normative models (Beck & Jahn, 2021) in social choice theory rest on both descriptive and normative assumptions. Some of the latter are made explicit—such as the canonical criteria articulated by Arrow's Impossibility Theorem show.

Arrow's Impossibility Theorem (1951/1963) is one of the most significant contributions to social choice theory, establishing that no social welfare function can satisfy a set of seemingly reasonable conditions (i.e., unrestricted domain, weak Pareto principle, independence of irrelevant alternatives, and non-dictatorship[6]) simultaneously, if there are three or more options to choose from (Thomson, 2001). A social welfare function takes as inputs preferences over outcomes from voters and outputs a ranking over these outcomes, often called social preferences (Morreau, 2019).

Arrow's Impossibility Theorem, and more generally the axiomatic approach in social choice theory that became dominant in the wake of it, serves to evaluate collective decision-making procedures against a set of formal crite-

[5]See Beck and Jahn (2021) for an analysis how normative models can have normative guidance.

[6]Unrestricted domain ensures that any set of individual preferences can be considered; the weak Pareto principle requires that if every individual prefers one option over another, then the outcome should reflect this preference; independence of irrelevant alternatives means that the social preference between any two options should depend only on individual preferences for those options, not on any other options; and non-dictatorship ensures that no single individual has the power to dictate the group's preference.

ria (i.e., axioms) that reflect normative ideals.[7]

By adopting this axiomatic approach, we typically regard properties like the Pareto principle and non-dictatorship as normatively desirable, and evaluate a social choice rule's fulfilment of these criteria as indicative of its overall desirability. However, for these properties to carry normative force, there must be some correspondence between the satisfaction of these desiderata within the model and the model's applicability to real-world elections. This does not imply that we need to have a specific election in mind; rather, we may seek general models with the expectation that they apply to a broad range of target systems. Naturally, there is a trade-off between the generality of a model and its accuracy in describing specific real-world applications. Nevertheless, this trade-off does not negate the necessity of justifying the normative force of our model's conclusion by establishing a clear link between the mathematical model and the real-world systems the model seeks to represent. In order to do so we need to scrutinise the (implicit) descriptive and normative assumptions made by social choice models.

In the following, I will outline how the chapters in this thesis contribute to this goal. Each chapter either tests descriptive or normative assumptions in social choice rules at the intersection of individual decision-making and collective decision-making. Chapter 1 *The Condorcet Jury Theorem under Ambiguity* and also chapter 4 *A Heuristic Approach to Manipulation: Simulating Simple Strategic Voting* assess the robustness of descriptive assumptions about voters and their impact on normative evaluations of social choice rules. Chapter 2 *On Equality, Power, and Strategic Voting* and chapter 3 *Manipulability, Measurement, and Mathematical Models* deal with implicit normative assumptions underpinning social choice theory.

[7] All of these desiderata remain standard criteria against which social welfare functions and social choice rules are evaluated. Social choice rules don't output a ranking over outcomes, instead the output is one winner or collection thereof (List, 2022).

Chapter 1 (The Condorcet Jury Theorem under Ambiguity) explores the implications of voters whose epistemic states cannot be represented by precise probabilities, and how this affects judgement aggregation under the majority rule. Judgement aggregation extends voting an to aggregation where individuals express judgements on propositions (e.g., legal decisions). The challenge, similarly to preference aggregation, lies in aggregating these judgements into an outcome that satisfies desirable properties, such as producing outcomes that tend to be correct.

Formal results in the field of judgement aggregation such as the Condorcet Jury Theorem explicate assumptions about voting behaviour. Yet, while the assumptions in the formal results are about behaviour, they are often informally motivated by assumptions about the agent's beliefs or credences, leaving a gap between the justification of an assumption and the assumption itself in the formal result. This chapter demonstrates that, under ambiguity, voters may fail to act competently, even when they are honest, rational, and epistemically competent. This contrasts with situations under risk, where voter competence is guaranteed under these conditions. Additionally, it identifies new collective decision-making procedures that may be better suited to less idealised uncertainty frameworks, showing that allowing abstention can enhance the epistemic benefits of voting, thereby extending the Condorcet Jury Theorem.

The next three chapters of this thesis examine a property of social choice rules called non-manipulability, with a particular focus on the satisfaction of strategy-proofness. Two of these chapters address the normative assumptions that play a role in the justification and measurement of non-manipulability in social choice rules.[8] In other words, they focus on normative assumptions

[8]Strategy-proofness is a desideratum across many collective decision mechanism for example in mechanism design where agents on both sides of the market (e.g., students

in social choice theory and their influence on the evaluation of social choice rules. The final chapter returns to descriptive assumptions, focusing on the descriptive idealizations of voters that influence judgements about the degree of manipulability.[9]

Chapter 2 (On Equality, Power, and Strategic Voting) addresses the puzzling nature of strategic voting. While there is broad consensus that strategic voting should be prevented, little attention has been given to why it is problematic. Building on early concerns in social choice theory that strategic voting turns elections into a *game of skill*, this chapter develops a normative argument: strategic voting can skew power relations among citizens. The chapter challenges the assumption that "one person, one vote" is the sole property needed to maintain equal power relations. By examining different theories of power in voting, it shows how strategic voting can distort these power relations. As such, normative theories concerned with power imbalances, like relational egalitarianism, should take a stance on strategic voting

and schools, or workers and firms) have preferences over whom they are matched with. Strategy-proofness in social choice rules combines decision and game theoretical considerations with the design of social choice rules. And yet, in the literature of social choice rules it is usually shied away from making explicit assumptions about individual decision makers (in contrast to the field of mechanism design and game theory). This means that the connection between the design of social choice rules and models of individual decision making is surprisingly often left implicit. It is then a particular curious balancing act of trying to retain a normative verdict while not defending descriptive assumptions of one's model.

[9] The Gibbard-Satterthwaite theorem, alongside Arrow's Impossibility Theorem, stands as one of the most influential results in social choice theory. Formulated independently by Allan Gibbard (1973) and Mark Satterthwaite (1975), the theorem establishes that any non-dictatorial, resolute social choice rule that ranks three or more options is necessarily susceptible to manipulation by voters—that is, it is not strategy-proof. For a comprehensive history of the study of manipulability in voting rules, see Barberà (2011) and Farquharson (1969). A resolute social choice rule always selects a unique winning alternative. For a social choice rule to be non-manipulable, i.e., strategy-proof, it must always be a (weakly) dominant strategy for voters to truthfully reveal their preferences. A weakly dominant strategy is one that is the best response across all possible preference profiles.

and influence the design and implementation of voting rules.

Chapter 3 (Manipulability, Measurement, and Mathematical Models) investigates how we can measure the vulnerability of institutions to strategic voting. Although there is general agreement that manipulation is bad, understanding why is it is bad is crucial to assessing how susceptible voting rules are to it. This chapter applies insights from the philosophy of science, particularly on the measurement of thick concepts, to show that measuring manipulation requires normative assumptions about what renders manipulation bad. The chapter argues that the current lack of engagement with the normative evaluation of manipulation hinders our ability to evaluate existing measures of manipulation. It then assesses two prominent frameworks in social choice theory—the Nitzan-Kelly index and the classification of voting rules by computational complexity—emphasising the need for a debate on the normative evaluation of manipulability in order to refine frameworks that better capture the normative dimensions of manipulation.

Chapter 4 (A Heuristic Approach to Manipulation: Simulating Simple Strategic Voting) focuses on the conditions that lead to manipulation. Specifically, it challenges the assumption that more expressive voting methods—those that allow voters to express more detailed preferences—necessarily lead to more manipulation. This is commonly understood to be the case because more expressive methods give the voters more opportunities to manipulate. However, through computer simulations, this chapter shows that while expressiveness leading to increased manipulation seems true under assumptions of utility maximising voters, it is not true under assumptions of heuristic voters. In fact, once rational utility maximisation is relaxed in favour of satisficing (one prominent heuristic), expressive voting methods can in fact minimise manipulation.

All four chapters show that a lack of attention to normative and descriptive assumptions can not only render these models inapplicable to certain target systems but also reverse the normative verdicts of social choice models. Therefore, if social choice theory aims to provide normative judgements on collective decision mechanisms, it must engage more deeply with its descriptive and normative assumptions, as well as their connection to the intended target systems. A failure to do so should not be mistaken for creating general models.

1

The Condorcet Jury Theorem under Ambiguity

Abstract

This chapter evaluates the Condorcet Jury Theorem under ambiguity. It explores the effects on voter competence assumption when voters are faced with situations they can't ascribe a single probability anymore. In contrast to voting in situations where voters can do so, this chapter shows that voters can fail to vote competently in situations under ambiguity even if they are honest, practically rational and epistemically competent. Thus, the conditions under which we can guarantee voter competence become obscure once we adopt a less idealised uncertainty framework. Namely, conditions that guarantee voter competence under risk do not guarantee voter competence under ambiguity. The second contribution is a more positive one. There is a fruitful research project that identifies collective decision procedures better suited for less idealised uncertainty frameworks. In relation to this, the chapter shows how allowing abstention can have positive effects on the epistemic benefits of voting and extends the Condorcet Jury Theorem.

1.1 Introduction

In this chapter I explore the applicability of one of the most well-known theorems in judgement aggregation to situations of ambiguity. Ambiguity is topic that has received much attention in decision theory but comparably little attention in social choice theory. When it has received attention in social choice theory it is more commonly modelled on the level of preferences, with little being said on decision principles that agents may use on the basis of such preferences. The aim of the chapter is to put pressure on the importance of the modelling choice of voter's epistemic states in judgement aggregation. Our choice of modelling framework can overturn our evaluation of voting rules.

This chapter evaluates the Condorcet Jury Theorem (CJT) under ambiguity. It demonstrates that under conditions of ambiguity, one of the theorem's key assumptions, namely the competence assumption, can fail in non-trivial ways. The CJT is one of the most important theorems in social choice theory and well-known beyond the field. Informally, the CJT in its classic form states that if there are two verdicts of which one is correct, and voters are competent (i.e., better than a coin flip at voting for the correct verdict) and independent of each other, then the probability that the majority vote outcome tracks the correct verdict increases with the size of the group, tending towards one if the group is infinite. Let us call this the epistemic benefits described by the CJT.

The influence of the CJT and its epistemic benefits is hard to overstate. The CJT has been applied to juries (e.g., Penrod & Hastie, 1979; Urken & Traflet, 1983), organizations (Nitzan & Paroush, 1982), crowd-sourced peer review (Arvan, Bright, & Heesen, 2022), methodological triangulation (Heesen, Bright, & Zucker, 2019), and generally serves as one of the arguments for democratic decision-making and the use of majority rule (Goodin & Spiekermann, 2018). In the literature, it is common practice to rigorously justify the applicability of the independence assumption when using the CJT. However, such rigorous justification is often deemed less necessary for the competence assumption in the absence of specific factors (e.g., misleading propaganda).[1] I demonstrate that this is a mistake and that the competence assumption warrants more attention. It is important not to conflate an assumption of epistemic competence with a behavioural assumption of competence, which is required for the CJT. In this regard, the chapter develops two important contributions to the literature: First, it shows that the conditions under which we can guarantee voter competence become obscure once we adopt a less idealised uncertainty framework. And second, there is a fruitful research project that identifies collective decision procedures better suited to less idealised uncertainty frameworks. For example, we will see that allowing abstention can help secure the epistemic benefits of voting.

There already exists a body of work checking the robustness of the CJT, leading to extensions of the original result and identifying the limits of its application. For example, it has been shown that not all individuals in the group need to be competent[2] (see, among others Grofman, Owen, & Feld, 1983; Boland, 1989). We also know that certain sorts of dependencies between judgements can retrieve the CJT (Ladha, 1992, 1993; Dietrich & Spiekermann, 2013). Moreover, List and Goodin (2001) show that the CJT can be generalized from the majority rule over two verdicts to the plurality rule over many verdicts; the plurality rule allows every voter to cast one vote, and the verdict with the most votes is elected.

However, the literature has yet to consider how the CJT holds up in an

[1] For notable exceptions, see Dietrich (2008), Dietrich and Spiekermann (2013) and Romeijn and Atkinson (2011).

[2] It suffices if individuals are, on average, competent.

ambiguous world. In decision theory, ambiguity refers to a type of uncertainty where we cannot assign a specific probability to an event. If we can assign a precise probability to an event, this is called decision-making under risk. For example, in roulette, you can assign a precise probability to each outcome. Traditionally, this is contrasted with decision-making under uncertainty. In such cases, you have no information about an event's probability. Typically, we are not completely ignorant of the truth but need more information to assign a precise probability. Arguably, we often find ourselves in this situation when we cast a vote. For instance, you might think that if a party is elected, it is more likely than not to support investment in green energy, even if you cannot assign a precise probability to it. In decision theory, there is an increasing acceptance of the descriptive as well as normative inadequacy of modeling agents' epistemic states through precise probabilities (e.g., R. Bradley, 2017; Seidenfeld, 2004; Joyce, 2010; Kyburg Jr, 1983; Levi, 1985). This chapter aims to take these concerns seriously and asks under which conditions we can still hope for competent voters and, thus, the epistemic benefits of voting. Prior work on voter competence modeled voters' epistemic states through precise probabilities. It also assumed some form of maximizing expected utility (e.g., Austen-Smith & Banks, 1996). Both assumptions are not directly applicable when voters need to make decisions based on ambiguous attitudes. Thus, this chapter will deviate from previous work by modeling epistemic states as sets of probabilities.

This chapter is structured as follows. I introduce the CJT with a simple model for agents who can be described using precise probabilities. For these agents, voter competence is guaranteed under assumptions of rationality, honesty, and epistemic competence. Next, I motivate and expand the model. I explicate how we model the epistemic states of rational agents that involve ambiguity and how to fill in our assumptions given those epistemic states. We explore why voter competence cannot be guaranteed under these conditions. Finally, we revisit assumptions, propose alternatives, and explore how to retrieve voter competence.

1.2 CJT and a Simple Model of Voter Competence

Imagine a researcher named Gyde who has to brief a policymaker about new insights in climate science. This briefing will include many agenda points, including whether the global temperature will rise over 1.5°C between 2030 and 2052 if we continue as we do now. As they are short on time, they ask their research assistants to read a paper each and email them back whether the statement that *global warming will reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate* or the statement that *global warming will not reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate* should be included in the briefing.[3] Would it be a good idea for them to simply follow whatever the majority of their research assistants advise? To help answer this, let us take a closer look at the CJT and the assumptions needed to apply it.

1.2.1 The Condorcet Jury Theorem

While the core idea of the CJT is simple, we will need to introduce some notation at this point to rely upon later on.[4] Let us start by assuming that we have n voters, labelled 1, ..., n. These voters have to make a decision over χ. In our example, χ consists of two opposing statements that can be included in a briefing, and our voters are the research assistants. Generally, χ is a set that contains multiple verdicts x, of which voters need to pick one. Yet, only one x is the correct verdict. The verdict may be correct according

[3]This set-up is similar to the example used by Steele (2012).

[4]This way of presenting the CJT is inspired by (List & Spiekermann, 2016).

to an epistemic, moral, or any other standard. In this chapter, we assume that to be correct is to match an (external) truth. Let us call the correct verdict in χ, x. We say that $\{T, F\} = \chi$, and thus $X = T$ or $X = F$. For each voter i, we write V_i to denote voter i's vote, where V_i can take the values T or F. Here, $V_i = T$ represents a vote for verdict T from voter i, while $V_i = F$ represents a vote for verdict F. We measure voter competence for each voter i and each possible truth x as $Pr(V_i = x|X = x)$. If, then, the two following assumptions hold, we can derive the CJT.

1. **Voter competence** is satisfied if for every voter i, $Pr(V_i = x|X = x)$ exceeds 0.5.

2. **Voter independence** is satisfied if the votes of all voters, $V_1, V_2, ..., V_n$, are mutually independent, conditional on the truth.[5]

Additionally, for any voters $i \neq j$, we assume $Pr(V_i = x|X = x) = Pr(V_j = x|X = x)$. Informally, this says that each voter is better than random at voting for the truth and equally proficient at doing so. While not needed to derive the CJT, we will also assume for any voter i, $Pr(V_i = T|X = T) = Pr(V_i = F|X = F)$.[6] Let us call the outcome of a majority vote of the n voters O. The outcome O is $O = T$ if there are more voters with $V_i = T$ than with $V_i = F$; $O = F$ if there are more voters with $V_i = F$ than with $V_i = T$; and $O = 0.5$ if there is a tie. The CJT then guarantees what we called the *epistemic benefits* of majority voting.

> **Epistemic benefits** are satisfied when, for each possible truth x, $Pr(O = x|X = x)$ increases and converges to 1 as the number n of voters increases.

[5] This means that, conditional on the truth of the proposition (i.e., whether $X = T$ or $X = F$), the knowledge of some voters' votes provides no additional information regarding the votes of others.

[6] This assumption merely serves to simplify our calculations later but is not needed for any substantial results.

1.2.2 A Simple Model of Voter Competence

Discussions of the CJT's applicability usually focus on the independence assumption (e.g., Ladha, 1992, 1993). Although crucial for deriving the CJT, the voter competence assumption is silent on how voters make decisions and is rarely questioned.[7] In fact, voter competence is a behavioural assumption that conflates several assumptions, which is useful to tease apart, especially when considering the effect of ambiguity.[8]

To reason about agents, we typically make three assumptions: what the voter *values*, the voter's *epistemic states*, and how the voter *acts* based on their values and epistemic states. Ambiguity affects voters' epistemic states rather than directly influencing their behaviour. An agent's epistemic state involve two assumptions that are often conflated. For our purposes, it is useful to separate them. The first assumption concerns the internal structure of epistemic states. Common ways to model the internal structure of the epistemic states of *epistemically competent* agents include, for example, all-out beliefs or credence functions. In principle, we can assume that the epistemic states of our agents are well-ordered without saying anything about how their epistemic states relate to the world. Ideally, however, we also want to address how epistemic states relate to the world or, at the very least, to the evidence available to the agent. An agent whose epistemic states are well-ordered but unresponsive to the evidence around them is only competent in a very thin sense. Thus, when I refer to the epistemic competence of agents, I am referring to the combination of two assumptions: the *internal structure of epistemic states* and the *relation of epistemic states to the truth*. In the following, I will outline assumptions about *honesty* and *practical rationality* that underlie *voter competence* in cases of risk, before addressing

[7]Notable exceptions are (e.g., Austen-Smith & Banks, 1996; R. Bradley & Thompson, 2012).) .

[8]I thank an anonymous referee for suggesting this framing.

more complex scenarios involving ambiguity.

1.2.2.1 Internal Structure of Epistemic States: Credence Functions

The first assumption concerns an agent's internal epistemic representation. We assume that the epistemic states of a voter can be represented by credence functions. These are rational credences in the sense that they are between 0 and 1 and follow the probability axioms.[9] We assume that voters have a certain credence over $X = T$ and $X = F$ (i.e., $p(T)$ and $p(F)$, respectively), and because $X = F$ and $X = T$ are exhaustive and mutually exclusive, it holds that $p(T) = 1 - p(F)$.[10] For illustrative purposes, let us assume that agents receive one probabilistic private signal that determines their epistemic state over the truth of T and F (e.g., Ladha, 1992).[11] Let us trace this assumption back to our previous example.

1.2.2.2 Relation of Epistemic States to the Truth: Minimal Fidelity of Credences

We assumed that individuals have private information about the true state of the world. Yet, to avoid trivial cases of voter competence failure, we need minimal safeguards in place to ensure the quality of these signals.[12] In other words, we need to state assumptions about *how* credences relate to the truth. Formally, this means imposing restrictions on the probability distributions of

[9] I will use "credence" and "probability function" interchangeably when referring to an agent's epistemic state.

[10] Because of this relationship, we will use only $p(T)$ for the rest of the chapter.

[11] A voting outcome may be seen as the result of aggregating private information of the voters (see Feddersen & Pesendorfer, 1997).

[12] We leave open how credences come about. However, one natural interpretation is that voters base their credences on evidence that satisfies similar constraints. A very simple underlying model would be that voters adopt expert opinions that reflect the evidence. Alternatively, a slightly more complicated model might combine priors and new evidence such that the above constraints are satisfied. This can be achieved in multiple ways by imposing stricter constraints on either priors or new evidence.

credence functions. A voter who can be modelled by credence functions but is exposed to highly misleading propaganda will, trivially, not be guaranteed to be competent at voting. For example, imagine a research environment where critical research is suppressed by state action. In the most extreme case, this would mean that no research indicating global warming is available in this country ($Pr(p_i(T) > 0.5|X = T) = 0$). Now consider a less extreme example, where research on global warming is not directly suppressed but there is a severe lack of funding for public research. Additionally, private actors (e.g., oil companies) produce research aimed at supporting the verdict that global warming is less severe than it truly is. It is not the case then that some research is by chance misleading but research is systematically misleading; we have a case of biased science ($Pr(p_i(T) > 0.5|X = T) < 0.5$). To exclude such trivial cases of voter competence failure, let us formulate some assumptions about *how* credence functions relate to the truth.

Minimal Fidelity of Credences: Formally, we say that individual i holds a credence, $p_i(T) \in [0, 1]$, about the true state of the world. Individual credences are independent draws from a state-dependent distribution satisfying $Pr(p_i(T) > 0.5|X = T) > 0.5$ and $Pr(p_i(T) < 0.5|X = F) > 0.5$, where $p_i(T)$ is the credence for individual i about the truth of statement T.

Minimal Fidelity of Credences (MFC) is a restriction on the probability distribution of credences—just as voter competence is a restriction on the probability distribution of voting behaviour. If the true state of the world is T, then it is more likely that an arbitrary voter assigns a probability between 0.5 and 1, whereas if the true state is F, it is more likely that such a voter assigns a probability between 0 and 0.5 to verdict T. Let us introduce the concept of a voter *favouring* a certain verdict to mean that the voter ascribes a probability higher than 0.5 to that verdict. To summarize, MFC shields voters from systematically misleading epistemic states, such as being in an epistemic environment where certain research is censored or unlikely to be published.[13] For now, this means we assume that the assistants reading the articles can sometimes favour the incorrect verdict. That is, Gyde's assistants deal with uncertainty and may tell us that the incorrect verdict is more probable. However, they are better than a coin flip at favouring the correct verdict.

1.2.2.3 Motivation: Honesty

Next, we address what motivates voters. Without this, little can be said about voting behaviour. Voter competence would fail trivially if voters enjoyed voting against their beliefs. To prevent these failures, we introduce *honesty*. We model voters' motivation with utility assignments, ensuring symmetric payoffs for the truth of $X = T$ and $X = F$. Any voter i gets a payoff of 1 if they vote for the correct outcome and -1 if they vote for the incorrect outcome. This means that if verdict T is correct, individuals receive a payoff of 1 if they voted for T and a payoff of -1 otherwise, and vice versa for verdict F. Formally, any voter i has utilities: $U_i(T,T) = U_i(F,F) = 1$ and $U_i(T,F) = U_i(F,T) = -1$. The first argument of U_i represents their vote, and the second represents the correct verdict.[14] Hence, all individuals would like to vote for verdict T if $X = T$ and vote for verdict F if $X = F$. Thus, Gyde may assume that their research assistants care only about sending the correct answer to them, and not about whether the research assistants, as a collective, got it correct.

[13] An alternative interpretation would be that this condition shields us from trivial failures of *voter competence*, where voters consist only of anti-science agents who anti-update on scientific evidence.

[14] In Austen-Smith and Banks (1996), the first argument stands for the election outcome. However, this can lead to 'noble lies' by voters, see Bright (2017).

1.2.2.4 Practical Rationality: Expected Utility Maximiser

However, note that even these assumptions about our voters' epistemic states and motivation still do not guarantee *voter competence*. Crucially, we need to make assumptions about how voters act on their epistemic states and motivation. In line with the standard rationality assumption in the literature, we assume that an individual determines which of the two verdicts provides the higher expected utility. Thus the decision of voter i is simply the one that maximises the expected utility given by $EU_i(V_i = T) = p_i(T) \cdot U_i(T,T) + (1 - p_i(T)) \cdot U_i(T,F)$, $EU_i(V_i = F) = p_i(T) \cdot U_i(F,T) + (1 - p_i(T)) \cdot U_i(F,F)$. Any voter i will maximize their expected utility and hence choose to vote for T if $EU_i(V_i = T) > EU_i(V_i = F)$ and vote for F if $EU_i(V_i = T) < EU_i(V_i = F)$.

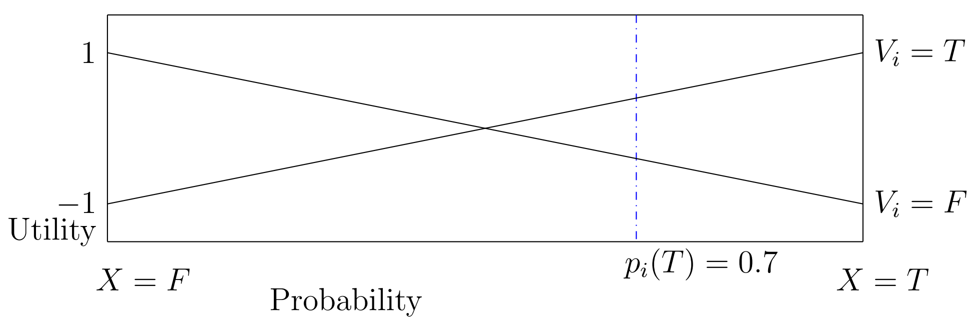

Figure 1.1: X-axis: Probability of $X = T$, left extreme $p(T) = 0$ and right extreme $p(T) = 1$. Y-axis: Expected utility of the respective actions. The blue dotted line marks the probability that voter i assigns to $X = T$ (i.e., $p_i(T) = 0.7$).

Of course, Gyde will never get the correct result with certainty, as they do not have infinite research assistants (or academic papers to draw upon). Yet, if epistemic competency, practical rationality, and honesty are satisfied, *voter competence* is satisfied (see appendix for proof). In that case, the accuracy of the voting result will improve with every research assistant (see

R. Bradley & Thompson, 2012).[15] Next, we will motivate why this model of voter competence may be too simple—descriptively or normatively—and propose a more general model.

1.3 Ambiguity and Voter Competence

Imagine you consider playing the lottery and ask your friend how likely it is to win. They reply: "I am not sure, but I am certain that it is something between 1 in a million and 1 in 10 million. Anyway, something incredibly small—definitely not worth your money!" Later, if someone asked you how likely it is to win the lottery, what would you answer?

In everyday life, the information we receive is often of the sort you received from your friend and is not necessarily suited to being reduced to a precise probability. Ellsberg (1961) used this type of uncertainty over probabilities in a series of experiments, demonstrating that participants' choices do not align with subjective expected utility theory. This type of problem has been discussed in decision theory and economics[16] under the term ambiguity (see, e.g., Barberis & Thaler, 2003; Frisch & Baron, 1988; Heath & Tversky, 1991; Gilboa & Schmeidler, 1989). The Ellsberg paradox[17] has inspired an extensive empirical literature. This literature highlights a concern in decision theory: precise probabilities fail to represent severe uncertainties correctly. They are inadequate on either descriptive or normative grounds (see, e.g.,

[15] Yet, as R. Bradley and Thompson (2012) also show, the unweighted majority rule is not necessarily the best-performing rule.

[16] Ambiguous evidence is also a subject of discussion in other sciences (e.g., see Budescu & Wallsten, 1995; Camerer & Weber, 1992; S. Bradley, Frigg, Du, & Smith, 2014; Augustin, Coolen, De Cooman, & Troffaes, 2014).

[17] In Ellsberg's paradox, agents are presented with two urns containing red and black balls: one urn has a known distribution (e.g., 50 red and 50 black), while the other urn has an unknown distribution. The participants are then presented with a series of bets, revealing an inconsistency with subjective expected utility theory that is now often interpreted as ambiguity aversion.

Kyburg Jr, 1983; Levi, 1985; R. Bradley, 2009; Seidenfeld, 2004; Joyce, 2010; Gilboa, Postlewaite, & Schmeidler, 2009). Among several concerns, one is that precise credences are not the rationally correct attitude to adopt given some types of evidence. If precise credences are not rationally required, voters may not base their decisions on them. Consequently, prior models that use precise credences to flesh out the conditions under which we can assume voter competence are not always applicable. Therefore, there is a lacuna in the literature regarding whether and when we can rely on *the epistemic benefits* of majority voting in the face of ambiguity. To close this lacuna, we should move on to models that can account for imprecision regarding the grounds on which voters presumably make a range of voting decisions. In the next section, we will expand the simple model introduced earlier. By incorporating the possibility of ambiguity, we will demonstrate why voter competence is no longer guaranteed under these assumptions.

1.3.1 A Less Simple Model of Voter Competence

1.3.1.1 Internal Structure of Epistemic States: Set of Credence Functions

Many economists and philosophers (Gilboa & Schmeidler, 1989; Levi, 1985; Joyce, 2010) argue that the most natural way to understand the agent's epistemic state in situations like Ellsberg's Paradox is their inability to determine which among a set of possible probability distributions is the true one. That is, while we know that there is one true probability distribution describing the lottery, we don't know which one it is. We now turn to how this translates into a formal model of epistemic states. When the probability distributions in the set of conceivable distributions can themselves be assigned probabilities, ambiguity in this sense can be expressed as second-order probability.

However, when the distributions cannot be assigned probabilities, ambiguity is often expressed by a set of probabilities (Camerer & Weber, 1992).[18] While not everyone in the literature agrees on what exactly is rationally required, there is some minimal agreement on what is rationally permissible. That is, it is always rationally permissible to adopt an epistemic state that covers the full range of probabilities compatible with the evidence that you have at your disposal (see Mahtani, 2019). This translates into a model where we have a set of probability distributions that cannot be reduced to second-order probabilities about the true probability distribution of the lottery. We will adopt this model of ambiguity to examine its impact on voter competence.

In our new model, voters may adopt a set of probability functions $\mathcal{P}$. Sets of probabilities are also sometimes called imprecise probabilities (e.g., Jeffrey, 1983; Joyce, 2005) or a credal set (Levi, 1974). $\mathcal{P}(x)$ is a set of numbers, e.g., any number between 0.6 and 0.9, without putting any weight on any particular number. In this case, we call their credal set convex.[19] A closed set includes its limit points (e.g., 0.6 and 0.9). A closed, convex set can formally represent what we intuitively mean by an agent adopting a *full range*. We will use this interval version of imprecise probabilities throughout this chapter. Convex sets can be represented by their extreme points. Thus, when we refer to $\mathcal{P}(x)$, we may write $[\underline{P}(x), \overline{P}(x)]$. Here, $\underline{P}(x)$ denotes the lowest probability in verdict x, while $\overline{P}(x)$ represents the highest probability in the set $\mathcal{P}(x)$.[20] Since every element of a credal set is a probability func-

[18]Note that this set may only contain one element p. Thus, this generalizes precise probabilities.

[19]A convex set is a set that includes all linear combinations of all members x, y in the set, such that $(1-w)x + wy$, with $w \in [0, 1]$.

[20]This is the simplest form of imprecise probabilities. Lower and upper probabilities cannot fully capture all the nuances of imprecise probabilities. Thus, usually, the literature refers to other frameworks, such as lower and upper previsions or lower and upper envelopes. However, for two mutually exclusive verdicts, the differences between frameworks are negligible. Thus, we simplify our analysis by sticking to the cruder notion of lower and upper probabilities.

tion, the highest probability of x equals 1 minus the lowest probability of $\neg x$ (i.e., $\overline{P}(T) = 1 - \underline{P}(F)$).[21] For example, suppose you have the epistemic state $[\underline{P}(T) = 0.6, \overline{P}(T) = 1]$. Then, your credal set contains the credence functions assigning any value between 0.6 and 1 to $X = T$ (e.g., $p(T) = 0.7$, $p(T) = 0.61$, $p(T) = 0.89$).[22]

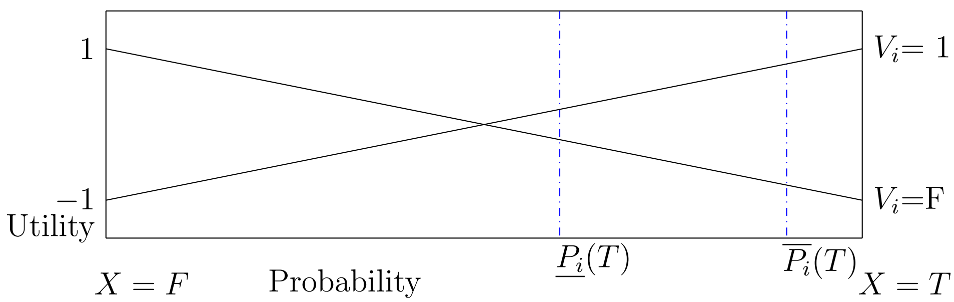

Figure 1.2: X-axis: Probability that $X = T$, with left extreme denoting probability 0 and right extreme 1. The dotted blue line indicate $\underline{P_i}(T)$ and $\overline{P_i}(T)$. Y-axis: Expected utility of the respective actions.

Let us now explicate how to spell out the conditions for how a set of credences relates to the truth.

1.3.1.2 Relation of Epistemic States to the Truth: Minimal Fidelity of Sets of Credence Functions

In the case of a single credence function, the intuition was quite straightforward. Voters have a better chance than a coin flip to favour (i.e., $p(T) > 0.5$) the true statement than to favour the incorrect statement (i.e., $p(T) < 0.5$). One can easily imagine constraints on credal set distributions that guarantee voter competence (e.g., $Pr(\underline{P}(T) = 1) = 1$). But what might be reasonable

[21] We use $\underline{P}(T)$ and $\overline{P}(T)$ as a shortcut for $\underline{P}(X = T)$ and $\overline{P}(X = T)$.

[22] Note that this implies, by the rationality constraints laid out above, that $[\underline{P}(F) = 0, \overline{P}(F) = 0.4]$.

minimal constraints on the distribution over sets of credences?

Here is one way to think about it. Consider the following scenario, which clearly violates minimal fidelity. Imagine a repressive state that wants to silence critical voices on climate change. One way minimal fidelity would surely be violated is if the state only allowed evidence entailing that it is more likely that global warming will *not* reach 1.5°C between 2030 and 2052 (i.e., $Pr(\overline{P}(T) < 0.5) = 1$) to be published. Alternatively, the state may permit non-conclusive evidence but suppress any evidence suggesting that global warming is more likely than not (i.e., $Pr(\underline{P}(T) > 0.5) = 0$). The rationale might be to avoid an epistemic environment where the only reasonable response to evidence is to favour the undesirable truth. Both scenarios surely would violate minimal fidelity. Minimal fidelity requires the possibility of evidence that *unequivocally favours* the true state (i.e., every credal function in your set favours the correct verdict, $\underline{P}(T) > 0.5$). Now imagine, as before, that there is no repressive state. However, publicly funded research is scarce, while privately funded research, influenced by special interests (e.g., oil companies), dominates. With little unbiased science, we face a troubling scenario. Whenever evidence unequivocally favours one verdict, it is more likely incorrect because private organisations selectively publish favourable results. These are cases where we don't expect voter competence to emerge. So let us exclude these cases with the following requirement on a set of credences.

Minimal Fidelity of Sets of Credences: Formally, we say that individual i holds a convex set of credences, $[\underline{P_i}(T), \overline{P_i}(T)]$, with $\underline{P_i}(T), \overline{P_i}(T) \in [0, 1]$ about the true state of the world. $\underline{P_i}(T)$ and $\overline{P_i}(T)$ are independent draws between voters satisfying $Pr(\underline{P_i}(T) > 0.5|X = T) > Pr(\overline{P_i}(T) < 0.5|X = T)$ and for $X = F$, $Pr(\underline{P_i}(T) < 0.5|X = F) > Pr(\overline{P_i}(T) > 0.5|X = F)$, where $\underline{P_{i_i}}(T)$ is the lower credence for individual i about the truth of statement T, and $\overline{P_i}(T)$ is the upper credence for individual i about the truth of statement T.[23]

One key feature of MFSC is that it generalises MFC. MFSC is satisfied if it is more probable that *every* credence function in a voter's credal set *favours* the correct verdict than that *every* credence function in their credal set *favours* the incorrect verdict. If the credal set contains only one credal function, MFSC just amounts to MFC. MFSC combines two constraints: 1. The epistemic state must allow for unequivocal favouring of the correct verdict (i.e., $Pr(\underline{P}(T) > 0.5)) \neq 0$), excluding extreme cases like a repressive state. 2. If evidence unequivocally favours one verdict (i.e., $\underline{P}(T) > 0.5$ or $\overline{P}(T) < 0.5$), it should not systematically favour the incorrect verdict, even though it may occur by chance. This excludes cases of biased science.[24]

To summarize what we have postulated so far, we no longer demand that voters ascribe a single probability to an event. However, MFSC ensures that agents' epistemic states are more likely to unequivocally favour the correct verdict than the other way around. This excludes epistemic environments (e.g., misleading propaganda, biased science) where we don't expect an agent to be epistemically competent.[25] The next section explores how rational agents make decisions using a set of credences.

[23]Note that this notion does not distinguish between X denoting a value or a proposition.

[24]Note that to satisfy this without assumption 1, a weak inequality would suffice: $Pr(\underline{P_i}(T) > 0.5|X = T) \geq Pr(\overline{P_i}(T) < 0.5|X = T)$. The strict inequality in MFSC ensures assumption 1 by ensuring that $Pr(\underline{P}(T) > 0.5) \neq 0$.

[25]If you believe MFSC does not make agents epistemically competent, it should not be surprising that voter competence fails under MFSC. However, MFSC is worth considering, as it can still guarantee the epistemic benefits of voting, even if voters' epistemic states satisfy only MFSC (see Section 5.3). In this case, the results show that the epistemic benefits of majority voting can still be guaranteed, even without epistemically competent voters.

1.3.1.3 Practical Rationality: Γ-Maximin

Imagine that you are one of the research assistants. You face the decision of advising Gyde on including one of the following statements in the briefing: (1) *Global warming will reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate* ($x = T$), or (2) *Global warming will not reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate* ($x = F$). Now, imagine reading in the report from the Intergovernmental Panel on Climate Change (IPCC) that "Global warming will *likely* reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate." Rather than assigning precise probabilities to each scenario, the IPCC presents intervals of probabilities. The IPCC uses standard intervals coded by calibrated language[26]. For example, *likely* indicates an assessed likelihood of 66%–100% (M. R. Allen et al., 2018). After reading the IPCC's explanation of *likely*, you adopt an epistemic state that can be described by $[\underline{P}(T) = 0.66, \overline{P}(T) = 1]$. As a rational agent, which principle should you follow to advise Gyde?

Unfortunately, there is no consensus on a single decision rule for rational decision-making with imprecise probabilities. However, fortunately, many of these rules converge when verdicts have symmetric payoffs. We take a decision rule as a stand-in for a whole class of decision rules that we may refer to as non-permissive (Mahtani, 2019), which typically require a single rational decision.[27] One of the most prominent decision rules (Seidenfeld, 2004; Troffaes, 2007) is called Γ-Maximin. Γ-Maximin prescribes agents to adopt pessimistic expectations (Gilboa & Schmeidler, 1989; Berger, 1985). Given a possible decision, agents assume the probability in their set that yields the lowest expected utility. Then they compare their most pessimistic

[26] For a more detailed discussion of uncertainty presentation in the IPCC, see R. Bradley, Helgeson, and Hill (2017); Dethier (2022).

[27] With symmetric payoffs, the verdict with the best expected payoff also maximizes worst-case expectations. Thus, rules based on either coincide.

expectations for each possible decision and pick the one that gives them the best-expected utility (given their pessimism). Translating this to a formal statement, we call the lowest expected payoff we achieve by picking the most pessimistic $p \in \mathcal{P}$ for a decision d $\underline{EU}_{\mathcal{P}}(d)$. We then call any decision $d \in D$ optimal if it maximizes $\underline{EU}_{\mathcal{P}}$ among all decisions in D. Thus, a Γ-Maximin decision in D is described as follows: $opt\underline{EU}_{\mathcal{P}}(D) := argmax_{d \in D}\underline{EU}_{\mathcal{P}}(d)$.

Imagine that you are such a cautious reasoner. As in the simpler model, you are still motivated by honesty, and your prior utility assignments remain in place. You consider if you should advise to include verdict T, then what is your most pessimistic expectation? Given your credal set $[0.66, 1]$, you use $p(T) = 0.66$ to calculate the expected utility of advising Gyde to include verdict T, resulting in an expected utility of 0.66. This is the lowest expected utility for verdict T that is compatible with your epistemic state. In a second step you consider whether advising to include the opposite verdict F results in the lowest expected utility compatible with your epistemic state. You use $p(T) = 1$ (i.e., $p(F) = 0$) to calculate the expected utility for verdict F, resulting in 0. In the next step, you compare 0.66 and 0 and decide to advise Gyde to include verdict T as this maximizes your expectation under a pessimistic assessment.

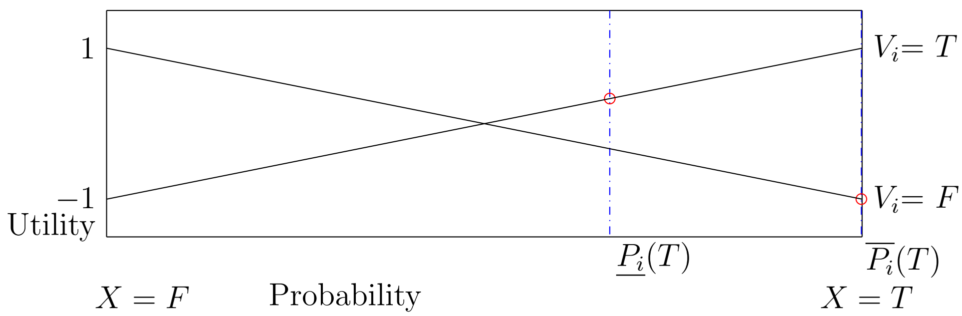

Figure 1.3: X-axis: Epistemic state of voter i of statement T, with left extreme denoting probability 0 and right extreme 1. Y-axis: Expected utility of the respective actions for voter i. Red circles indicate $\underline{EU_{i_{\mathcal{P}_i}}}(d)$ for each d (i.e., T, F).

At this point, let us stop once more and summarize the model that we built: Every voter has an epistemic state, described by $\underline{P_i}(x)$ and $\overline{P_i}(x)$ over the truth of verdict x. These epistemic states are drawn from a distribution satisfying MFSC, defined as $Pr(\underline{P_i}(x) > 0.5|X = x) > Pr(\overline{P_i}(x) < 0.5|X = x)$. Each voter then makes one of two possible decisions by adhering to Γ-Maximin, caring only to vote correctly. Next, I will show why the assumptions chosen so far do not guarantee voter competence, and thus fall short of ensuring the epistemic benefits of voting.

1.3.2 A Failure of Voter Competence

Voter competence is not guaranteed under the conditions of epistemic competence, honesty, and practical rationality, as shown by a simple counterexample. Assume that for any research assistant i of Gyde, their epistemic states are drawn from a distribution satisfying the following: $Pr(\underline{P_i}(T) > 0.5|X = T) = .45$, $Pr(\underline{P_i}(T) < 0.5 < \overline{P_i}(T)|X = T) = .38$, $Pr(\overline{P_i}(T) < 0.5|X = T) = .17$. Additionally, whenever $Pr(\underline{P_i}(T) < 0.5 < \overline{P_i}(T)|X = T)$, $\underline{P_i}(T) = 0.33$, $\overline{P_i}(T) = 0.6$. Whenever $\underline{P_i}(T) > 0.5$, Γ-Maximin will select voting for T as optimal. Similarly, if $\overline{P_i}(T) < 0.5$, Γ-Maximin selects voting for F as optimal. For $[\underline{P_i}(T) = 0.33, \overline{P_i}(T) = 0.6]$, voting for F will be the decision that gives the best-expected utility given being reasonably pessimistic i.e., maximizes the minimum EU. Thus, with a probability of 45%, a voter will vote for the correct verdict, but with a probability of 55%, they will vote for the wrong verdict, failing to satisfy voter competence.

Hence, letting the research assistants email them back which verdict to include might not be wise. One factor driving this result is that our decision rule Γ-Maximin singles out the extreme points of the credal set. However, our definition of MFSC imposes few restrictions on these extreme points. There is a mismatch in the requirements: epistemic competence and practical rationality do not necessarily translate into competent behavior under uncertainty, as they do under risk.

In the next section, we address this mismatch by spelling out practical rationality and epistemic competency differently. We then explore their effects on voter competence. Unfortunately, changing practical rationality alone is not promising. However, spelling out epistemic competency differently shows greater promise for improving voter competence. Lastly, I show that reformulating these concepts is not needed to guarantee the epistemic benefits of voting. Instead, a more practical and promising solution lies in changing the voting rule.

1.4 Possible Escape Routes

The failure of voter competence means we cannot guarantee the correct result, even as the number of voters approaches infinity. Adding more voters could even worsen the outcome. Failure of voter competence might be common, but surely we don't want to add to this predicament by expecting voter incompetence when agents are epistemically competent, practically rational, and honest.

This section explores three ways to restore the epistemic benefits of voting. The failure of voter competence arises from the combination of practical rationality and epistemic competence. More precisely, the underlying issue is a decision rule that puts weight on the extreme points of the credal set and a requirement on epistemic states that only indirectly puts some restrictions on these extreme points. The first route explores restoring voting competence by spelling out practical rationality differently. The second route examines changing the minimal fidelity of credal sets. The former will not turn out to be a promising route, while the latter has some promising results. Finally, the tension between practical rationality and epistemic competency can be resolved by changing the available actions for voters. Allowing abstention as part of the voting rule offers a promising third route.

1.4.1 Revisiting Practical Rationality: E-admissibility

Given the tension between a decision rule that focuses solely on the extreme points of the credal set and an epistemic competence requirement that imposes few restrictions on those extreme points, let us consider another contender in rational decision making with imprecise probabilities. Another class of decision rules filters the available decisions and typically singles out a set of rationally permissible decisions. We will use the decision rule e-admissibility as a stand-in for these decision rules.[28] E-admissibility demands that you maximize expected utility according to some probability function $p \in \mathcal{P}$ (Good, 1952; Levi, 1974; Seidenfeld, 2004) and thus is not specifically looking at the extreme points of the credal set. It generalizes the principle

[28]For example, decision rules such as interval dominance or maximality. For more details about these decision rules and why they are at least as permissive as e-admissibility in our example, see (Troffaes, 2007).

of maximizing expected utility but does not impose a complete ordering over decisions. Differences between e-admissibility and Γ-Maximin occur, in principle, with the type of epistemic state (i.e., [0.3, 0.66]) we utilized to show that Γ-Maximin will not result in voter competence.[29] We denote the set of $d \in D$ that maximizes expected utility according to p $opt_{EU_p}(D)$. Now we call any decision $d \in D$ e-admissible if it is in $opt_{EU_p}(D)$ for some $p \in \mathcal{P}$. Thus, all e-admissible decisions can be denoted by the following union:

$$opt_{\mathcal{P}}(D) := \bigcup_{p \in \mathcal{P}} opt_{EU_p}(D)$$

In our previous example, you, one of the research assistants, adopted the credal set $[0.66, 1]$. You now reason as follows, according to any probability function p in $\mathcal{P}$: $EU(T) > EU(F)$ and thus $opt_{\mathcal{P}}(D)$ contains T but does not contain F. Hence, in this case, you would also advise Gyde to include verdict T. However, like Γ-Maximin, e-admissibility does not guarantee voter competence, though for different reasons. E-admissibility is a permissive decision rule. In the cases where rationality alone does not prescribe a unique decision, picking any e-admissible verdict is acceptable for the voter. This can result in voters failing to select the correct verdict.

[29] E-admissibility may exclude the Γ-Maximin decision, because that decision may not maximize expected utility according to any $p \in \mathcal{P}$.

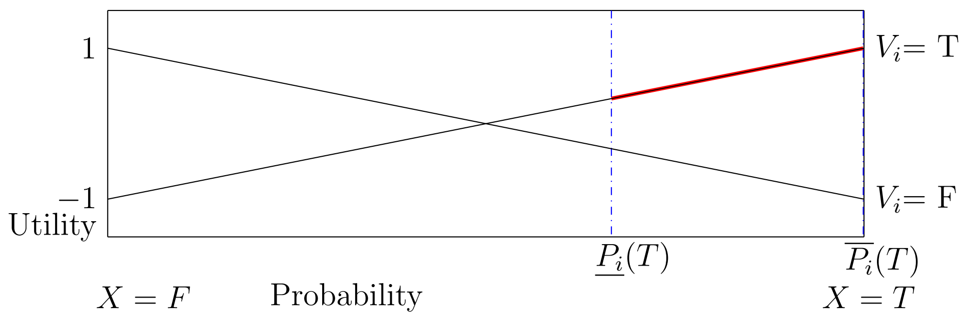

Figure 1.4: X-axis: Probability of statement T (0 to 1). The dotted lines indicate the voter's credal set. Y-axis: Expected utility of decisions; horizontal lines represent the upper and lower probabilities for T. Red lines show the decision maximizing expected utility for the corresponding probabilities.

E-admissibility fails to retrieve voter competence, as shown by the same example as Γ-Maximin.[30] E-admissibility and Γ-Maximin differ only for $[\underline{P_i}(T) = 0.33, \overline{P_i}(T) = 0.6]$. In these cases, e-admissibility allows voting for either T or F. Voters must vote for T with a probability of 45%, for F with 17%, and may vote either way with 38%. To guarantee voter competence, it must be rationally required, not merely permissible, to vote for the correct verdict with a probability over 50%.

One might object that even if different choices are permissible under e-admissibility, voters' choice processes require further explanation. Voters may have other second-order rationality criteria that impose structure on their choices. For example, one common criterion would be to be risk-averse among the rationally permissible verdicts. Yet, as seen in the previous section, risk aversion would require them to vote for F in this case, resulting in the same dilemma as with Γ-Maximin.

[30] We assumed: $Pr(\underline{P_i}(T) > 0.5|X = T) = .45$, $Pr(\underline{P_i}(T) < 0.5 < \overline{P_i}(T)|X = T) = .38$, and $Pr(\overline{P_i}(T) < 0.5|X = T) = .17$. Additionally, whenever $Pr(\underline{P_i}(T) < 0.5 < \overline{P_i}(T)|X = T)$, $\underline{P_i}(T) = 0.33$ and $\overline{P_i}(T) = 0.6$.

Another objection is that assuming mere rationality is not the most charitable approach. Typical voter behavior is relevant for applying the CJT. Voters may randomize uniformly if rationality does not dictate a choice. Imagine choosing between two equally good bottles of water at the supermarket; you consistently pick the one at eye level. The arrangement influences your choice without changing your preferences. Display height is irrelevant if one bottle is sparkling and the other still. Similarly, we might say that if you are undecided between policies you are more likely to choose whichever verdict is listed first on the ballot. If, for instance, the status quo verdict is always listed first, voters may vote for the status quo verdict more often. If the status quo is often wrong (e.g., based on outdated science), this creates a non-random selection of verdicts without requiring second-order criteria or irrationality. Establishing that voters randomize over rationally permissible actions is not trivial. Hence, changing the decision rule will not affect voter competence, suggesting a need to reconsider assumptions about MFSC.

1.4.2 Revisiting the Relation of Epistemic States to the Truth: Symmetric Minimal Fidelity of Sets of Credence Functions

Since altering practical rationality is unpromising, let us turn to epistemic competency. One might argue that our requirement on credal sets is too weak to qualify as minimal. MFSC ignores the asymmetry in the spread of possible verdict values (e.g., $\underline{P}(T) = 0.51$ vs. $\underline{P}(T) = 0.9$). To address this, we could consider how strongly a verdict is supported. This means imposing restrictions directly on the extreme points of the credal set. What minimal restrictions on credal sets respect the epistemic importance of these extreme points?

Imagine again the repressive state. They aim to create an environment where the strongest credences compatible with the epistemic state deny climate change. That is, the state censors evidence where the strongest credence compatible with the evidence favours the undesirable truth (i.e., $Pr(\underline{P}(T) > 1 - \overline{P}(T)|X = T) = 0$). Now imagine that instead of the repressive state, we have predominantly private companies conducting research with an interest in evidence where the strongest signals favour the incorrect verdict. Science is again biased, systematically producing evidence where the strongest signals favour the incorrect verdict ($Pr(\underline{P}(T) > 1 - \overline{P}(T)|X = T) < Pr(\underline{P}(T) \leq 1 - \overline{P}(T)|X = T)$). This suggests another intuition about what constitutes minimal fidelity in the case of credal sets. Let us exclude these cases.

Symmetric Minimal Fidelity of Sets of Credence Functions (SMFSC): $Pr(\underline{P}(T) > 1 - \overline{P}(T)|X = T) > Pr(\underline{P}(T) \leq 1 - \overline{P}(T)|X = T)$ (and similarly for $X = F$). The Symmetric Minimal Fidelity of Sets of Credence Functions (SMFSC) captures the intuition that the correct verdict is more often associated with the highest probability than the incorrect one. SMFSC ensures voter competence for Γ-Maximin but not for e-admissibility (see proofs in the appendix). The reasons for this are outlined below.

1.4.2.1 Voter Competence under SMFSC with Γ-Maximin

SMFSC is not merely a condition under which Γ-Maximin guarantees voter competence. SMFSC is the minimal epistemic condition that you, as a cautious reasoner with credal sets, need to fulfill to ensure competent voting. The intuition is straightforward. When outcomes are symmetrical in severity (i.e., being correct or incorrect is equally good or bad for both verdicts), Γ-Maximin selects the verdict with the highest credence compatible with *favouring* it. By SMFSC, it is guaranteed that for any voter i, $Pr(\underline{P_i}(T) > 1 - \overline{P_i}(T)) > 0.5$. By Γ-Maximin, it is guaranteed that for any voter i, $Pr(\underline{EU}_{i_{\mathcal{P}_i}}(T) > \underline{EU}_{i_{\mathcal{P}_i}}(F)) > 0.5$. Thus, it follows that $Pr(V_i = T) > 0.5$ for voter i. Thus, voters are guaranteed to be competent.

This is good news. This means that rethinking our requirements on credal sets ensures ambiguity does not hinder the epistemic benefits of voting. We can advise Gyde to pursue their approach to the briefing. However, if rationality permits permissive decision rules like e-admissibility, voters might not be competent under SMFSC (see proof in the appendix).

1.4.2.2 Failure of Voter Competence under SMFSC with E-admissibility

Roughly, the failure of voter competence is driven by the fact that e-admissibility does not take into account how strongly some probability p favours one verdict, but merely which verdicts are favoured by some p in your credal set. In fact, the minimal epistemic requirement under e-admissibility for guaranteeing voter competence would be that $Pr(\underline{P}(1) > 0.5) > 0.5$ (see full proof in the appendix). SMFSC does not guarantee that epistemic states ever unequivocally favour one verdict,[31] yet this is required to single out one rationally required action under e-admissibility. However, combining e-admissibility as a first-order criterion with Γ-Maximin as a second-order criterion restores voter competence. Although e-admissibility can, in principle, exclude actions chosen by Γ-Maximin, this does not occur in our scenarios. When a credal set unequivocally favours a verdict, e-admissibility and Γ-Maximin agree. Since e-admissibility does not exclude any actions at $p(T) = 0.5$, both actions are optimal. Thus, the action selected by Γ-Maximin as a first-order criterion remains unchanged when applied as a second-order criterion after e-admissibility.

There is, however, a third route to resolve the tension between epistemic competency and practical rationality. Next, we will expand our model to

[31] The same counterexample from the previous section demonstrates this.

account for the possibility that voters may abstain. If voters can abstain, it becomes possible to guarantee that voting is epistemically beneficial with MFSC and Γ-Maximin.

1.4.3 Revisiting Voter Competence and Majority Rule

Let us consider voter competence differently from how it is defined in the CJT. Instead, let us focus on whether a voter is more likely to vote for the correct verdict than for the incorrect verdict. Let us call this concept *generalized voter competence*. With only two ballot choices, both definitions of voter competence coincide. However, they diverge if abstention is included. When determining the outcome of an election, abstentions are disregarded. The outcome O is $O = T$ if there are more voters with $V_i = T$ than $V_i = F$, $O = F$ if there are more voters with $V_i = F$ than $V_i = T$, and $O = \{\}$ in the event of a tie.

Formally, the difference is as follows. Given that $X = T$, generalized voter competence holds that $Pr(V_i = T|X = T) > Pr(V_i = F|X = T)$, whereas voter competence holds that $Pr(V_i = T|X = T) = c$ with $c > 0.5$. The rationale for changing the voting rule is that voting is a form of expressing attitudes. If a voting rule allows a more fine-grained expression of attitudes, a voter is not forced to guess incorrectly. In other words, this approach reduces information loss and distortion, allowing voters to express ambiguity to some degree. However, to conclude that rational voters will use this opportunity in a way that will guarantee the epistemic benefits of voting, more assumptions are needed. A voting rule that allows more nuanced expression does not guarantee voters will use it. Let us begin by introducing a third verdict to the decision space, D. D now includes the verdict to abstain, denoted by A. In the second step, we need to decide what utility a voter may get from abstaining. Surely, voting for the correct verdict should be better than avoiding a wrong choice. Furthermore, avoiding a wrong choice should be better than making one. Thus, we are left with a modeling choice between $-1 < U_i(V_i = A|X = x) < 1$. Let us fix the utility to 0 (i.e., $U_i(V_i = A|X = x) = 0$).

1.4.3.1 General Voter Competence under MFSC with Γ-Maximin

Abstention, as outlined above, does guarantee the epistemic benefits of voting (see full proof in the appendix). There are two key features that drive this result. The utility of abstaining is not outcome-dependent, so its expected utility is independent of the credence function. Second, the expected utility is always 0. This means that whenever the credal set contains the credence $p(T) = 0.5$, the extreme points of the credal set become irrelevant for decision making. Γ-Maximin will always pick to abstain when a credal set does not unequivocally favour one verdict over the other. Yet, since MFSC requires that there is some probability of epistemic states unequivocally favouring one verdict, voters will not always abstain. Additionally, from MFSC, it also follows that we expect voters to vote for the correct verdict when they do not abstain. If voters face severe uncertainty, their method of expressing epistemic states needs to be more fine-grained. It is somewhat surprising that having merely three possible ways to express your epistemic states (V=T, V=F, V=A) suffices for us to retrieve the epistemic benefits of voting. Given a certain decision rule, the epistemic requirements on the credal sets can be lessened by changing the voting rule.

1.4.3.2 Failure of General Voter Competence under MFSC with E-admissibility

Abstaining will not guarantee general voter competence if voters can freely choose between e-admissible actions. The reason for this is quite simple: e-admissibility excludes actions that do not maximize expected utility under any credal function in your credal set. However, if the credal set contains $p(T) = 0.5$, voting for verdict T, verdict F, or abstaining is rationally permissible. Hence, the same counterexample from the previous sections illustrates the failure of general voter competence.[32] Adding Γ-Maximin as a second-order criterion under SMFSC would also retrieve general voter competence. This means that in the example of our research assistants, even if only MFSC is satisfied, Gyde can divide the work among their research assistants. It might be best to allow their research assistants to email that their evidence is insufficient for advice. If Gyde knows their research assistants are cautious reasoners, this procedure guarantees the benefits of voting. If not, it at least avoids worsening the outcome. While Gyde cannot control the epistemic environment, they can adjust the voting procedure to improve the chances of a good epistemic outcome.

1.5 Conclusion

In many vital areas of collective decision-making, we face severe uncertainty and often rely on ambiguous evidence (e.g., climate science). As a result, rational agents' epistemic states reflect the ambiguity of the evidence in such decision-making scenarios. Unfortunately, our assessments of collective decision-making procedures seldom do the reality justice.

This chapter demonstrated a mismatch between scenarios where a simple majority vote seems to guarantee epistemic benefits and when the conditions for such a guarantee actually hold. More precisely, I showed that in contrast to situations of risk, voter competence is not guaranteed in situations of ambiguity, even if we assume practical rationality, honesty, and epistemic competency. The results depend on how practical rationality, honesty, and epistemic competency are defined together. Whether voting competence holds depends on the interplay of all three assumptions. The first conclusion of this chapter is that considering different kinds of uncertainty significantly

[32] To check that this allows for a failure of general voter competence is left to the reader.

affects the conditions required for achieving voting competence.

The second conclusion is that collective decision-making rules that perform well in situations of risk might not apply to decision-making under ambiguous prospects. If ambiguity is the norm rather than the exception, we must reconsider and revise the rules for collective decision-making that perform well epistemically. As a start to this research project, I proposed extending the ballot choices and showed that including abstention can retrieve similar results to the CJT. The intuition behind this proposal was that procedures designed without ambiguity in mind force more precision than is warranted. The novel aspect of this result was not that abstention allows for a more fine-grained description of voters' attitudes but how rationality guarantees its use.

2

On Equality, Power and Strategic Voting

Abstract

Strategic voting is an odd phenomenon. On the one hand, there is widespread agreement that we should seek to prevent it; on the other hand, there is scarcely any work on why we should regard strategic voting as bad. This chapter takes early concerns in social choice theory seriously that strategic voting turns voting into a game of skill. It provides one possible grounding of this concern with the help of contemporary normative theory by arguing that strategic voting can undermine the equal standing between citizens. The core of the chapter challenges the assumption that "one person one vote" is the only property of our voting rules required to ensure equal power relations. Through exploring various approaches to power in voting, the chapter illustrates the potential effects that strategic voting can have on power relations among voters. Consequently normative theories that are concerned with such power imbalances (e.g., relational egalitarianism) ought to weigh in on strategic voting and on the details of the design and implementation of voting rules.

2.1 Introduction

Strategic voting has a bad rep, but public sources and even academic ones scarcely attempt to justify the bad rep and just occasionally mention reasons for that rep in passing (for a notable exception see Dowding & Van Hees, 2008). This is especially surprising as there is a lot of ink spilled in social choice theory on the differences between voting rules regarding their vulnerability to strategic voting (for example Brams & Fishburn, 1978; Saari, 1990; Bartholdi III & Orlin, 1991). With social choice theory being increasingly picked up in algorithmic decision making and playing a role in AI design (e.g., Zuckerman, Procaccia, & Rosenschein, 2009; Boutilier et al., 2012; Betzler, Slinko, & Uhlmann, 2013; Shrestha & Yang, 2019) old questions about how to design voting rules become pressing again. And while there are many new approaches addressing how to decrease if not stop strategic voting (e.g., Elkind & Lipmaa, 2005; Filos-Ratsikas & Miltersen, 2014; Chakrabarty & Swamy, 2014; Slinko, 2004) the normative justifications for such designs remain underdeveloped. What is more, the few philosophical works that engage with the normative status of strategic voting see strategic voting as normatively acceptable if not desirable (Dowding & Van Hees, 2008; Lehtinen, 2015).

In this chapter, I take early concerns in social choice theory seriously that strategic voting turns voting into a game of skill (Dodgson 1876, reprinted in Black 1958, for a similar concern see Satterthwaite 1973). I provide one possible grounding of this concern with the help of contemporary normative theory by arguing that strategic voting can lead to power imbalances between citizens. I argue that normative theories that are concerned with such imbalances (e.g., relational egalitarianism), ought to weigh in on strategic voting. More precisely, if relational egalitarians (e.g., Kolodny, 2014b, 2014a) argue for the principle “one person one vote” they also have a pro tanto reason to endorsing voting rules that reduce if not prevent strategic voting.

The chapter is structured as follows. First, I give a short overview of the theory of strategic voting and the early concerns voiced by Dodgson (1876/1958) and Satterthwaite (1973) regarding turning voting into a game of skill. Second, I will motivate why the influence of difference in skill on power relations among citizens is normatively relevant. For this, I will sketch relational egalitarianism as one justification for democratic procedures and point out where equal power through the formal democratic procedure is central for its justification. Here, I will especially rely on Kolodny's (2014a, 2014b) argument for the principle of "one person one vote". Third, I will argue that the possibility of strategic voting puts these arguments in question by showing how "one person one vote" does not guarantee equal power relations among citizens via the voting rule. For this I will explore different approaches to power in the literature and show the implications of these different understandings of power on the harm that comes with the possibility of strategic voting. Lastly, I conclude that there are grounds to worry about the equal standing of citizens if strategic voting is possible.

2.2 Strategy-proofness in Social Choice Theory

When social choice theorists talk about strategic voting, they are using a technical term. Strategic voting then refers to the misrepresentation of your preferences to achieve an outcome that you actually prefer. We follow this literature by adopting the standardly assumed minimal rationality requirements implying that voters have an ordinal ranking over all candidates, and these rankings are complete and transitive.[1]

[1]There is also a game-theoretical literature that assumes cardinal utilities and utility maximizing behavioural (e.g., Myerson & Weber, 1993) and thus there are other incentive compatible properties one might be interested in. However, for simplicity we will stick with

For an example of strategic voting, let us imagine a committee consisting of you and me and three policies that we vote over: A *progressive policy* (p), a *conservative policy* (c) and the *status quo policy* (s). Let's say that you and I will vote via a simple majority vote among these policies. That is, both of us can vote for one policy and the policy with the most votes will be implemented. In the event of a tie the option will be broken according to a lexicographical rule (i.e., the policy that comes first in alphabetical order is elected). An instance of strategic voting in the technical sense of social choice theory would then be the following: Imagine that I am a progressive and rank the policies $p \succ s \succ c$, whereas you dislike any change and thus rank the policies $s \succ c \succ p$. Given that I vote sincerely (i.e., vote for my first-ranked candidate) I will vote for the progressive policy. Then you, however, will do better by giving your vote not for the status quo policy but for the conservative policy instead. This would be the case, because in a tie of the progressive policy with the status quo policy, the lexicographical rule will break in favour of the progressive policy. However, in a tie with the conservative policy and the progressive policy the lexicographical tie breaker rule favours the conservative policy. If you voted for the conservative policy in this case, you voted strategically. If a voting rule allows for some preference profiles (i.e., for some combination of preferences[2]) where voters like you can profit from strategic voting, we call the voting rule not strategy-proof

the notion of ordinal preferences and strategy-proofness so we can focus our attention to the normative analysis.

[2]More specifically, strategic voting is defined for social choice rules. A social choice rule is a function that assigns a specific candidate as the outcome for every combination of preference orderings (one for each individual), where each individual expresses their preferences over a set of candidates. The Gibbard-Satterthwaite result, establishes that any social choice rule over more than two candidates, that satisfies a few other minimal desiderata, cannot be strategy-proof. In other words, it demonstrates that it is not always advantageous for individuals to reveal their true preferences because there will always exist a preference profile where at least one person has an incentive to deviate from their true preferences.

(Gibbard, 1973; Satterthwaite, 1973) Voting rules typically allow for some possibility of strategic voting and thus are not strategy-proof.

Strategic voting is, however, not just a technical term in the mathematical study of voting. Strategic voting as laid out above serves as the theoretical foundation for empirical work (see for example, Fisher, 2004) and is frequently observed. For example, in one of the primary locations for the empirical study of strategic voting, the UK (Aldrich, Blais, & Stevenson, 2018, p.31), Alvarez, Boehmke, and Nagler (2006) estimate that around 50% of all voters will, given the opportunity, vote strategically.[3][4] Major newspapers in the UK reflect and inform readers about strategic voting (e.g., Drewett & Richards, 2021; Herbert, 2019; Elgot, 2022) and occasionally give explicit advice on how to vote strategically (e.g., Davies, 2019; Times, 2015; Clark, 2019)[5]. In the US, campaigns against current voting rules stress that the current rules encourages strategic voting and alternative voting rules would lessen this predicament.[6] This evaluation is not new, Borda (1784/1958) famously exclaimed: "My scheme is intended only for honest men" (quoted in Black, 1958, p.215) realising that the Borda voting rule allows for preference profiles that incentives strategic voting.

And yet while a lot of effort is spent to find the relation between voting rules and the incentivisation of strategic voting, very few have explored the normative status of strategic voting (see Dowding & Van Hees, 2008). This is especially surprising as social choice theory is full of impossibility results

[3]Measuring strategic voting is a complex task and in practice methodologies and definitions of strategic voting differ slightly. Yet, that strategic voting of some for happens is a widespread finding across the literature.

[4]Of course, strategic voting is not just a UK issue but has been observed in many democratic elections (see Aldrich et al., 2018, for data on Israel, Germany, Japan, Belgium, Switzerland and Canada).

[5]The UK has a history of strategic voting campaigns give where you can get tailored advice how to vote strategically in your district (e.g., https://tacticalvote.co.uk/).

[6]E.g., https://fairvote.org/our-reforms/ranked-choice-voting/.

that warrant trade-offs between properties that voting rules can satisfy. A systematic procedure for making such a trade-off requires a good grip of the normative foundations underlying the desiderata employed in social choice theory. And yet again, literature that goes beyond noting a reason in passing for strategy-proofness is hard to come by. Consider the diverse applications of social choice theory and strategy-proofness, ranging from political elections to recommender systems, hiring committees and allocation problems like school or organ assignments. It appears more reasonable than not to adopt a pluralistic approach when considering the justifications for implementing strategy-proofness. Hence, the objective of this chapter is not to assert a singular inherent normative value of strategy-proofness. Instead, this chapter focuses on one reason of strategic voting mentioned in the literature[7], that strategic voting turns voting into a game of skill. The objective of the chapter is to develop this reason further and to clarify which assumptions are needed to hold such that it is in fact a normative concern. One of the earliest concerns that explicitly refers to turning voting into a game of skill is from Charles Dodgson who wrote:

> "[A] principle of voting which makes an election more of a game of skill than a real test of the wishes of the electors, and [...] my opinion is that it is better for elections to be decided according to the wish of the majority than of those who happen to have the most skill at the game." - (Dodgson quoted in Black 1958, p.265)."

Satterthwaite (1973) also seemed to have a similar concern, when he wrote:

> "Committee members [...] will in general be unequally skilled in the employment of sophisticated strategies. Those members who are more skilled will, in effect, have a greater weight on the

[7]Other reasons mentioned in the literature include among others honesty (Monkovic, 2016) and non-transparency (e.g., Elkind & Lipmaa, 2005).

> committee's decision than those members who are less skilled. This inequality of weighting may contradict whatever principle of representation – such as one man one vote – that may have guided selection of the committee."

Whereas Dodgson and Satterthwaite both assumed that voters are typically unequally skilled and thus voting would turns a *game of skill*, they didn't provide substantial normative reasons for their concern. In the rest of the chapter, I will spell out one possible grounding for these concerns that rests on strategy-proofness protecting the equal power of citizens over each other via their voting rule.

For our full normative argument, we need to establish first a) a normative claim specifying what is harmful about making voting a game of skill and b) a descriptive claim that shows how strategic voting will induce the harm specified in a). We will start with a) in the next section and describe what could be harmful if it were the case that greater skill translates into effecting the outcome to a greater degree.

More precisely, our strategy is the following: we will follow Satterthwaite in taking the principle of "one person one vote" as a starting point. First, we will spell out how one may defend this principle according to one normative theory, i.e., Niko Kolodny's (2014a, 2014b) version of relational egalitarianism, where this principle arises from the concern for equal standing between citizens and consequently power balances that are an integral part of determining the standing between persons. In a second step, we will show how strategic voting can systematically interfere with power balances even though everyone has the same number of votes. We will then conclude that under this normative framework a considerable harm of allowing strategic voting is that, when voters have unequal skill in casting strategic votes, power between voters becomes unequal and thus the standing between them as citizens is at least potentially in jeopardy. Thus, turning voting into a game of skill is threatening the equal standing of voters, a principle that many are already committed to by endorsing the principle of "one person one vote".

Establishing this serves two purposes. First, it can serve to guide trade-offs between desiderata in social choice theory. One way to decide whether (and how) to give up strategy-proofness or other desiderata such as dictatorship or monotonicity[8] is to evaluate their violation on a common value such as power. Second it shows that normative theory has an important place to play to determine details about aggregation procedures. This chapter shows that once we adopt a non-ideal assumption of a *difference in skill* relational egalitarianists have grounds to worry about strategic voting.[9]

2.3 Relational Egalitarianism and One Person One Vote

In this section we are going to spell out one possible justification for "one person one vote"; the idea that everyone's voice should be weighed equally. Our justification rests upon the claim that equal power over each other is an important aspect between citizens that requires the implementation of voting to organise common aspects of living together. We will use this later in the chapter to show that if you are obligated to "one person one vote" in virtue of power considerations the very same reason commits you to worry about

[8] Dictatorship means that there can be one voter i whose (strict) preferences always prevail in the election outcome. That is, it must be in principle possible that the voting rule can decide against voter i's preferences. Monotonicity means that a vote for a certain candidate should never be of disadvantage for that candidate i.e., a voter should not be able to hurt a candidate by voting for them (or ranking them higher).

[9] Note however that relational egalitarianism is arguably not the only normative theory whose scope includes strategy-proofness (see for e.g., Lehtinen, 2008, for a connection between welfare and strategic voting and Austen-Smith and Banks, 1996 for a connection between epistemic properties of voting outcomes and strategic voting) or other desiderata in social choice theory.

strategic voting as well. In other words, turning voting into a game of skill may undermine the very same reason why voting is a legitimate procedure in the first place.

Imagine that you were to discover that in the last election my vote counted twice as much as yours; you may not be amused. After all, "one person one vote"[10] is one of the fundamental principles of modern democracies. Yet, while it may seem non-negotiable to many nowadays, not everyone would subscribe to this principle. For one, consider Plato's argument that democracy undermines the appropriate deference to expertise necessary for the proper governance of societies (Plato, 2003, Book VI). John Stuart Mill also defended a form of epistocracy, sometimes referred to as the "plural voting" scheme (see Christiano & Bajaj, 2022). Both Plato and Mill argue against the principle "one person one vote" on epistemic grounds. "one person one vote" is thus not a trivial assumption, it needs to be grounded in a broader normative framework.

One way to do that emerges from reflection on why many nowadays would not be swayed by talk of epistemic harm induced by the principle of "one person one vote". They may believe in an intrinsic or procedural value of democracy instead of a purely instrumental one (e.g., its usefulness for achieving epistemic ends). That is, you hold that voting is not just a means to produce some outcomes but is valuable in itself. You could argue that giving me more votes than you is a violation of this value.

You could construct a robust argument for this claim grounded in the concept of relational equality (Anderson, 1999; Christiano, 2008; González-Ricoy & Queralt, 2018; Kolodny, 2014b, 2014a; Viehoff, 2014; Wilson, 2019). Rela-

[10]Note that the term "one person one vote" refers to the weight of votes, not to how votes are expressed. Thus, a voting rule that gives everyone two votes would satisfy this principle.

tional equality is not the only value that generates the claim that our voting rules ought to be designed in such a way that it gives every citizen via the voting rule equal weight (i.e. "one person one vote"). Some others appeal to liberty (e.g., Gould, 1990), domination (e.g. Pettit, 1997), normative authority and rational independence (e.g., Cordelli, 2020) or even make the argument on epistemic grounds (Goodin & Spiekermann, 2018). However, for the sake of providing one possible normative grounding of the concerns of Dodgson or Satterthwaite, let's make the argument in an exemplary way on relational egalitarian grounds.

Let us focus on Kolodny's version of relational egalitarianism (2014a, 2014b) to make our argument more precise. Kolodny, like other relational egalitarians, doesn't focus on outcomes but takes how we relate to each other as fundamental. For example, Kolodny (2014a, 2014b) very roughly argues that certain power, consideration, and authority differences can constitute relationships that we deem as highly problematic and bad in themselves, since they treat one party as inferior and the other party as superior. What is objectionable about such arrangements is not merely their instrumental consequences, or the fact that those deemed inferior are treated in ways that are anyway problematic quite apart from the fact that others are superior. For example, that some are treated unfairly is bad in itself. But even in the absence of consequences that we regard as bad, the very fact that some are above and others below is bad in its own right. Given that we deem such relationships as bad, and power imbalances[11] are at the heart of such relationships, at the very least our formal procedure should not give some more power over the rest (see Kolodny 2014b). Moreover, since political decisions are binding and can be expected to influence people lives, power over political outcomes is always also power over others and thus contributes to

[11]Kolodny is not the only relational egalitarian who takes power balances as central for equal standing (see Viehoff, 2014, 2019; Scheffler, 2015; Anderson, 1999; Christiano, 2008).

people's unequal status in society.[12]

The principle "one person one vote" is supposed to ensure that every citizen can at least in principle influence political decisions to the same degree as their employer, their neighbour, or any other member in the society and thus everyone has the same power over each other via the voting rule (see Kolodny 2014b). There would be something especially bad about endorsing and nourishing power inequalities by the structures and institutions used to organise ourselves. In this sense, "one person one vote" is a critical mechanism for preventing any one person from having an inappropriate degree of control over others. Voting mechanisms are often fundamentally relational, in that they shape the ways in which individuals are able to relate to one another as equals.

If you were to discover that my vote is counted twice, or generally everyone whose name starts with the letter A or everyone who read Kafka's Trial is counted twice, you may complain in the following way: The institutions that govern us, give you more power over me than they give me power over you. Having additional weight by principle treats you as above me and thus the equal standing between us has been violated.

In a next step we want to focus on how strategic voting can disturb power balances between citizens and thus show that strategic voting potentially inflicts the harm of creating unequal standing between citizens. In other words, I argue that if you are committed to "one person one vote" in virtue of power considerations (may they be based on relational egalitarian concerns or any other concerns that single out power balances as a central property) the very same reason commits you to worry about strategic voting as well.

[12]For an overview over power to and power over see A. Allen (2022).

2.4 Power and Voting

In this part of the chapter, we will focus on the second part of establishing our normative argument. That is we are set out to explore the connections between power and strategic voting and thus how the harm specified in the prior part may come about. This won't be a trivial task: in traditional methods of analysing voting power in social choice theory, strategic voting is assumed away. Hence, the contribution in this part of the chapter is two-fold. First, we engage with social choice theory's approach to voting power and highlight which descriptive idealisations affect engagement with normative theories, especially with Kolodny's version of relational egalitarianism. Accordingly, the first point in this section is about voting power measures, their interpretations but also their limitations for normative theorising.

The second objective of this section is to determine how strategic voting and power are interconnected, and to specify which aspects of the voting procedure contribute to power imbalances. To this end, we will consider two different conceptions of voting power. Firstly, we will examine the established understanding of voting power: a priori voting power indices. Secondly, we explore another perspective on voting power within social choice theory: preference-based voting indices. We will demonstrate that strategic voting induces power imbalances under both conceptions and explore whether strategy-proofness can mitigate the power differences. It is important to note that while strategic voting induces power imbalances across both conceptions of voting power, the conditions and reasons for these imbalances differ which in turn has implications when strategy-proofness is a desirable property of voting rules.

2.4.1 Voting Power and One Person One Vote

Voting power is standardly measured in social choice theory by so-called a priori voting indices (e.g., the Penrose-Banzhaf[13] index or the Shapley-Shubik[14] index). These power indices are important for us to consider for two reasons. First, they are developed within a subfield of social choice theory, thus providing a natural resource for social choice theorists to determine trade-offs between desiderata in terms of power considerations. Second, they also are regularly cited in philosophy, political science and even law to evaluate voting rules, see for example, Morriss (2002, p.184) and Kolodny (2014a)[15][16] thereby already influencing how a range of researchers think about voting rules in terms of power relations. For instance, Kolodny (2014b) explicitly relies on these indices to argue for the permissiveness of the requirements on formal voting rules.[17]

[13]To calculate the scores for a voter according to the Penrose-Banzhaf index, it's necessary to evaluate all potential minimum winning coalitions. These are groups of voters just large enough to elect a candidate (i.e., the coalition would fail if even one member left the coalition). The Penrose-Banzhaf score is determined by counting the instances where that voter's withdrawal could change a coalition from winning to losing. To normalise these scores so that their sum equals one, we first sum the individual scores and then divide each score by this total sum.

[14]To determine each voter's power using the Shapley-Shubik index under specific voting rules, we examine every conceivable sequence in which all voters could cast their votes. For each sequence, we pinpoint the voter whose vote is decisive in ensuring a candidate's election. This voter is termed pivotal in that specific sequence. The power of a voter is calculated as the proportion of all possible voting sequences in which they play this pivotal role.

[15]For an argument against the applicability in a particular case see Potthoff and Brams Potthoff and Brams 1998)

[16]A priori voting power indices conceive of voting power in terms of agents' ability to determine or control outcomes (Felsenthal & Machover, 1998, p.2 also p.35). Within this context, voting power is construed as the influence one wields purely by virtue of the "formal institutional structure of the decision-making procedure." This structure encompasses a set of candidates, the voters, action profiles, and a voting rule, which transmutes the actions of agents into a collective outcome—such as an election result (Felsenthal and Machover, 1998, pp.11-12).

[17]Kolodny (2014b) states that his definition of contributory influence, the kind of influence relevant to social equality is in line with the Shapley-Shubik index. However, as

To get a rough idea of these indices, let me lay out some features of the Penrose-Banzhaf index (many but not all are shared by the Shapley-Shubik index) and the typical justification of these features. These voting indices gauge the relative frequency with which you become a difference maker. In other words, they zero in on instances where you might alter the outcome from what it would have been, keeping fixed how everyone else voted. The difference maker concept aims to encapsulate your opportunities to affect the outcome, regardless of whether your vote does, in fact, enact change. Moreover, this concept underscores your potential to influence others' behaviour. By determining the outcome, you compel other voters to act in a manner they might not have otherwise (usually exactly half of the electorate). We can already see why these indices would be of interest for relational egalitarians – they aim to formalize in voting rules the opportunities to influence other people. They are a formalization of having power over others.

Determining which opportunities are available, however, necessitates making some assumptions. Because these indices aim to determine your voting power via the voting rule, they try to abstract away from any particular election. Instead, they presume that the probability of any voter casting a vote for a given candidate is equal to the probability they vote for any other candidate and also assume voting independence (i.e., my votes cannot help you to predict someone else's vote). The justification typically found in the literature for these assumptions is that in order to measure power solely in virtue of the formal institutional rules one requires abstracting from the distribution of preferences. Once such information is excluded, equiprobability are supposed to be justified by the principle of insufficient reason (see Garrett & Tsebelis, 1999).

noted by Abizadeh (2021), his argument is actually is referring to the Penrose-Banzhaf index.

It would go beyond the scope of this chapter to examine voting power indices comprehensively (for a comprehensive introduction see Felsenthal and Machover 1998). Nonetheless, a cursory understanding can be gained by revisiting our initial example. We again have an election of the three policies: the progressive policy, the conservative policy and the status quo policy. Our actual preferences in the example are irrelevant to determine voting power according to these a priori indices. To determine whether we have equal power over the outcome (and hence over each other) we look at all possible elections. In total, there are 36 possible elections, derived from the six distinct (strict) preference orderings that each of us can have, and the various combinations thereof. Under each of these scenarios we now analyse where you could change the outcome. As it turns out, you are always in a position to change the outcome when I vote either for the status quo policy or the progressive policy. That is, in two-thirds of the cases you are a difference maker. If we repeat this analysis, we see that I am also a difference maker whenever you voted for the progressive policy or the status quo policy. Thus, we both enjoy the same power over the outcome and over each other. The standard interpretation of such an analysis is that voting power is represented through the probability that a voter will be decisive (Felsenthal and Machover 1998, pp.37–38; Penrose, 1946, p.53).

Let us change the voting rule in our illustrative example. We transition from a majoritarian rule, wherein every vote is counted equally, to one whereby my vote counts twice and yours, merely once. Now, I do not merely possess more power; rather, I wield absolute power, effectively rendering me a dictator. There exists no vote you might cast that I could not overturn with mine. To make the example less extreme, let's add another voter. Suppose both you and this new voter are allocated one vote, while I retain my two. Now, contemplating all potential elections, there exist several where, had you voted differently, the outcome would have been different. For instance, were the other voter and I to be divided – they opting for the conservative policy and I for the progressive – the electoral outcome hinges upon your vote. Consequently, you possess some voting power. Nevertheless, the number of elections where the outcome would have depended upon my vote is greater. For instance, I can also be decisive if you and the other voter align on a policy. Thus, an inequality persists: I more frequently emerge as a difference maker and, therefore, possess greater voting power than you.

This approach to voting power already permits us to cogently address your previous grievance of me having twice as many votes (see Kolodny 2014b): namely that giving me more votes than you inherently subverts our equal status as citizens. You having more votes than me puts you in a situation where you can make me do more things against my will than I can make you do things against your will. Namely there are more possible elections where you are a difference maker, giving you more power over me than I have power over you: this violates our equal standing!

2.4.2 Inequality of Skill, Strategic Voting and Power

We have seen in the previous section that voting power indices substantiate the claim that "one person one vote" is needed to assure equal power over the voting outcome. In this section we lay the groundwork that "one person one vote" is not sufficient to guarantee equal voting power. Strategic voting can disturb power relations in a similarly structural fashion like a violation of "one person one vote" and thus can undermine the equal standing of voters. Before we make this argument in full though, we need to walk through another descriptive assumption in voting power indices: preferences and votes are treated as interchangeable and thus, albeit implicitly, strategic voting is idealised away.[18] So trivially, in this framework strategic voting does not affect voting power – we assumed away the very tools to theorise about strategic voting.

In order to theorise about strategic voting and the skill thereof we, at the very least, need to spell out the relationship between votes, outcomes and preferences, where votes cannot straightforwardly be read off from preferences (e.g., you also need beliefs and behavioural assumptions). Strategic voting is about how one may use votes such that one's preferences match the voting outcome. In other words, in order to model assumptions of unequal skill in strategic voting as posited by Satterthwaite and Dodgson, we need a conception of not merely of how potential actions (i.e., votes) relate to an outcome (e.g., the elected candidate), but how potential preferences or intentions relate to the outcome. The former described how we understood being a difference maker in the previous section, the latter is the understanding of a difference maker that we want to adopt in this section. Note, that in the following we will use the term skill quite loosely to describe the efficacy with which a voter may misrepresent their preferences in order to alter the outcome to one they prefer. Therefore, our usage of skill also includes the idea of circumstantial asymmetries in ability such as informational asymmetries.

To convince you that we need to adjust our definition of difference maker, let us consider the following example by Kolodny (2014b).[19] Let's imagine

[18] This may not necessarily mirror an underlying commitment in the literature; rather, the development of these power indices predated the onset of a systematic study of strategic voting within social choice theory. Although, Napel (2018, p.2) notes that strategic skill can change actual outcomes but notes that this is a posteriori perspective. However, our argument does not rely on the actual strategic skill of a particular voter or group of voters but instead relies on a class of distributions over strategic skill. In fact, we expand the implicit assumption of equiprobable skill of strategic skill and expand it to any other distribution.

[19] Kolodny's perspective is echoed broadly in existing literature; others also establish a connection between ability or how intention and outcome relate to power (e.g., Scott,

being restrained by a combination lock. You can unlock it by entering the correct numbers, and if you possess knowledge of the lock's combination, you inherently hold more power to influence the outcome — your intentions can be translated into action. You still possess the same range of possibilities to act as anyone else, and the number of combinations available for the lock remains the same. Moreover, given a specific action (i.e., combination of numbers), your chances of success are equal. However, if someone else lacks knowledge of the lock's combination, their attempts to unlock it can only amount to guessing. Consequently, Kolodny concludes that "social equality requires equal opportunity to knowingly influence political decisions in line with one's judgements." (Kolodny, 2014a, p.332)

Just as one might know which numbers will unlock a combination lock, one can have greater skill (or knowledge) in determining which choice on the ballot will result in the election of a particular candidate. A measure that just defines the possibility of such disparities away cannot serve as an adequate tool in our endeavour. To convince you that one needs to account for such disparities of skill even if one is merely interested in voting power through the voting rule, think about how the order of voting creates informational asymmetries and thereby can affect the skill of voting strategically. If a voting rule allows voting subsequently and openly the last voter will always be in a better epistemic position than the first voter and therefore be better skilled at voting strategically - they have full knowledge over ballot choices. This is an advantage a voter gains purely through the voting rule. Hence, in the following sections we will understand difference maker as not just someone who can change the outcome through their vote but someone who can knowingly change the outcome through their vote. While two votes seem to afford more influence than one, possessing one vote, coupled with complete understanding of its impact, provides greater influence over the outcome than

2018; Abizadeh, 2021; Dowding, 1996; Morriss, 2002) two votes used indiscriminately.

2.4.3 Strategy-Proofness and its Effect on the Power to Affect Change

In the following we will establish two effects that arise from a difference in skill to vote strategically on voting power. Both of them are what I call broad effects. We call them broad effects because they warrant assumptions that go beyond what is formalized by voting power indices. The first effect materializes once we include the content of voting outcomes in our consideration (contrary, for instance, to only observing structural features between preferences and outcomes). The argument is roughly that getting what you want, at least in a political context, systematically correlates with gaining more influence over others. The second broad effect occurs when we include in our considerations incentive structures for being skilled in voting. Then a violation of strategy-proofness may lead to a difference in the relative frequency of being a difference maker and thus in power differences in the very sense of voting power indices.

Establishing this achieves the following objectives. It reveals how descriptive assumptions in formalising power influence our judgements about when strategic voting is problematic and when it is not. The argument is not that these levels of descriptive idealisation are invariably correct, but rather to demonstrate that the connection between strategic voting and voting power cannot be dismissed outright. In an ideal world, the principle of "one person, one vote" might suffice for our voting rule. However, in a less than ideal world, we must re-evaluate the requirements that our voting rules should meet. Furthermore, the expansion and relaxation of these assumptions are meant to illustrate the conditions under which power differences through strategic voting arise systematically. Thereby presenting compelling cases for relational egalitarians, irrespective of whether we regard these as changes in formal voting power, or as considerations prompting us to reassess the formal structure of our voting rules. In any case they seem to provide a pro tanto reason to adopt voting rules that embody characteristics reducing, if not altogether preventing strategic voting.

2.4.3.1 Broad effect I - The Incentive to Realise a Difference in Skill

Our first broad effect draws on an observation made by Dowding and Van Hees (2008) but situates it within a context of power relations. The possibility to vote strategically encourages voters to learn more about how others vote as they can use that knowledge to get what they want. While Dowding and van Hees see this incentivisation as a reason to praise manipulation as it leads to more informed voters— through a lens of power it becomes a reason to be wary of manipulation.

If every voter possesses equal means to gather information about others, the possibility for strategic voting would merely result in enhanced mutual understanding of preferences and a universal elevation in the likelihood of knowingly influencing outcomes.[20] Yet, in a less ideal setting, varying recourses such as wealth, time or access to information would indeed exacerbate informational asymmetries, particularly given the plausible assumption that information acquisition bears cost, affordable by some and prohibitive for others. In essence, being a difference maker only pays off when strategic voting is possible. In a context where information acquisition is bearable only by some, a non-strategy-proof voting rule amplifies informational asymmetries, and these informational asymmetries, in turn, give rise to power discrepancies among voters by changing the relative frequency of being a dif-

[20]Note that from a power perspective, learning about other's preferences because one wants to understand someone else's' perspective or even change one's judgements is not necessarily an issue. The incentive here though works through one wanting to overturn the expressed preferences of others.

ference maker. Some are now more often in a position to knowingly change the outcome while others are not.

Let's once again circle back to our example of an election between three policies to underscore this point. Now suppose you completed an online test that recommends a policy for you to support and you signalled your commitment to follow this recommendation. Unfortunately, you didn't read the fine print when you accepted the conditions and cookies on that website. It turns out that your information and answers are sellable to third parties, and in this case, I am the third party. Given that I can profit by choosing my vote strategically and this information helps me do this successfully, I have a clear incentive to acquire this data on you. I now become someone who can knowingly change the outcome in accordance with my judgement and you cannot. For instance it turns out that you, as a progressive, rank policies as $p \succ s \succ c$, while I, opposing change, have the following preferences $s \succ c \succ p$. Uninformed about my preferences, you cast an honest vote for the progressive policy. I now know that only a vote for the conservative policy will overturn your vote. With this knowledge, I vote for the conservative policy in order to avert the election of the progressive policy. Note, that me possessing power over you emanates not from the of me achieving my preferred outcomes per se, but from my ability to alter outcomes that are obligatorily binding upon you, a phenomenon that arises from informational asymmetries not from me casting a strategic vote. Thus, not the strategic voting itself, but the anticipation of strategic voting is the driver for the shift in power relations. This is important to note, as it wouldn't be of concern to introduce a non-strategy-proof voting rule in a world where I wouldn't have any possibility to gather further information (or other means to further my skill) before I cast my vote.

Let's once again circle back to our example of an election between three policies to underscore this point. Now suppose you completed an online test

that recommends a policy for you to support and you signalled your commitment to follow this recommendation. Unfortunately, you didn't read the fine print when you accepted the conditions and cookies on that website. It turns out that your information and answers are sellable to third parties, and in this case, I am the third party. Given that I can profit by choosing my vote strategically and this information helps me do this successfully, I have a clear incentive to acquire this data on you. I now become someone who can knowingly change the outcome in accordance with my judgement and you cannot. For instance it turns out that you, as a progressive, rank policies as $p \succ s \succ c$, while I, opposing change, have the following preferences $s \succ c \succ p$. Uninformed about my preferences, you cast an honest vote for the progressive policy. I now know that only a vote for the conservative policy will overturn your vote. With this knowledge, I vote for the conservative party in order to advert the progressive policy to be elected. Note, that me possessing power over you emanates not from the of me achieving my preferred outcomes per se, but from my ability to alter outcomes that are obligatorily binding upon you, a phenomenon that arises from informational asymmetries not from me casting a strategic vote. Thus, not the strategic voting itself, but the anticipation of strategic voting is the driver for the shift in power relations. This is important to note, as it wouldn't be of concern to introduce a non-strategy-proof voting rule in a world where I wouldn't have any possibility to gather further information (or other means to further my skill) before I cast my vote.

In contrast, imagine a binary policy choice: a progressive policy and a status quo policy. Information about your voting intent provides no advantage to me. The election is strategy-proof; irrespective of your vote, I can only fare worse by not voting for my preferred policy. For example, were you to vote for the progressive policy no matter whether I prefer the progressive or the status quo policy my best course of action is to vote straightforwardly and the same holds true if you had voted for the status quo policy instead. Thus, I wouldn't benefit from getting data on you as it has no influence on getting what I want in the election. I wouldn't choose to acquire this data on you and our relative frequency of being a difference maker stays the same.

2.4.3.2 Broad effect II - Wanting Power

The second broad effect of strategic voting on power is the systematic correlation between achieving one's desires and subsequently acquiring power over one's fellow voters. Getting what you want is usually not seen as inherently synonymous with power (see Barry, 1980; Dowding, 1996). You can get lucky and just find the world to be such that the outcomes correspond to your preferences. Thus, it seems a mistake to equate getting what you want with power.

However, it seems equally mistaken to deny that there is a relation between getting what you want and accumulating power. Especially political elections pertain to political landscapes wherein electoral outcomes are inherently capable of reshaping power relations among voters. This is a point that is usually regarded beyond the scope of social choice theory but nevertheless important to make explicit, especially if one wants to theorise about normative foundations of voting institutions. There is voting power, but there are also other kinds of power structures that are important to normative theorists. For instance, Kolodny (2014b) recognises wealth as a form of informal power. It enables individuals to sculpt public opinion, lobby policymakers, and potentially sway electoral results. Moreover, we have good reasons to assume that skill in voting strategically correlates with financial means (Eggers & Vivyan, 2020) and thus with possessing informal power.

Consider that the three policies under discussion before are policies about inheritance tax schemes and as such directly affect the distribution of informal power through wealth among citizens. A reasonable presumption is that individuals typically prefer policies amplifying their power share and align their voting behaviour with such preferences. For the sake of the argument assume that voters prefer policies amplifying their power and vote accordingly – e.g., those with greater wealth voting for low taxes on high inheritances. If we assume that individuals with greater resources, such as wealth, are better equipped to engage in strategic voting (see Eggers and Vivyan 2020), then strategic behavioural would contribute to an exacerbation of informal power inequalities. Informal power imbalances are already notoriously difficult to eliminate (see Kolodny 2014b), and it is undesirable to have a voting rule that incentives their increase. If one follows this line of reasoning, the design of institutions such as voting rules should aim to prevent strategic voting in order to address and rectify informal power disparities. For a non-ideal theorists this seems a highly relevant context. Outcomes of political elections are intrinsically power-shaping and, within a political realm, securing what one prefers will typically manifest in an amplification of power.[21]

However, it is worth noting that not all elections are characterised by their impact on informal power structures. Some elections (e.g., election by the Norwegian Nobel Committee) may solely revolve around achieving personal preferences without shaping informal power dynamics between the voters. In such cases, there might be no harm in having a voting rule that allows for the possibility of strategic voting.

In sum, there will be scenarios where strategic voting will not result in power differences. However, these scenarios are of limited interest to us.[22] For example, we seem to have to exclude any political elections as well as any scenario where voters have different opportunities to react to the anticipation

[21]Note that this is not a point relying on actual skills and actual preferences but merely relies on distributions thereof.

[22]Similarly there are also elections where giving me more votes than you does not give me more influence over you.

of voting strategically. This leaves us with a significant collection of voting scenarios where we have grounds to worry about power and strategic voting – and if there is a potential for power imbalances there is also a potential for a violation of equal standing. It seems that we should at least try to respond to this in the design of our voting rules.

Similarly, just as one could have previously argued that giving voters different numbers of votes creates power imbalances and thus threatens equal standing, a complaint could be formulated against non-strategy-proof voting rules. The argument could proceed as follows:

The prospect of strategic voting incentivised us to learn about other's preferences, yet while you possess the resources to do so, I do not. Consequently, you will more frequently find yourself in a position to knowingly influence outcomes, thereby wielding greater power over me. This may threaten our equal standing as citizens in the same way as when those who read Kafka's Trial are accorded multiple votes. If the voting proposition is itself power shaping you may go on and assert: Furthermore, it is not just that you more often get what you want, getting what you want will deepen our power relations even further as you will vote in your favour when it comes to the distribution of informal power. Hence, your heightened chance of realising desired outcomes through strategic voting not only assures you get what you want but also systematically increases your (informal) power. This, in turn, might ultimately jeopardise our equal standing.

In the next section, we want to address a rival way to understand voting power in the voting power literature. These measures address how power and strategic voting relate if you think it doesn't just matter that you can make a difference but also what difference you can make. Under this conception of power we don't need to make any additional assumptions that go beyond the measures. Strategic voting increases power differences directly and thus have what I call, a narrow effect.

2.4.4 Strategy-Proofness and Its Effect on the Power to Realise Desired Change

In the preceding section we delved into the ramifications of strategic voting with voting power defined as the opportunity to influence an outcome through one's vote. In this section, we aim to address the opportunity to influence an outcome through one's vote in a way that aligns with one's preferences. It is crucial to distinguish this from the mere correspondence of outcomes with preferences, and it is also not the same as being a difference maker. While the former can be achieved randomly, as in a lottery, the latter could conceivably occur without any corresponding with one's preferences. Hence, they both differ from the opportunity to change an outcome in a direction that aligns with one's preferences.

Critiques of classical voting indices such as the Penrose-Banzhaf index are typically motivated by the conviction that the assumption of equiprobable distribution is untenable. These critics maintain that there are valid reasons to consider alternative preference distributions, arguing that the principle of insufficient reason is misapplied (Abizadeh, 2021). This shift in perspective often accompanies the acknowledgement that voters who are a member of a persistent minority are less likely to influence outcomes in their favour, despite possessing an equal number of votes than those who are not. Our aim is not to advocate for one conception of voting power over another, but rather to delineate the implications of each conception for the principle of strategy-proofness. Thus, in this section, we will explore the ramifications of this alternative notion of voting power.

As previously noted, a clear advantage of proficiency in strategic voting is the increased probability of realising preferred outcomes, despite an equal number of votes. Furthermore, while the correspondence of preferences with outcomes may be relevant to considerations such as welfare, it is typically not construed as directly a form of power but luck (e.g., Dowding, 1996). Instead, let us consider what strategic voting does to your opportunities to influence the outcome in a way that accords with your preferences. There are some approaches that argue that voting power is inextricably linked to the opportunity to influence outcomes in a desired direction, as opposed to merely influencing outcomes in any capacity (Barry, 1980; Steunenberg, Schmidtchen, & Koboldt, 1999; Napel & Widgrén, 2005, 2009). Thus, in contrast to those traditional power measures, these measures' notion of power "as the ability to get what you want" (see Schmidtchen & Steunenberg, 2014). In this debate this conceptual shift towards a preference-oriented view of power is inextricably tied to the assumed distribution of preferences. To see why, note that in an idealised scenario where preference is equated with ballot choices, a multitude of effects collapse into a single dimension. Abstracting from misrepresenting preferences, the sole means by which one could be a difference maker hinges upon *having* different preferences. Consequently, in this framework the opportunity to influence outcomes and the opportunity to influence outcomes in a preferred direction become synonymous.[23]

[23]While this idealisation (i.e. that preferences and ballot choices are identical) usually does not really change anything under the traditional voting power indices and thus is often employed, it really does matter for preference-based indices. Under this idealisation those in persistent minorities are less frequently positioned to affect change of any kind. To illustrate why, imagine that you advocate for progressive policies in a committee within a predominantly conservative field. Given this set-up, the probability of a tie occurring between the status quo and a conservative policy is significantly higher than a tie involving a progressive policy. Consequently, as you cannot vote strategically the probability of your vote making a difference is correspondingly low. Conversely, if your preferences align with the status quo or a conservative policy, your chances of breaking a tie increase substantially. This encapsulates the dilemma faced by persistent minorities, who not only experience a lack of correspondence between their preferences and electoral outcomes but also find their voices to be less likely to influence outcomes. Note that minorities are deemed "persistent" due to the recurring low probability of a tie involving a progressive policy in subsequent elections.

That being noted, let us reintroduce strategic voting and a difference in skill thereof. Then we are in a position to judge how strategy-proofness relates to this conception of voting power.

2.4.4.1 Narrow effect – Controlling the Outcome in Your Favour

With the aforementioned conception of power, i.e., it matters not just that you can make a difference but that you can make a difference in your favour, the relation between strategic voting and power becomes immediately obvious.

Consider the scenario in which we vote among three policies. Being skilled in voting strategically will always improve your ability to get what you want – independently whether you think the principle of insufficient reason over the distribution of preferences applies or not. To see this let's assume an equiprobable distribution again while maintaining the view that voting power is fundamentally about the ability to get what you want.

Under full knowledge of everyone's vote, your opportunity to be a difference maker is simply the combinatorial possibility of swinging the outcome by making a different choice on the ballot. Yet, in the absence of full knowledge, this doesn't hold true anymore. There is a parallel phenomenon happening for the opportunity and the possibility to change the outcome in a direction that you desire. Under the assumption of an equiprobable distribution over preferences, it is still true that you and I equally often have the possibility to change the outcome. Yet, if I am better skilled (e.g., because I bought your data) I can utilise these possibilities, because I know what vote I need to cast to turn the voting outcome. It then follows that in a subset of these ties I can overturn the outcome in my favour (i.e., I can change the outcome to one that I actually prefer).

To illustrate this point, consider the following: In contrast to you, I am a difference maker once there is a tie, and I can change the outcome by voting strategically. Not all of these potential changes would be in my favour. Let us assume that you voted for the progressive policy and I am the one disliking change. This time around the conservative policy is quite an extreme one, actually proposing tremendous changes to the political landscape. Thus my preferences are the following $s \succ p \succ c$. My knowledge on how you are voting does not help me to change the outcome in a way that I desire. My only way to change the outcome is by voting for the conservative policy. This makes matters worse! While under the prior conception of power this scenario would count toward my voting power under our current conception of voting power it does not. Nonetheless, I will be able to change the outcome in a direction that I desire more often than you. That is because if the conservative policy where not as extreme and I had the preferences such that $s \succ c \succ p$ I could use that knowledge to change the outcome in a way that I desire, whereas, you, in the absence of knowledge about my votes couldn't knowingly change the outcome in either of the two scenarios.

To summarise if we were to understand power as sketched out in this section, we could now formulate your complaint in the following way:

Having a voting rule that allows for strategic voting puts you in a situation where you can make me do more things against my will in a direction that you desire than I can make you do things against your will in a direction that I desire. Namely there are more possible preference profiles where you can change the outcome in the direction that you prefer, granting you more voting power than me. Thereby our institutions favour those who are better at voting strategically. This is equally inadmissible as just giving everyone more influence who read Kafka's Trial or whose name starts with A.

To isolate this new effect from the broad effects we have identified before, imagine that the outcomes of this election are not themselves power-shaping. Furthermore, note that, in contrast to the last conception of power, introducing a voting rule that allows strategic voting remains problematic even after I have already decided to acquire potential information about your vote (e.g., buying your data online). In other words, not the anticipation of the possibility to vote strategically is the issue but the opportunity of voting strategically. Imagine that I want to sell you a product and your political inclinations are beneficial for this. Independent of the election, I have an interest in your vote and thus have already acquired data on your voting behavioural. Under the previously considered conception of power, introducing a voting rule in this scenario that allows strategic voting wouldn't be problematic because it wouldn't lead to an amplification of the relative frequency of us being difference makers. The possibility of voting strategically has no effect on me wanting to get information about you. In contrast, the crux under this section's concept of voting power is that strategic voting allows me to utilise the information to change the outcome to one that I want, not that the anticipation of strategic voting nor the fulfilment of my preferences has negative downstream effects for power relations.

Let me give you an example on how this point extends beyond employing voting rules that are less sensitive to strategic voting. Let us follow the literature around this conception of power and assume some distributions are more likely than others. If one subscribes to the logic of preference-based voting indices, then it follows that persistent minorities wield less voting power than their majority counterparts. One way to counter this power imbalance would be for persistent minorities to be well-versed in strategic voting. Consequently, if a voting rule is not strategy-proof, this line of argument would lend support to policies that heighten persistent minorities' skill to vote strategically (e.g., by making information about other groups voting behaviour more accessible). While we focused in this chapter on individual strategic voting similar reasoning might be applied to different types of voting manipulations such as coalition building. If that is the case we would also have reasons to support policies that enable persistent minorities to a effectively utilise their votes and to coordinate their voting strategies.

Let us now circle back to what we established in this section about not just the connection between strategic voting and power but the normative evaluation of strategy-proofness. To summarise, we can distinguish between two approaches to understanding voting power: one that is preference-independent and another that is preference-dependent. Both frameworks enable us to establish a link between strategic voting and power relations and draw the implications for strategy-proofness. According to our first conception of voting power, strategy-proofness indirectly influences power relations by reducing the opportunities to exploit strategic voting skills for informal power gains and diminishing the incentives for the asymmetric accumulation of information. In contrast, under the second conception, strategy-proofness directly addresses the disparities in power that arise from unequal skills in strategic voting. I hope to have presented a persuasive argument that, in non-ideal scenarios, the differences in power relevant to relational egalitarians suggest that more than the principle of "one person, one vote" is required from our voting rules. It appears that if we disregard strategic voting, our theories about voting and the demands we place upon them fail to apply to any real-world voting context.

2.5 Conclusion

In conclusion, I explored the concerns within social choice theory regarding strategic voting and its transformation of voting into a game of skill. We have not just established that strategic voting can create power imbalances among citizens but that a non-strategy-proof voting rule can intensify these power imbalances. Non-ideal theorists should worry about strategic voting on power grounds and have a pro tanto reason to adopt a voting rule that satisfies strategy-proofness. However, depending on the conception of power one adopts, the circumstances where this pro tanto reason for strategy-proofness holds will vary. These insights are essential for normative theorists aiming to comprehend and mitigate the effects of strategic voting within democratic frameworks.

Looking ahead, this chapter establishes a groundwork for subsequent research in social choice theory. Future studies could strive to enhance our understanding of the interplay between various conceptions of power and strategic voting and voting rules. Furthermore, articulating normative objectives more clearly can assist in navigating the necessary trade-offs in social choice theory. For example, explicitly defining what justifies voting (e.g., the principle of equal standing) can facilitate a balance between conflicting desiderata of a voting rule and inform which relaxations of properties, such as strategy-proofness, are justifiable.

3

Manipulability, Measurement and Mathematical Models

Abstract

The answer to the question why manipulation is bad is not only interesting in itself but also a prerequisite for assessing how vulnerable our institutions are to manipulation. In this chapter, I explore how to measure the vulnerability of institutions to manipulative strategic voting via voting rules. First, I examine candidate answers to the normative question of why manipulation is harmful by discussing two main objections to strategic voting found in the literature: non-transparency and the notion of voting as a "game of skill". In the second part of the chapter, I assess two frameworks used in social choice theory to explore vulnerability to strategic voting, namely the Nitzan-Kelly Index and the practice of categorising voting rules according to their computational complexity. Using these examples, I argue that the literature plausibly needs to develop several metrics that track different normative concerns about manipulation.

3.1 Introduction

> "It can be argued that some of the difficulties in the general theory of social choice arise from a desire to fit essentially different classes of group aggregation problems into one uniform framework and from seeking excessive generality" - (Sen, 1977)

Manipulation leaves a bad taste. We associate it with illegitimate forms of influence. When we design institutions for collective decision-making we naturally worry about manipulation. One such behaviour that is repeatedly labelled manipulation is the strategic misrepresentation of preferences to achieve a preferred outcome.

In voting theory, this type of misrepresentation is better known as strategic voting and may be the most publicly discussed form of individual manipulation.[1] In the context of voting theory, it is the default assumption that manipulation is bad, and therefore, the manipulability of a voting rule is also considered bad. However, as noted in the previous chapter the literature does not often argue explicitly for why strategic voting is bad. At times, it seems that strategic voting is assumed to be harmful merely because it is labelled as manipulation. Unsurprisingly, the prevailing view in the literature is that systems should be designed to be as resistant to manipulation as possible (List, 2022; Chamberlin, 1985; Saari, 1990).

The aim of this chapter is to take seriously the normative and descriptive entanglement of the term "manipulation", and with it, the implications for measuring manipulation in voting theory. I argue that measuring a voting rule's vulnerability to manipulation involves measuring a thick concept. Thick concepts are dual in nature, entailing both descriptive and normative claims (see, for example, Foot, 1958; Williams, 1985; Elstein & Hurka, 2009;

[1] See chapter 2 for examples of public debates about strategic voting.

Dancy, 1995; Putnam, 2002). Therefore, when assessing a voting rule's vulnerability to manipulation, we must account for the normative, evaluative dimension of manipulation. Moreover, the assessment must get the normative evaluation right in order to have normative force. What complicates this task, however, is that the reasons put forward in the literature against manipulation differ. I will argue that these different reasons against manipulation in the literature can be understood as a difference in values. In light of this, I contend that current debates around metrics for evaluating the vulnerability to manipulation in voting rules are misguided. If there is a plurality of value judgements regarding why strategic voting is bad, then we may also need multiple metrics to assess the vulnerability of voting rules to manipulation in ways that reflect these values.

The chapter is structured as follows. First, I introduce the concept of strategic voting as understood in social choice theory. Second, I discuss two of the main objections to strategic voting found in the literature: non-transparency (Dowding & Van Hees, 2008) and the notion of voting as a "game of skill" (Satterthwaite, 1973)[2]. I will argue that non-transparency is a concern for welfarists in aggregation procedures, while the worry about voting becoming a "game of skill" arises for relational egalitarians, but not necessarily vice versa. Third, I will argue that adopting these different values changes the object of normative concern in discussions about manipulation. Welfarists are concerned with how the election outcome relates to voter preferences, whereas relational egalitarians focus on the unequal opportunities to influ-

[2]Those do not exhaust the space of potential reasons against strategic voting. Another reason, perhaps most frequently mentioned in the philosophical literature, is the sincerity argument (Dowding & Van Hees, 2008). A similar argument can be made that insincerity is a character-based value, distinct from both welfarist and egalitarian values, and therefore another consideration that must be taken into account when constructing metrics for the vulnerability of voting rules to manipulation. For those concerned with dishonesty, the object of normative significance is typically a function of the voters' preferences and their ballot choices.

ence the election result and, by extension, each other. Fourth, I will sketch how these two different evaluative perspectives impact the debate about how to assess vulnerability to manipulation in voting rules. I will use the discussion between proponents of two different approaches—the Nitzan-Kelly Index and computational complexity in social choice theory—to illustrate this point. I conclude that the literature will plausibly need to develop multiple metrics if it wishes to avoid prescribing a single value judgement across all applications of voting rules.

3.2 Strategic Voting

Let me remind you that an instance of strategic voting in the technical sense of social choice theory is as follows. Imagine the three policies from chapter 2. I am a progressive and rank the policies $p \succ s \succ c$, whereas you dislike any change and thus rank the policies $s \succ c \succ p$. Given that I vote sincerely (i.e., for my first-ranked candidate), I will vote for the progressive policy. You, however, will do better by casting your vote not for the status quo policy, but for the conservative policy instead. This is because, in the event of a tie between the progressive policy and the status quo policy, the lexicographical rule breaks in favour of the progressive policy. However, in a tie between the conservative and progressive policies, the lexicographical tie-breaker rule favours the conservative policy. If you vote for the conservative policy in this case, you have voted strategically. If a voting rule allows situations (i.e., for some combinations of preferences) where voters like you can profit from strategic voting, the voting rule is not strategy-proof (Gibbard, 1973; Satterthwaite, 1973). Voting rules typically allow some possibility of strategic voting and thus are not strategy-proof.

More specifically, strategic voting is defined for social choice rules. A social choice rule assigns a specific candidate as the outcome for every combination of preference orderings (one for each individual), where each individual expresses their preferences over a set of candidates. The Gibbard-Satterthwaite theorem establishes that any voting rule over more than two candidates that satisfies a few minimal criteria cannot be strategy-proof. In other words, it demonstrates that it is not always advantageous for individuals to reveal their true preferences, because there will always exist a scenario where at least one person has an incentive to deviate from their true preferences.

If one wants to adopt a voting rule that is strategy-proof, one must be willing to accept some drastic restrictions on the voting rule, such as limiting the choices to two options or adopting a dictatorship. Thus, since we cannot construct strategy-proof voting rules, the literature has naturally shifted to the question of how to achieve the second-best option.

What is curious, however, is that despite the significant effort in the literature to find solutions to circumvent the Gibbard-Satterthwaite theorem, the reasons why this is desirable are rarely fully explored. In the following, I will address two reasons from the literature—non-transparency and turning voting into a "game of skill"—and flesh them out in terms of value commitments.

3.3 Reasons against Manipulation

3.3.1 Non-transparency

One of the most commonly mentioned arguments against strategic voting is its non-transparency (Dowding and Van Hees, 2008). The non-transparency created by strategic voting primarily concerns the true preferences that voters have over the candidates. This, in turn, can lead to other consequences that worry some (Satterthwaite, 1973) while others are unconcerned (Dowding and Van Hees, 2008).[3][4] If a voting outcome is non-transparent, this poses an initial epistemic problem for a potential social planner (Conitzer and Walsh, 2016). It becomes a more significant concern though if one seeks to establish a voting rule as superior because it produces outcomes that aggregate the preferences of voters in a specific way (e.g. Mueller, 2003, for an argument for the majority rule on utilitarian grounds). Such arguments are often employed when advocating for a voting rule based on welfare considerations.

While welfare is explicitly a concern for social welfare functions that produce a ranking of social welfare, it is also pertinent to many voting rules (see Lehtinen, 2015, for an argument for strategic voting on welfare grounds). One of the most significant criterion against which voting rules are evaluated is the weak Pareto principle. This principle requires that if every individual prefers candidate A over candidate B, then the voting rule should not elect candidate B over candidate A. The weak Pareto principle is commonly justified as a welfare principle in ordinal frameworks, where preference orderings contain no information about the intensity of individual preferences. It is difficult to reject the weak Pareto principle, as electing a candidate that is preferred by all seems compelling (List, 2022).[5]

One could argue that an outcome that is a Pareto improvement over another outcome is superior, as it is better or at least not worse for every voter. Moreover, voting rules guaranteed to produce such outcomes, if they exist, are preferable to those that are not. Strategic voting, as defined in social choice theory, changes the outcome compared to what it would have been

[3]For instance, knowing that revealing your true preference before voting can be used against you provides an incentive not to reveal them (Satterthwaite, 1973). This could be problematic if deliberation occurs before voting.

[4]If one aggregates beliefs rather than preferences, the relationship would concern the outcome and the beliefs held about a proposition.

[5]The weak Pareto principle is sometimes also referred to as the unanimity principle. Note that I am not suggesting it can only be argued for on welfare grounds.

under straightforward voting. Therefore, if the legitimacy of voting depends on the outcomes produced by the voting rule, strategic voting has the potential to undermine these justifications (see Conitzer and Walsh, 2016 for a discussion on "bad equilibria" due to strategic voting). If one justifies an aggregation mechanism such as a voting rule based at least partly on welfare considerations, then losing epistemic access to welfare gains undermines the justification of that mechanism.

To sum up, the object of normative significance concerning manipulation is the welfare loss resulting from the changed outcome.[6] Strategy-proofness is seen as instrumentally valuable because, for instance, the weak Pareto principle[7] as a welfare criterion is defined based on truthful preferences (Barberà & Jackson, 1992; Dasgupta & Maskin, 2020).[8]

Significant portions of social choice theory concentrate on welfare aggregation and aim to rank aggregation methods based on their social welfare outcomes. Through the weak Pareto principle, one can indirectly advocate for strategy-proof voting rules on welfare grounds. However, there exists another tradition in social choice theory that aligns more closely with political philosophy than with welfare economics. In the following section, I will outline how one can object to manipulable voting rules on egalitarian grounds.

3.3.2 Game of Skill

Not all legitimations of voting rest on the welfare gains it may produce or, for that matter, the outcomes voting tends to produce. Voting can also be

[6]One could make a similar argument with judgement aggregation. Plausibly, the object of normative significance lies in the epistemic loss through the change of outcome.

[7]If it is explicitly used as a welfare criterion it is also often called weak Pareto efficiency.

[8]In mechanism design, there is work on the trade-offs between welfare loss and strategy-proofness (see Dubins & Freedman, 1981; Roth, 1982).

argued for on non-instrumental grounds. We can argue that voting is justified by the procedure itself, independently of the outcomes it produces, for instance, on egalitarian grounds.

One sentiment about voting is that it allows everyone to have an equal say over decisions that are of common concern (Christiano & Bajaj, 2022). Voting is perhaps the only decision mechanism that respects each person's point of view on matters of common concern by granting everyone an equal say in cases of disagreement (see, for example, Singer, 1973; Waldron, 1999).

The principle of "one person, one vote" is supposed to ensure that every citizen can, at least in principle, influence political decisions to the same degree. In this sense, "one person, one vote" is a critical mechanism for preventing any one person from having an inappropriate degree of control over others. Voting mechanisms are often fundamentally relational in that they shape the ways in which individuals are able to relate to one another as equals. There would be something especially problematic about endorsing and nourishing power inequalities through the structures and institutions used to organise ourselves.[9]

Given unequal "skill" or information about others, some votes are more effective than others regarding casting their votes (see Dowding and Van Hees, 2008), which undermines the principle that everyone has equal influence in the process of voting. This concern is echoed in the literature on manipulation, where, for instance, Satterthwaite (1973) is worried that strategic voting makes voting a "game of skill" (for a similar concern, see Dodgson 1876, reprinted in Black 1958 and Conitzer and Walsh 2016), who argue that those who are less well-informed will cast less effective votes, or Eggers and

[9] Abdulkadiroglu, Pathak, Roth, and Sönmez (2006) argue for a strategy-proof matching mechanism because it equalises the influence of the outcome. However, they do not invoke egalitarian but fairness criteria.

Vivyan (2020), who claim that the existence of inequalities in strategic voting speaks against electoral systems that reward it.

Normative theories like relational egalitarianism legitimise voting partly on the grounds that equal power over each other is an important aspect among equals that requires the implementation of voting to organise common aspects of living together (see Kolodny 2014 for an example). In other words, turning voting into a game of skill may undermine the very reason why voting is a legitimate procedure in the first place. For a more detailed argument how a difference in skill to vote strategically can influence voting power see the previous chapter.

If voting is justified by the idea that the process should give everyone the same opportunities to influence decisions, then looking at the effects of election outcomes is not directly informative of the harms of manipulation. Note that, in contrast to the argument based on non-transparency, here we are more interested in dispositional concepts such as power over others or the opportunity to influence the outcome (Dahl, 1957; A. Allen, 2022; Dowding, 1996; Felsenthal & Machover, 1998; Holler & Nurmi, 2014; Lukes, 2005; Kolodny, 2014b). Furthermore, we are interested not just in the absolute opportunities to vote strategically but also in the opportunities relative to those of other voters. In the most extreme case, this would mean that the election outcome could be overturned every time without harming the legitimacy of the voting procedure, provided that everyone had the same opportunities.

To sum up, the objective of normative significance for strategic voting is the distribution of opportunities to influence the electoral outcome. We thus see that there are two different concerns about strategic voting. One is the worry about the non-transparency of the outcome that may arise if one is concerned with welfare. The other is that strategic voting turns into a game of skill, which gives some voters more control over the outcome (and thus potentially over each other), undermining values such as equality. This section aimed to establish that manipulation in the form of strategic voting can have different normative evaluations grounded in different values. In the next section, I will discuss the concept of manipulation as a thick concept with both evaluative and descriptive dimensions. I will establish that the measurement of a thick concept will only have normative force if it also takes into account the evaluative dimension of the term manipulation.

3.4 Measuring Manipulability as a Thick Concept

3.4.1 Thick Concepts

Manipulability is a thick concept. Since Williams (1985) introduced the term "thick concepts", it has been discussed in philosophy of language, ethics and philosophy of science (Putnam, 2002).[10] Evaluative terms are typically categorised as either thin or thick (Väyrynen, 2021). Thin terms such as "good" or "bad" offer broad, general evaluations, often without detailed descriptive content. For instance, labelling an action as "bad" conveys a negative evaluation but does not provide much specific information about the action itself. This lack of detail makes thin terms broadly applicable but limited in descriptive richness.

In contrast, thick concepts incorporate both evaluative and descriptive ele-

[10] Alexandrova (2018) makes a convincing case that "mixed claim" that resemble thick claims, yet evade some of the controversies, especially about moral realism about thick claims, belong in science. We continue with the term of thick terms though to keep with the terminology most common in the literature, while being sympathetic to her.

ments. An example of a descriptive term would, for instance, be "square" or "gold" (Väyrynen, 2021). Examples of thick concepts include moral virtues, like "generous" or "selfish," but also span other categories such as "imprudent," or "banal". When we describe someone as "selfish," we are not just negatively evaluating their actions; we are also indicating that their actions prioritise their interests over others' (Väyrynen, 2021). This dual nature of thick concepts—their combination of evaluative judgement and specific descriptive content—is what distinguishes them from thin concepts.[11] The measurement of thick concepts poses a particular challenge as one needs to account for the evaluative dimension of the term that one measures as well (Alexandrova, 2018). The measurement of manipulability in voting is no exception in this regard.

The challenge in measuring thick concepts lies in answering the question: How do we devise measurement tools that accurately capture both their descriptive and normative aspects? For instance, to identify a metric with normative force in the science of well-being, one must clarify how the quantity being measured relates to what constitutes well-being. What causes well-being can vary greatly depending on views that one may ascribe to, such as 'positive hedonic profile', 'life satisfaction', or a sense of 'flourishing' (see Alexandrova 2018). For example, the question whether observing choice behaviour is a good measure for preference satisfaction does not settle whether choice behavioural is informative about well-being. That is the case because different views about well-being assume different relations between well-being and preference satisfaction. Hence, merely settling whether choice behaviour is a good measure of preference satisfaction is not enough for a measure of well-being to have normative force.

[11]Characterising thick concepts beyond their characteristic of blending evaluative and descriptive elements proves difficult in the literature. Thus we will use thick concepts will trying not to remain neutral on other positions around thick concepts.

There are no fleshed-out accounts of what makes manipulation via strategic voting bad in the same way in which there are positions about what constitutes well-being. However, there is an abundance of different ways to assess the manipulability of voting rules (e.g. Brams & Fishburn, 1978; Saari, 1990; Aleskerov & Kurbanov, 1999; Chamberlin, 1985; Favardin & Lepelley, 2006; Favardin, Lepelley, & Serais, 2002; Kim & Roush, 1996; Lepelley & Mbih, 1994; Peleg, 1979; Pritchard & Wilson, 2007). Some make restrictions on the type of preferences voters can have (Black, 1958; Moulin, 1980; Barberà, Gul, & Stacchetti, 1993; Nehring & Puppe, 2007a, 2007b). Others move to weaker versions of strategy-proof solutions such as tolerable manipulability (Feigenbaum & Shenker, 2004), threshold strategy-proofness (Roberts & Postlewaite, 1976), strategy-proofness with counterthreats (Pattanaik, 1978) or asymptotic strategy-proofness (Slinko, 2004). These are weakenings of strategy-proofness in the sense that if two voting rules violate strategy-proofness one rule may still be held to be less manipulable than the other voting rule if one satisfies a weakened requirement and the other voting rule does not.

Different approaches to assessing the manipulability of voting rules do, in fact, diverge on which voting rule is more susceptible to manipulation. Yet, the literature that aims to measure the manipulability of voting rules does not engage with the normative dimension of manipulation. In the following, I will outline two approaches to assessing the vulnerability of voting rules to strategic voting: the Nitzan-Kelly Index and the use of computational complexity to determine the difficulty of voting strategically. We will then use these two approaches in the subsequent section to illustrate the argumentative gaps that arise from an underappreciation of normative evaluation in the discussion between proponents of different metrics. I will outline how a disagreement over which assessment to use can be framed in terms of differing value judgements—namely, welfare versus egalitarian concerns—in the evaluation of manipulation.

3.5 The Nitzan-Kelly Index vs. Computational Complexity

3.5.1 The Nitzan-Kelly Index

One of the most established approaches to measuring manipulability is the Nitzan-Kelly Index (Kelly, 1993; Nitzan, 1985). If a voting rule is manipulable, we are saying that there exists at least one possible preference profile under which an agent can vote strategically. A preference profile is a possible combination of preferences between all agents. The Nitzan-Kelly Index considers the ratio of preference profiles where manipulation is possible to the total number of preference profiles. According to the Nitzan-Kelly Index we then say that a voting rule is less manipulable if it is manipulable at fewer preference profiles, or equivalently, if it has a smaller index according to the Nitzan-Kelly Index.

To get a first informal understanding of what the Nitzan-Kelly index does, imagine again that we were to vote about the three different policies: the progressive policy (p), the conservative policy (c) and the status quo policy (s). And imagine again that we fixed the voting rule to the majority rule with a lexicographical tie breaker. The Nitzan-Kelly Index then counts every preference profile where at least one of us could do better by not voting straightforwardly and then divides the number of preference profiles where one us could profit from voting straightforwardly by the number of all preference profiles. The Nitzan-Kelly index is usually not interpreted just as the combinatorial possibilities of strategic voting but given a probabilistic interpretation (Nitzan, 1985; Kelly, 1993; Aleskerov & Kurbanov, 1999; Smith, 1999). There are multiple assumptions one may make about the preference profiles and their probability distributions. The most popular one is called impartial culture (IC). Under IC it is assumed that any strict ranking of the three policies is equally likely for you, that is $p \succ s \succ c$ is as likely as $c \succ p \succ s$ and so on. Furthermore, we assume independence of our preferences. Your preferences are not informative for what preferences I may have and so on. Under this interpretation we would get the following version of the Nitzan-Kelly index:

The Nitzan-Kelly Index is given by:

$$\text{Nitzan-Kelly Index} = \frac{d_0}{(m!)^n}$$

where m is the number of candidates, $m!$ is the number of all possible strict preference rankings on the set of options, n is the number of voters, and $(m!)^n$ is the number of all possible preference profiles. Finally, d_0 is the number of manipulable profiles. In our example, this would be filled out as follows:

$$\frac{d_0}{(m!)^n} = \frac{4}{(6)^2} = \frac{1}{9}.$$

Thus, our voting rule has an index of $\frac{1}{9}$.

The four preference profiles where you or I have an incentive to vote strategically are the following: I am a progressive and rank the policies $p \succ s \succ c$, whereas you dislike any change and thus rank the policies $s \succ c \succ p$. There is also the preference profile where our preferences are just flipped. A similar situation occurs when one of us has the preferences $p \succ c \succ s$ and the other has $s \succ c \succ p$.

Let's slightly change the tie-breaker rule for the next example. Instead of a lexicographical tie-breaker rule over candidates, we adopt a lexicographical tie-breaker rule over voters, where the voter whose last name comes first in the alphabet resolves the tie. Let's assume that this is you in our case.

Then the index for this voting rule is the following:

$$\text{Nitzan-Kelly Index} = \frac{0}{(6)^2}$$

There is no possibility for me or you to manipulate. The reason is that the tie-breaking rule made you a de facto dictator. Thus, this second voting rule fares better according to manipulation. Note that we don't say it is overall better. However, we say it is better or more desirable along the dimension of manipulability. Yet, this is not the only way to evaluate the vulnerability to strategic voting. In a next step, let's contrast the Nitzan-Kelly Index with the computational complexity approach.

3.5.2 The Computational Complexity Approach

In recent years, interest in the computational aspects of social choice theory, and in particular in the computational aspects of voting, especially for manipulability (Zuckerman et al., 2009) has sharply increased. Generally, algorithmic complexity tries to approximate the running time of an algorithm for a given type of problem. This is usually done by assuming a smallest unit of calculation, an input variable to the algorithm, and then calculating how fast your units of calculations multiply depending on the size of the algorithm input. Depending on how rapidly calculations scale with the input, we can classify the problem into a complexity class, which in turn indicates whether, in principle, an algorithm could be found to solve it within a certain runtime.

Complexity measures of manipulation sort voting rules into complexity classes and then call some voting method computationally resistant or "hard to manipulate" if finding a successful path to vote strategically falls, for instance, into a complexity class called NP-complete (Bartholdi III, Tovey, & Trick, 1989).

NP-completeness tells us that a problem is infeasible to solve quickly[12] in the general case, which implies that for any algorithm we know today, and for sufficiently large instances of the problem, it takes impractically long to find a solution. However, note that this doesn't mean that every instance of an NP-complete problem is difficult to solve. Additionally, even though finding a solution is hard, verifying whether a proposed solution is correct can be done quickly (in polynomial time). So if we happen to stumble upon the correct solution, we will know it quickly.

More concretely, what the computational complexity approach to strategic voting tests is whether there is an algorithm that can solve the following decision problem: given that all other voters' ballots are fixed and known, can you misrepresent your preferences in such a way that a particular candidate c wins? This question is at the heart of strategic voting, since determining whether you can benefit candidate c by mispresenting your preferences on your ballot requires exactly this kind of analysis. Moreover, because we ask this question for any candidate c, we are effectively testing whether finding a successful misrepresentation is computationally easy or difficult.

[12] In computational complexity theory, polynomial time refers to a class of problems that can be solved by an algorithm whose running time grows polynomially with the size of the input. Specifically, an algorithm is said to run in polynomial time if its execution time can be expressed as a polynomial function of the length of the input n, denoted as $O(n^k)$ for some non-negative integer k. A problem is classified as easy to solve if it can be solved in polynomial time.

Let us circle back to the example of you and me voting. Imagine we were to vote again with majority rule with a tie-breaker that favours you. Now we consider the task of finding a voting strategy that elects a certain candidate under this rule. We would either find an algorithm that solves this task in at most polynomial time or prove that such an algorithm does not exist. Given our scenario we can find an algorithm that could solve the task to determine whether you can elect some candidate c given my ballot choices. This algorithm is a very simple instance of what is called a greedy algorithm. A greedy algorithm is typically not computationally hard.

i) Place c on top of your preference ranking.

ii) Put any not yet positioned candidate on the next free position in your preference order, given that they don't prevent c from winning.

iii) If (ii) is not possible, return `false`.

If we apply this type of algorithm to strategic voting under the majority rule, the steps become trivial: vote for candidate c and observe whether c is elected. This algorithm would guarantee either that voting for c strategically elects c, or, if not, confirm that it is not possible.

However, such an algorithm would not work for another voting rule such as Single Transferable Voting (STV). STV proceeds in a number of rounds. Unless one candidate has a majority of first place votes, we eliminate the candidate with the least number of votes first. Any ballots placing the eliminated candidate in first place are re-assigned to the second place candidate. We then repeat until one candidate has a majority. It can be proven that this voting rule does not have an algorithm like the one described above (Bartholdi III & Orlin, 1991). That is to say that we can prove that if strategic voting under a given preference profile is possible just voting for that candidate is not guaranteed to be a successful strategy. Moreover, we can prove that no other algorithm exists where the calculation steps at most grow polynomially with the input size (number of candidates and voters).

Our simple example of majority voting with a tie breaking rule in your favour already shows where in principle these two approaches can diverge in their verdict about the manipulability of a voting rule. Majority voting with our tie breaking rule does very well according to the Nitzan-Kelly Index but very badly according the computational complexity approach.[13] In other words, there is no possibility to vote strategically (i.e., the Nitzan-Kelly index is zero) but to figure this out is maximally easy (the problem can be solved in polynomial time). This may be a trivial example as this is achieved by making you a de facto dictator. However, it nicely shows how a different verdict can be reached even on more complex cases.

3.6 Metrics of Manipulability and Normative Guidance

Having described two distinct frameworks for measuring manipulability in voting rules, it is evident that they capture different aspects of manipulability, and their evaluations can indeed diverge. However, if both purport to measure manipulability but yield different verdicts, we require a means to determine which of their verdict to follow. In what follows, I outline why the current state of the debate fails to provide a satisfactory resolution. I will argue that the current debate between proponents of the Nitzan-Kelly index and the computational complexity approach can be reconstructed as a debate about descriptive assumptions in their models. However, the debate fails to sufficiently recognise that normative models inevitably rest on normative assumptions. The issue is exacerbated by the fact that reasonable differences

[13]Note that proponents of the computational complexity approach wouldn't take their approach in this instance as a good measure of manipulability because the input variables are very low. Yet, for illustrative principle we will use this example.

in these normative assumptions—such as manipulation being objectionable on welfare grounds versus egalitarian grounds—can exist. To make matters worse, the decision on which normative assumption to adopt can determine which descriptive assumptions in a model are appropriate.

Let me start to reconstruct the debate about which framework to use. We regard the satisfaction of strategy-proofness as a desirable property and when a voting rule is deemed less manipulable, it is usually taken to be closer to what makes strategy-proofness desirable.[14] The normative force of manipulability does not derive from a property of the voting rule in our mathematical model, but rather from how that property transfers to something normatively desirable in the target system. Social choice rules are collections of mathematical models of real-world voting systems. As such, they are susceptible to criticisms regarding the descriptive inaccuracy of their assumptions. Descriptive idealisations involve false assumptions, introduced to simplify the model or make it more mathematically tractable (Beck & Jahn, 2021). These descriptive assumptions are not only important to scrutinise in predictive or descriptive models but also present a puzzle for normative models, where normative verdicts must remain valid despite involving factually inaccurate idealisations (Beck & Jahn, 2021). If we want a better-than-judgement according to manipulability, the metric employed must track something of normative relevance.

The debate of manipulability in social choice theory can be understood as a dispute over which descriptive assumptions are appropriate for models of manipulability to exert normative force. That is, which descriptive assumptions about voters are appropriate so that the model measures something desirable in the target system (e.g., welfare). We then understand the debate about

[14]For a more in depth treatment how models may deliver normative guidance see Beck and Jahn, 2021 and Roussos, 2022).

manipulability as an instance of a long-standing and recurring debate regarding how realistic assumptions need to be within models (e.g., M. Friedman, 1953; Hausman, 1994; Mäki et al., 1992; Morgan, 2012; Cartwright, 2009). For example, the IC assumption is routinely criticised as unrealistic, and that more realistic distributions over preference profiles should be employed to measure manipulability (see Tsetlin, Regenwetter, & Grofman, 2003; Mattei, Forshee, & Goldsmith, 2012). If the distribution over preference profiles is unrealistic then we may get wrong verdicts about how probable it is that manipulation is possible under a voting rule. Moreover, one of the justifications for the computational complexity approach to assessing manipulability is the conviction that the Nitzan-Kelly Index does not account for how the 'logical' possibility of manipulation translates to reality (Chamberlin, 1985). As a consequence, we may overestimate how probable it is that manipulation happens in our target system when using the Nitzan-Kelly index because we presuppose an overly demanding view of rationality or reasoning capacities (Kube & Puppe, 2009; Tal, Meir, & Gal, 2015; Conitzer & Walsh, 2016).[15] In light of this, it is not surprising that some researchers have utilised empirical evidence to argue that the computational complexity approach is predictively successful, and that, consequently, its descriptive idealisations about the reasoning capacities of voters are more appropriate (for example Harrison & McDaniel, 2008; Tal et al., 2015; Flanigan, Liang, Procaccia, & Wang, 2024).

Scrutinising one's models for descriptive accuracy is surely an important task in evaluating which metric is well-suited to evaluate a voting rule, but note that this debate does not suffice to establish that a metric provides normative guidance. Models that offer normative verdicts involve both descriptive and normative idealisations (Colyvan, 2013). We also need to scrutinise the

[15] Note that proponents of the Nitzan-Kelly Index may respond that they are interested in worst-case scenarios for manipulation. That is, they may present an argument that highlights value judgements they may make in addition to the evaluative dimension of manipulation.

normative idealisation in our model.

Returning to the broader discussion surrounding welfare concerns and egalitarian objections to strategic voting, we observed that normative evaluations regarding manipulation can differ. From a welfare perspective, the primary concern is the potential loss incurred from the overturning of election outcomes due to manipulation. In contrast, an egalitarian's primary concern is the unequal opportunities to influence election results. Our normative idealisation in the model could be, for example, that the only thing that normatively matters is the satisfaction of preferences, as expressed by the weak Pareto principle. Or alternatively we may use the normative idealisation that a difference in opportunity to influence the voting outcome are always problematic.

That means the normative assumptions about which part of manipulation tracks normative significance can diverge. If assumptions about what normatively matters about manipulation can diverge then so may what descriptive assumptions are appropriate in a model. To illustrate this consider that you are concerned with welfare. Then you may want your descriptive assumptions in your model to accurately reflect how often the outcome is overturned by strategic voting. For this, an acceptable descriptive idealisation might be to assume that everyone is equally skilled at voting strategically. After all, what you are not concerned with who is overturning the outcome.

Yet, assuming away the difference in skill in voting strategically would definitely not be an appropriate descriptive idealisation if you are concerned with the difference of opportunities to change the outcome (for an argument why this is the case see the previous chapter). In the next step, let me provide an example, of how a welfarist and an egalitarian might argue for each rule, respectively even when we fix several key descriptive assumptions about our target system.

The following is an illustrative example how a difference in normative commitments can give rise to an argument for following whether the Nitzan-Kelly Index or the computational complexity approach: A committee is formed, including both a philosophy department and a big technology firm. The question arises about how to vote within that committee. There are two voting rules: voting rule A and voting rule B. Voting rule A has a lower Nitzan-Kelly index but also lower computational complexity than voting rule B. To explore this in more detail, let us assume that nothing is known about preference distributions at the time of selecting the voting rule. Once there is an actual election, however, we assume that all agents have access to information about the preferences of other voters. Additionally, we assume that the process of determining optimal strategic votes is computationally costly, as suggested by computational complexity theory. Furthermore, let us assume that voters are unequally skilled in voting strategically. Despite fixing several key descriptive assumptions about our target system, there is an argument to be made for either metric based on how we may evaluate manipulation normatively. Let me provide an example of how a welfarist and an egalitarian might argue for each rule, respectively.

Here is an argument a welfarist might make for voting rule B: If a voting rule is NP-complete to manipulate, the computational difficulty of finding a successful strategic vote increases significantly. In practical terms, this means that even if voters possess full information about others' preferences, the time and resources required to compute a manipulative strategy may become prohibitive. As a result, fewer voters are likely to engage in strategic behaviour. Consequently, more voters will cast their ballots honestly, leading to outcomes that better reflect the true preferences of the electorate and, hence, the "honest" outcome. While fewer preference profiles that incentives at least one voter to cast a strategic vote are theoretically also positive, the computational complexity proves to be a better predictor than the Nitzan-Kelly index of how often we get an outcome that is the “honest” outcome. Thus, we should prefer voting rule B over voting rule A. It is less vulnerable to the harms of manipulation.

Here is an argument an egalitarian might make for voting rule A: voting rule A is easily manipulable and guarantees all members of the committee the ability to identify a strategic vote within a reasonable time frame, thus keeping the playing field relatively level. Conversely, voting rule B, which is hard to manipulate, poses challenges for both institutions but disproportionately favours the big technology firm, which has superior computational resources. Although the big technology firm is not guaranteed to find a strategy to misrepresent effectively, it still has a higher chance due to its greater computational power.[16] In effect, a higher computational complexity, under the concern of differing skills, increases the inequality in manipulability between the members of the committee. The difficulty of finding an effective strategy may lead to a lower expectation of submitted strategic votes, as both institutions are now not guaranteed to find a way to do so. Yet, it heightens the inequality of the opportunity to do so. While the Nitzan-Kelly index does not directly account for the inequality in strategic voting abilities among voters, a lower index reduces the upper bound on the number of opportunities for manipulation, thereby indirectly mitigating potential disparities. Thus, we should prefer voting rule A over voting rule B. It is less vulnerable to the harms of manipulation.

To sum up, manipulation is a thick concept. For a metric to have any

[16]This is the case because computational complexity only tells us that it would take a long time to *guarantee* finding a solution. Your odds still improve with the more computational resources as you can allocate more computational resources to trying to find a solution.

normative guidance, it needs to track a normative concern. There are different normative concerns about manipulation, which may make a difference in which metric provides normative guidance. Unfortunately, this has been overlooked by the literature so far. Merely a discussion about descriptive idealisations cannot resolve the question which metric provides normative guidance if there is potential for a difference in the normative evaluation of manipulation. The question now left to answer is, given our analysis, what should the literature on the manipulability of voting rules do? In the next section, I will outline the options available.

3.7 The Case for a Plurality of Metrics

In this section, I want to outline the different responses one could take to this issue and suggest that the best course of action is to develop multiple metrics to assess the manipulability of voting rules. We have established that manipulation via strategic voting can be evaluated along different values, leading to different objects of normative concern. There are several ways one could respond to the challenges that these different values may pose to metrics of manipulability:

1. The literature could demonstrate that differing value judgements about strategic voting do not translate into differences in measurement.

2. The literature could adopt one normative evaluation of manipulation, disregarding others.

3. The literature could develop several metrics that suit different normative evaluations of manipulation.

The first claim one could make is that, even though the objective of normative concern changes empirically, the objects of measurement remain correlated. One might argue that despite the diversity in value judgements concerning strategic voting, these differences do not necessarily translate into differences in how manipulability is measured. This approach suggests that different normative concerns about voting may correlate with the same empirical objects of measurement. Thus, adopting new metrics for every value framework might not be necessary. Similar to how descriptive models of voting do not require a new model for every slight variation in target systems, it is plausible that different evaluative frameworks could employ similar metrics.[17]

However, much like models that lose predictive accuracy when misaligned with reality, we must be cautious about using normative models that diverge too far from the values we care about. For instance, it may turn out that in our target system the that the Nitzan-Kelly Index is a good predictor of how often the election result is overturned and thus serves as a proxy for welfare concerns. Such a claim that that differing value judgements about strategic voting do not translate into differences in measurement is ultimately an empirical one, and further research would be needed to confirm it. Given the argument sketched above, I think it is questionable that this would be true of every possible value judgement.

In the absence of such an empirical argument, the second stance one can take is to adopt a value judgement on what makes manipulation via strategic voting harmful. In the absence of the possibility of a value-free stance, we cannot investigate the manipulability of voting rules without making such a choice. There are two issues with this approach. First, the object of dis-

[17] There are various ways to justify voting, many of which are supported by formal results about the outcomes that voting rules tend to produce. One such justification involves jury theorems, which suggest that voting can, in theory, be the best means to achieve outcomes where the result can be correct according to some standard (e.g., finding the most effective medical intervention for a patient). However, these theorems typically do not account for strategic voting. Assuming away the complexities of strategic voting can undermine the justification for voting, particularly when that justification relies on the idealised outcomes predicted by these theorems.

agreement is the evaluation of a collective decision mechanism itself. If we try to resolve these differences in value judgements by using a collective decision mechanism—such as a certain voting rule—we are essentially employing the very object under dispute to resolve a normative disagreement about its own evaluation. This introduces a circularity because we would be using a contested collective decision mechanism to justify that very collective decision mechanism. Instead of neutrally arbitrating between different value judgements, the mechanism already embodies certain normative assumptions, which means the dispute is being "resolved" by the mechanism we are trying to assess in the first place. Moreover, even if we were to ignore it, it seems a very strong claim to make that voting always triggers the same normative concerns. The same voting rule may be employed in political elections, hiring committees, and recommender systems. It would be peculiar if we thought that what justifies a certain voting rule is the same across all these applications.

The final, and in my view the most compelling, response is to develop multiple metrics that track different objects of normative concern. That is, we need a kind of value pluralism in the development of metrics that measure the vulnerability to manipulation in voting rules. This does not, of course, settle which values we engage with or which metrics we develop. This idea is supported by arguments for pluralism in other domains, such as well-being (Mitchell & Alexandrova, 2021) and social scientific indicators (Thoma, 2024). In this pluralistic approach, metrics would be designed to reflect the various value judgements about manipulation, capturing different dimensions of the issue. There are various future questions about how to implement value pluralism in measuring manipulability in voting rules. However, even before we are in a position to do this, there is much work to be done on the normative evaluation of strategic voting, which is understudied in philosophy and underappreciated in social choice theory. Developing a set of metrics that align with different normative concerns would offer a more comprehensive and nuanced understanding of voting rule manipulability.

In the absence of empirical support for the first option, and given the limitations of normative monism, metric pluralism emerges as the most viable path forward. By developing metrics that reflect various legitimising concerns about voting, we can ensure that the evaluative tools we use are robust and reflective of the complexity of manipulability in voting rules.

3.8 Conclusion

To sum up, manipulation is a thick concept, and as such, it brings with it all the usual complexities associated with measuring thick concepts. To determine what justifies moving from a lower index (or classification) to a normative evaluation of a target model, we must first understand why manipulation is considered a bad (or good) thing in the specific context. This understanding is crucial in the development of tools used to assess our institutional designs. If we fail to consider the full range of normative concerns that manipulation can raise, we risk creating assessment tools that only capture some concerns, leading to biased normative evaluations that favour those concerns simply because they are easier to measure.

I have demonstrated how this oversight has led to misguided discussions in the literature about measuring the vulnerability to manipulation in voting contexts. The Nitzan-Kelly Index and the computational complexity approach are among the most widely used metrics in the literature, but they can lead to different recommendations and rankings of voting rules. I have shown that the disagreement between these metrics cannot be resolved by merely pointing out the descriptive idealisations that each approach employs in its models. Instead, we need to understand the specific normative concerns at play and assess how well each approach can track those concerns.

As efforts to reduce or prevent manipulation continue, it remains difficult to determine whether these metrics are truly capturing what we value or whether they address the conflicts of interest that arise when choosing one method over another. To make progress, we need a clearer understanding of manipulation and the normative role it plays in different institutional designs.

4

A Heuristic Approach to Manipulation: Simulating Simple Strategic Voting

Abstract

The current approaches to studying strategic voting within social choice theory often maintain very strong assumptions about rationality, which consequently lead to implications about manipulation that are not necessarily generalisable across all applications of social choice theory. In light of this, the chapter aims to show that examining manipulation under the assumption of heuristic voters can overturn the normative assessment of vulnerability to manipulation. Through computer simulations, I demonstrate that more expressive voting rules (i.e., voting rules that allocate more votes to voters) do not necessarily lead to increased manipulation, contrary to the prevailing view in the literature. Instead, when we drop the assumption that agents maximise utility and adopt a satisficing approach, expressive voting rules can reliably result in almost no manipulation at all. Thus, the frequently cited argument against the implementation of more expressive voting rules, based on increased manipulability, rests on specific descriptive assumptions about voters.

4.1 Introduction

While it can be rational from an individual perspective to misrepresent preferences, from a social planner's perspective it is usually undesirable. Why this is the case can vary[1], but certainly one reason is that the significance of the result of an election may depend on voters having revealed their "true preferences".[2] Thus, the question of how vulnerable a voting rule is to manipulation becomes an important question and feeds into our evaluation of voting rules.

There is a variety of different voting rules to pick from, therefore when we need to choose one, we typically want to establish two things. First, that the voting rule aggregates the "right" subset of individual preferences. Different voting rules use different information of the preference rankings of the voters, and there can be reasonable dispute over which information is appropriate to use for aggregation. For instance we may think that we should include the information of which option is disliked the most as it matters that the elected option is not too polarising. However, one could also argue that what solely matters is giving the majority their most preferred option (i.e., majority rule).[3] Second we usually want that the voting rule satisfies desirable normative properties (e.g., Pareto Principle, Non-dictatorship). In this context, the property that relates to our worry about strategic voting is called strategy-proofness. Let me remind you that for a voting rule to be strategy-proof, there has to be no preference profile at which the voting

[1]For a more in depth discussion see chapter 2 and chapter 3.

[2]See Bernheim (2016) for a sceptical take on the idea that agents have "true preferences". Moreover it is well known that choices may not reveal preferences if agents have false beliefs about the available options (see for instance Hausman, 2000). A more detailed discussion of these issues is outside of the scope of this chapter. We assume that the private information about valuation in combination with the agent's other beliefs (true or false) are what we intend to aggregate and thus what we need agents to express.

[3]For more examples which considerations may play a role to determine which information we consider relevant see Risse (2009).

rule is manipulable by some individual j. Manipulation here means that if j misrepresents her preference the voting rule chooses an alternative that j strictly prefers to the alternative that would win if j submitted her true preference. Since, strategy-proofness is not easy to obtain and conflicts with a bundle of other desirable properties of voting rules the literature has worked out more explicitly which trade-offs are required to establish strategy-proof voting rules. For instance for a wide class of voting rules to be strategy-proof some form of monotonicity (i.e., receiving more support from the voters is always better for an option) and Independence of Irrelevant of Alternatives (i.e. the group outcome over any two alternatives x and y depend only on the individual preferences between x and y) need to be satisfied (see Sengupta, 1980). Another branch of the literature proposed different desiderata that are suitable once one makes some additional assumptions about the voters. Voters may not be able utilise the opportunity to manipulate due to a lack of information (e.g., Myerson, 1979) or complexity boundaries (Bartholdi III et al., 1989).

In what follows, I will assume that the primary concern of manipulability is the overturning of the election result and examine the relationship between more expressive voting rules and the likelihood of such overturning.[4] See chapter 3 for a discussion why making normative commitments underlying the evaluation of manipulability is essential to formulate a measure thereof. To examine the relationship between more expressive voting rules and the likelihood of overturning the election result, it is necessary to introduce certain assumptions regarding voter behaviour. In the existing literature, game-theoretic models are commonly employed to capture how voters strategise and form beliefs based on incomplete information. While game theory provides a useful framework for modelling strategic interactions,

[4]For why this may be a reasonable concern in the context of strategic voting see chapter 3.

it may not fully capture the behaviour of all voter types in every context. Traditional game-theoretical models rely on assumptions such as common knowledge of rationality, cardinal rankings, and utility maximisation. A whole body of experimental data is now available that suggests that people do not typically act in accordance with such game-theoretical predictions but are rather "boundedly" rational. For an overview of the experimental literature both confirming and disagreeing with analytical results from game theory see Camerer (2011).

In the light of this, this chapter aims at starting with behavioural assumptions that are more directly informed by the empirical literature in political science. It is well established in the political science literature that voters make use of heuristics to deal with the complex information and decisions voting accompanies (see Popkin, 1995; Fiorina, 1981; Lupia, 1994; Sniderman, Brody, & Tetlock, 1993; Dancey & Sheagley, 2013; Baldassarri & Schadee, 2006; McDermott, 2005; Lau & Redlawsk, 2006). Especially voting with multiple votes can become cognitively complex fast (Darcy & Marsh, 1994) and is hence prone to voters falling back on simpler rules. Following the empirical literature, the model presented in this chapter focuses on a heuristic meant to capture the fact that voters may rather satisfice than maximize (Simon, 1955). This assumption has some claim to central importance because it is at the core of many heuristics as well as arguments for bounded rationality (Gigerenzer & Gaissmaier, 2011; Aumann, 1997; Wheeler, 2020). Satisficing here means that people look for an option that is good enough — defined by their aspiration level — and then stick with the first option which is above that level. In this chapter I work with a model simulating heuristic voters in repeated elections, showing that more expressive voting rules (i.e., voting rules that give more votes to voters) do not usually lead to a higher frequency of the election result being overturned. Many approaches in the literature look at the manipulability of voting rules (Saari, 1992; Nitzan, 1985; Kelly, 1993), inferring that a voting rule is more susceptible to manipulation based on the increased combinatorial possibilities of overturning the election outcome. In what follows, manipulation is defined as the frequency with which strategic voting alters the election outcome compared to the result that would emerge if all voters cast sincere ballots. Thus, the normative concern extends beyond the mere existence of manipulation, focusing instead on how often the election outcome fails to reflect the electorate's true preferences. To evaluate this, a model of individual voter decision-making is required.

The results of the model demonstrate that introducing satisficing as a behavioural assumption can significantly alter the evaluation of voting rules that allow voters to cast a relatively large number of votes. In particular, the findings challenge the existing body of literature in social choice theory (Saari, 1992), which suggests that more expressive voting rules (i.e., voting rules that allocate more votes to voters) typically result in increased manipulation. Once the assumption of rational maximisation is abandoned in favour of a satisficing approach, expressive voting rules can reliably lead to minimal manipulation (i.e., overturning the "sincere" election result becomes rare). This is crucial, as it has practical implications for which voting rules should be selected. If we believe that voters are more likely to satisfice than maximise, it may be that previous formal analyses of strategic voting fail to capture what we ought to be concerned with—namely, reducing the likelihood of overturning the "sincere" election result.

4.2 Simple Strategic Voting - The Model

4.2.1 The Model in Relation to Previous Literature

While there is a substantial body of literature on voter heuristics, much of it focuses on heuristics related to information processing or those that structure preferences over candidates. Comparatively less research has been devoted to understanding the heuristics involved in the actual process of casting votes. To highlight a few exceptions, Meffert and Gschwend (2007) argue that voters may cast an insincere vote if they want to keep a party over a certain threshold under which they do not find representation in parliament even though this party does not rank first in the voter's preference ranking. Fiorina (1981) argues that if voters are satisfied with the candidates in power, they keep voting for them and otherwise shift their votes for the opposition.

The use of simple rules as strategies is also familiar in evolutionary game theory where a whole population of agents often executes certain strategies which sometimes can be described as heuristics in the above-mentioned sense. In the process of applying game theory to biology, assumptions such as ideally rational agents have been modified to better fit non-humans. In order to meet critiques of traditional game theory these changes were later imported back to economics and other social sciences to better model dynamic games and bounded rationality (see Hofbauer & Sigmund, 2003; D. Friedman, 1991). There are also a few applications of evolutionary game theory to voting. Yet they tend to concentrate on voting participation (Sieg & Schulz, 1995; Conley, Toossi, & Wooders, 2006). In the context of evolutionary game theory, a heuristic that can capture aspiration levels is called Win-Stay Loose-Switch (WSLS). WSLS is a simple, yet effective strategy known for outperforming Tit for Tat in repeated prisoner dilemmas with a noisy environment (Nowak

& Sigmund, 1993). WSLS is a variation of a so-called trigger strategy with a memory of one. The model in this chapter will concentrate on repeated voting with voters using variations of WSLS. Agents using a WSLS will continue to perform an action if the previous outcome was successful and switch their action if the outcome was not. What is considered successful, or a "win" can vary depending on their aspiration level. There is evidence that WSLS fits a lot of behaviour outside of voting quite well. That is, in certain circumstances it can better describe how agents behave than more complex decision mechanisms such as maximisation or more complex learning mechanisms such as reinforcement learning (see Otto, Taylor, & Markman, 2011; Worthy, Hawthorne, & Otto, 2013; Wang, Xu, & Zhou, 2014). Therefore, WSLS is one of the simplest plausible strategies of voting which deviates from sincere voting and incorporates aspiration levels and, thus, will serve as a starting point in this chapter to study the effects of satisficing as a behavioural assumption in voting.

4.2.2 Interpretation of the Model

Voting is often modelled as a one-shot game, and consequently the analysis of strategy-proofness is usually formulated in terms of a one-shot game. However, repeated voting has a couple of interesting interpretations and gives time for voters to learn over time. For instance, elections that may appear to be isolated incidents can be viewed as repeated events, assuming stable preferences. Alternatively, some voting processes involve a dynamic expression of opinions on options, coupled with the updating of beliefs about how other voters will behave. The following examples illustrate scenarios that can be modelled using the mathematical framework introduced in the next section. One interpretation of repeated voting under a stable preference profile is a recurring event where preferences over options remain unchanged. For example, consider repeatedly going to lunch with the same group of people (e.g., at a conference) and voting to decide the venue. If you have a gluten intolerance or follow a vegan diet, repeated interactions would reflect your satisfaction with restaurant choices that accommodate these dietary preferences. Similarly, consider elections over a lifetime where a population votes based not on specific candidates but on ideological alignment. In this case, the model may describe partisan voters who consistently support their party (or parties) in repeated elections, without concern for the individual candidate representing the party.

Another noteworthy interpretation of repeated voting is that election-labelled events can capture the dynamics of belief formation leading up to a single election. In this case, the model reflects only one actual election, but repeated voting illustrates how voters adjust their decisions based on periodic polling data. Thus, repeated voting shows how polls influence the final outcome, as voters base their decisions in the upcoming election on the latest available poll, which in turn is shaped by prior polling results. To illustrate how an individual voter may vote, imagine there being three parties A, B and C. You have the following preference order over these parties $A \succ B \succ C$. You would be satisfied if either A or B would be elected. If asked who you would vote for if this Sunday were the election you say A. On Monday, you learn that if the previous Sunday's election had been held, candidate C would have won. When asked again who you would vote for if the election were this coming Sunday, you decide to switch your vote to B. The following Monday, you read that if the election had taken place the previous Sunday, candidate A would have won. You continue voting for B until C wins again, at which point you switch your vote once more. With these possible interpretations in mind, let us go into more details regarding the model.

4.2.3 A Closer Look at Voter Behaviour

To test whether adding votes will lead to more manipulation the model will use different variations of k-approval voting rules. That is voters cast all together k votes with the limitation that they can at most cast one vote per option.

Voters in this model will vote one hundred times. In their first round, all voters will start with voting sincerely, reflecting a disposition to vote sincerely if not given a reason to do otherwise. All voters will vote simultaneously and, after each round, receive information on which option was selected. Then based on whether they are satisfied with that chosen option or not, they will change the selection of options they vote for. More precisely in this model, they will stop voting for their lowest-ranked option they previously voted for and cast that vote randomly to someone else. This, of course raises the question why we should model voters to switch their votes in this way. It may seem counterintuitive that, in the scenario above, someone might vote for candidate C, even when they rank C the lowest. However, there are situations where doing so could be advantageous, making it difficult to dismiss the possibility of voting for one's least preferred option. This approach to modelling vote changes offers one distinct advantage: it prevents voters, after several iterations, from failing to vote for any candidate with whom they are satisfied. Additionally, it ensures that voters experiment with votes that reflect their least preferred choices. Prima facie, although not directly implied by the model, one might speculate that voters possess a disposition to vote at least partially sincerely when possible. While it would be valuable to explore different methods of vote selection and casting in an extension of the model, this is unfortunately beyond the scope of this chapter. A key assumption tested later in this chapter is the extent to which voters may change their vote.

Let me illustrate how these behavioural assumptions play out with several voters in a simple example. Assume that there are 4 options A, B, C, and D, and 4 voters with the following strict preferences: [5] Assume that there are 4 options A, B, C, and D, and 4 voters with the following strict preferences:

Rankings	**Number of Voters**
$A \succ B \succ C \succ D$	2
$B \succ C \succ D \succ A$	1
$D \succ B \succ A \succ C$	1

Table 4.1: Voter preferences and rankings

Additionally, assume, for this example, that the voters only consider it a "win" if their favourite option is elected. The tables below illustrate the change from the first to the second round of voting for the above preferences. For simplicity, the table includes only three versions of k-approval voting rules: Plurality ($k = 1$), Vote-for-Two ($k = 2$), and Minus (N-1) with N describing the number of options.

Generally, to cast $N - 1$ votes is equivalent to casting a vote against one option. That is, for four options whether A, B, and C all receive one vote and D none, or A, B, and C receiving no vote and D a negative vote, is mathematically equivalent. Thus, instead of 3 positive votes in the case of four options, Minus is illustrated by one negative vote.

The table below shows how many votes an option receives in the first round when everyone votes sincerely and the expected votes an option receives in the second round. The expected votes are broken down into multiple parts i.) the votes an option will receive for sure, ii.) the number of votes which can potentially be cast for the option, and iii.) the chance for a vote to be cast on that option. While the expected number of votes can include fractions, the actual resulting votes can only be an integer. The next two columns in the table show the expected change of the elected option and the possible

[5] See `https://github.com/belewo/WSLS_voting` for supplementary material for the following simulation work.

changes, respectively.

Voting Rule	**Round 1**	**Round 2**	**Expected Change of Winner**	**Possible Winner**
Plurality	A: 2 B: 1 C: 0 D: 1	A : 2 + 2 * $\frac{1}{3}$ B : 0 + 1 * $\frac{1}{3}$ C : 0 + 2 * $\frac{1}{3}$ D : 0 + 1 * $\frac{1}{3}$	A $\rightarrow$ A	A, C
Vote-For-Two	A: 2 B: 4 C: 1 D: 1	A : 2 + 1 * $\frac{1}{2}$ B : 1 C : 1 + 3 * $\frac{1}{2}$ D : 1 + 2 * $\frac{1}{2}$	B $\rightarrow$ A/C	A, C, D
Minus	A: -1 B: 0 C: -1 D: -2	A : -1 - 3 * $\frac{1}{3}$ B : 0 - 3 * $\frac{1}{3}$ C : 0 - 2 * $\frac{1}{3}$ D : 0 - 1 * $\frac{1}{3}$	B $\rightarrow$ D	A, B, C, D

Table 4.2: Example: Change of Votes

For Plurality, we can see that even though A wins in the first round and is also expected to win in the second round C could also win in the second round. In contrast, while D could win under Vote-For-Two the expected winner in the second round is A or C. While B could get at most 1 vote and, hence, would always lose against A who is guaranteed 2 votes in round 2, C could potentially get 4 votes and A could get 3 votes. That means that the winner that would have won under sincere voting will certainly not be elected in the second round under Vote-For-Two, while for Plurality the sincere winner is expected to win in the second round. Under Minus, the sincere winner is B and B could possibly win in the second round. However, the expected winner is A and not B. Thus, the sincere winner is not expected to win but possibly could.

4.3 Results

In the above table, we assumed that voters care only about their top-ranked option and observed significantly different expectations across voting rules regarding whether the sincere winner 'survives' the first round. Let us see how these results hold when we draw from random preference profiles. The below table shows the average results from 500 different preference profiles for which every voter votes 100 times respectively. If we run the simulation for 100 voters and draw from a uniform distribution over all preference profiles we see that Plurality is almost always electing the sincere winner, while Vote-For-Two is in over half the cases electing an option that wouldn't have won if everyone had voted sincerely. Minus does considerably better than Vote-For-Two but is still far away from almost always electing the sincere winner as Plurality does.

Yet, these results change once we assume that voters are not just satisfied with their top-ranked option but also with their second-ranked option. Then, Vote-For-Two and Plurality both almost always elect the sincere winner, and Minus improves significantly as well. Interestingly, Vote-For-Two had the worst percentage of electing the sincere winner and now ties for the best percentage. This already shows that it matters which satisfaction threshold we ascribe to voters.

Voting Rule	**Percentage of Sincere Winner Elected**
Plurality	99.8
Vote-For-Two	44.3
Minus	87.4

Table 4.3: Satisfaction Threshold 1: k-Approval and percentage of sincere winner elected

If we shift the threshold of satisfaction one option further, i.e. voters are satisfied with anyone as long as their last-ranked option is not chosen, Minus

Voting Rule	**Percentage of Sincere Winner Elected**
Plurality	99.9
Vote-For-Two	99.9
Minus	100

Table 4.4: Satisfaction Threshold 2: k-Approval and percentage of sincere winner elected

always elects the sincere winner. These results hint at a connection between

Voting Rule	**Percentage of Sincere Winner Elected**
Plurality	98.6
Vote-For-Two	99.9
Minus	100

Table 4.5: Satisfaction Threshold 3: k-Approval and percentage of sincere winner elected

the satisfaction levels of voters and the effects of insincere voting.[6] In our case, k-approval voting would reflect the number of votes each voter wishes to cast regardless. However, note that not all voting rules elect the sincere winner more often once voters are satisfied with more options. Plurality elects the sincere winner less often if voters are satisfied with any of their top three ranked options. Generally, it makes sense to expect a tendency that voting rules are more stable, i.e., electing more often the sincere winner, as the satisfaction threshold increases. Hence, we should also expect that the differences in the percentage of electing the sincere winner between voting rules tend to shrink.

For instance, for satisfaction threshold 1, the percentages of electing the sincere winner ranged from 99.8% (Plurality) to 44.3% (Minus). In the third condition, with the satisfaction threshold set to 3, the difference between the

[6]Such a connection is not novel in the literature. When voters' preferences are dichotomous—i.e., they divide the alternatives into two indifference classes: preferred and not preferred—approval voting is strategy-proof, giving each voter an incentive to vote for precisely their preferred alternatives (Brams & Fishburn, 1978)

best and worst-performing voting rules is merely 1.4% (Plurality 98.6% and Minus 100%). This makes sense as the set of options a voter is satisfied with grows as the threshold increases. Thus, we expect voters to be generally more satisfied with the result of the election. Yet, this does not automatically mean that every voting rule will elect the sincere winner more often as the threshold increases. As we see with Plurality, it performs best when the satisfaction threshold is set to one (99.9%) and elects the sincere winner less often as the set of options a voter is satisfied with grows (99.8% and 98.6%).[7]

Consequently, we might want to consider another factor that drives the percentage of the sincere winner up for Vote-For-Two and Minus. Under the same preference rankings, voters are satisfied with different sets of options if the threshold of satisfaction is changed. If voters are only satisfied with their first ranked option, then under sincere voting Plurality picks out whichever option is most often first ranked. Hence, in the first round the biggest chunk of voters will be satisfied with the elected winner and thus the biggest chunk of voters is not incentivised to change to insincere voting. For four options that means that at least a quarter of the voters will not change their vote. Given that voters are satisfied with either their first ranked option or their second-ranked option, Vote-For-Two elects the option which is most often in the first or the second rank among all voters, i.e., in the case of four options at least 50 percent will not want to switch to insincere voting in the second round. Also, under Plurality the number of the voters that are satisfied with the winner of the first round increases but it will be somewhere between a 25 percent of the voters and up to the percentage that are satisfied with the Voting-For-Two sincere winner. Yet, the number of voters that are satisfied with the Plurality sincere winner can never exceed the amount of voters that are satisfied with the Vote-For-Two sincere winner (note that sincere winner

[7]Note, however, that running these simulations was very computationally demanding. The results presented are thus averages over only 500 runs, and the exact figures should not be interpreted too strictly.

can be the same). Hence, every k-approval voting rule has a "matching" variant of the WSLS heuristic, for Plurality (k=1) it is the WSLS heuristic with the satisfaction threshold 1 for Vote-For-Two (k=2) the WSLS heuristic with the satisfaction threshold 2, and for Minus (k=N-1) the WSLS heuristic with the satisfaction threshold (N-1).

To demonstrate this hypothesis let us scale up the simulation. We now test elections with 5, 6, 7, 8, 9, and 10 options and include k-approval voting rules with k=1, k=2, k=3, k=4, and k=N-1. Again, the population will consist out of 100 voters whose preference rankings are generated from an impartial culture. If we fix the satisfaction threshold at 1 for the population, so that every voter is merely satisfied with their first-ranked option we get the following result. Note that the category k=4, N=4 would trivially always be 100 percent, as everyone is always satisfied with every option. Instead of putting 100 percent in there, this category is denoted by a "/". Moreover, note that k=3 and k=N-1 and k=4 and k=N-1 under N=4 and N=5 are the same respectively.

k-Approval	**N=4**	**N=5**	**N=6**	**N=7**	**N=8**	**N=9**	**N=10**
k=1	99.9	100	100	99.8	100	100	100
k=2	44.8	37.8	31.9	30.0	16.4	28.6	16.1
k=3	87.4	31.1	28.7	25.7	18.5	26.3	36.0
k=4	/	70.9	24.6	30.2	36.9	29.3	47.9
k=N-1	87.3	70.9	53.0	20.8	13.5	12.7	10.9

Table 4.6: Satisfaction Threshold 1: k-Approval and percentage of sincere winner elected

We can see that Plurality is by far the most stable (i.e. most often elects the "sincere" winner) voting rule and almost always elects the sincere winner. The percentage of electing the sincere winner ranges from 10.9 % (k=N-1, N=10) to 100 % (k=1, N=5,6,8,9,10). However, as suggested by our example from above these results change quickly once we change the satisfaction threshold to 2.

k-Approval	N=4	N=5	N=6	N=7	N=8	N=9	N=10
k=1	99.9	100	100	97.0	99.9	98.2	100
k=2	100	100	99.9	100	100	99.5	100
k=3	96.3	34.2	32.4	23.7	31.5	36.6	31.9
k=4	/	97.8	36.7	37.5	22.4	27.4	63.4
k=N-1	96.3	97.8	91.8	83.5	36.1	47.3	32.3

Table 4.7: Satisfaction Threshold 2: k-Approval and percentage of sincere winner elected

With the satisfaction threshold set to 2 we see that all voting rules elect more often the sincere winner with the exception of Plurality. However, the decrease in percentage is slight. The difference between the voting rules shrinks now from 31,5 (k=3, N=7) and 100 (k=1, N=5,6,10; k=2, N=4,5,7,8,10). Most notably, now Vote-For-Two is electing most often the sincere winner, while Plurality is still close but elects less often the sincere winner. Below are the results once we move the satisfaction threshold to 3, 4, and N-1 meaning that voters are satisfied with anyone who they rank second and above, third and above, and so on.

k-Approval	N=4	N=5	N=6	N=7	N=8	N=9	N=10
k=1	100	99.3	99.1	99.3	99.3	99.5	99.2
k=2	100	100	100	99.3	100	99.8	100
k=3	100	100	100	100	100	100	99.3
k=4	/	97.2	43.9	49.1	48.6	46.8	40.3
k=N-1	100	97.2	98.3	95.8	93.7	86.5	81.1

Table 4.8: Satisfaction Threshold 3: k-Approval and percentage of sincere winner elected

We can now confirm that there is a tendency for all tested voting rules to become more stable once the satisfaction threshold includes a bigger sub-

k-Approval	N=4	N=5	N=6	N=7	N=8	N=9	N=10
k=1	100	100	99.9	99.3	98.7	99.5	99.4
k=2	100	100	100	99.2	99.6	99.8	99.2
k=3	100	100	100	98.3	99.9	100	100
k=4	/	100	100	100	100	100	100
k=N-1	100	100	99.5	97.1	93.0	90.2	91.6

Table 4.9: Satisfaction Threshold 4: k-Approval and percentage of sincere winner elected

k-Approval	N=4	N=5	N=6	N=7	N=8	N=9	N=10
k=1	98.4	100	99.6	94.4	96.9	97.9	97.8
k=2	99.8	100	99.9	99.8	97.1	99.8	99.2
k=3	100	98.8	97.4	97.1	99.9	98.8	98.7
k=4	/	100	100	98.2	98.5	99.9	99.3
k=N-1	100	100	100	100	100	100	100

Table 4.10: Satisfaction Threshold N-1: k-Approval and percentage of sincere winner elected

set of voters ranking. Yet, no voting rule always delivers the best result in terms of electing the sincere winner. Our other hypothesis that there is a meaningful relation between voting rules picking up the "relevant" subset of options and these voting rules being stable seems to hold as well. More so, it seems that the voting rules that allow voters to exactly vote for the subset of options they are satisfied are most often elects the sincere winner. An explanation for this seems to be that if voting rules elect the options most voters are satisfied with, this reduces the number of voters that are switching to misrepresenting preferences.

One could criticise that in the results presented above the voters are only allowed to change one vote of their available votes and one would expect different results for the voting rules that allow for relatively many votes if voters could change more than one vote. While a certain limitation of votes changing might be psychologically plausible, there is not much reason to assume why that limitation should be at one vote. Hence the next result repeats the simulation under the same conditions as from above but voters change two votes instead of one vote if their voting rule permits them to. That means that if voters are unsatisfied with the option elected, they will select among the options they voted for last time, the two options they rank lowest and cast them randomly to other options.

k-Approval	**N=4**	**N=5**	**N=6**	**N=7**	**N=8**	**N=9**	**N=10**
k=1	100	100	100	100	100	100	99.7
k=2	100	99.7	98.1	98.9	98.9	99.1	96.8
k=3	97.2	25.4	23.3	32.6	25.9	20.2	96.8
k=4	/	93.3	34.8	21.8	14.7	14.2	29.8
k=N-1	97.2	93.3	69.6	39.4	20.3	15.9	12.2

Table 4.11: Satisfaction Threshold 1 II: k-Approval and percentage of sincere winner elected

k-Approval	**N=4**	**N=5**	**N=6**	**N=7**	**N=8**	**N=9**	**N=10**
k=1	100	99.9	100	100	100	100	100
k=2	100	100	100	100	100	100	100
k=3	100	98.6	98.2	98.7	98.3	98.2	98.4
k=4	/	99.3	18.4	14.8	17.1	17.6	28.2
k=N-1	100	99.3	94.0	89.2	80.1	67.5	52.9

k-Approval	**N=4**	**N=5**	**N=6**	**N=7**	**N=8**	**N=9**	**N=10**
k=1	100	100	100	100	100	99.7	100
k=2	100	100	100	100	100	99.7	100
k=3	100	100	100	100	100	100	100
k=4	/	100	98.5	98.9	98.5	98.6	98.1
k=N-1	100	100	93.2	91.2	82.6	80.2	79.9

Table 4.14: Satisfaction Threshold 3 II: k-Approval and percentage of sincere winner elected

Table 4.12: Satisfaction Threshold 3 II: k-Approval and Percentage of Sincere Winner Elected

k-Approval	**N=4**	**N=5**	**N=6**	**N=7**	**N=8**	**N=9**	**N=10**
k=1	100	100	100	100	100	99.7	100
k=2	100	100	100	100	100	99.7	100
k=3	100	100	100	100	100	100	100
k=4	/	100	98.5	98.9	98.5	98.6	98.1
k=N-1	100	100	93.2	91.2	82.6	80.2	79.9

Table 4.13: Satisfaction Threshold 2 II: k-Approval and percentage of sincere winner elected

k-Approval	**N=4**	**N=5**	**N=6**	**N=7**	**N=8**	**N=9**	**N=10**
k=1	99.8	99.7	98.6	99.6	98.1	99.2	100
k=2	100	100	100	100	100	99.8	100
k=3	100	100	100	100	100	100	100
k=4	/	100	100	100	100	100	100
k=N-1	100	100	98.8	97.9	94.7	89.0	85.5

Table 4.15: Satisfaction Threshold 4 II: k-Approval and percentage of sincere winner elected

k-Approval	**N=4**	**N=5**	**N=6**	**N=7**	**N=8**	**N=9**	**N=10**
k=1	99.5	100	98.6	95.9	98.2	95.1	96.4
k=2	100	100	100	99.5	99.3	98.1	99.9
k=3	100	99.6	99.3	98.5	100	98.7	99.5
k=4	/	100	100	98.4	98.7	99.2	96.2
k=N-1	100	100	100	100	100	100	100

Table 4.16: Satisfaction threshold N-1 II: k-Approval and percentage of sincere winner elected

While generally, for some of the voting rules the percentage changes with which they elect the sincere winner certain key lessons do not. Hence, we may conclude with the following three key lessons from these simulations:

Lesson 1: No k-approval voting rule is always, i.e., across all satisfaction thresholds, most stable or always least stable.

Lesson 2: No (non-trivial k = N) satisfaction threshold makes every k-approval voting rule most stable to strategic voting.

Lesson 3: The combination of satisfaction threshold and k-approval voting rule counts. Each satisfaction threshold has a matching k-approval voting rule that is most stable for that satisfaction threshold. Moreover, the percentage of electing the sincere winner is about the same for each matching pair.

A natural question emerging from the simulation results concerns their generalisability. It is expected that the proportion of elections in which the sincere winner is selected will vary with the distribution of preference profiles under a given voting rule. However, the qualitative insight appears robust: specific satisfaction thresholds correspond to particular k-approval voting rules. This correspondence holds because each satisfaction threshold has a matching k-approval rule that selects the sincere winner who satisfies the greatest number of voters at that threshold. This relationship is independent of other assumptions that might be introduced. Whether this matching rule consistently emerges as the most stable option likely depends on the preference profile distribution and the strategies voters use to switch their votes. Nevertheless, the fact that a significant proportion of voters will not seek to switch their votes in the initial round suggests that such a rule would generally exhibit a degree of stability and thus elect more often the sincere winner.

4.4 Conclusion

When selecting a voting rule, we typically aim to ensure that the chosen rule satisfies desirable normative properties. If lesson 3 generalises beyond this simulation, there is an interesting connection between aggregating the "right" subset of individual preferences and minimising successful strategic voting. Since voters operate with satisfaction thresholds and are not maximisers, it seems reasonable to argue that the appropriate subset to aggregate consists only of the options that voters are satisfied with. This approach maximises the number of voters satisfied with the election outcome. The simulation results show that when a voting rule aggregates these satisfaction-based preferences, it consistently minimises manipulation compared to other k-approval voting rules.

Hence, I propose that an interesting takeaway from this simulation is that, to the extent we believe satisfaction is linked to strategic voting, if a voting rule elects an option that many voters are satisfied with, we can expect fewer instances of strategic voting. This conclusion is based on the assumption that satisfaction levels and strategic voting are related, though not necessarily as closely as in this model. The model discussed in this chapter has several limitations, and there is ample room for extension. For example, one could introduce varied satisfaction thresholds to allow for more sophisticated strategic voting, where voters cast votes randomly but not uniformly across their options. Additionally, further exploration of how voters might change their votes could be pursued. For instance, one approach could involve voters casting votes randomly but always selecting their highest-ranked option. Different information states could also be introduced, with voters reacting to these states in varied ways. Beyond that, one could employ entirely different classes of heuristics or adjust the number of strategically cast votes. While the real-world dynamics are more complex than those captured in this model, the key insight is that if a voting rule captures meaningful compromise, the good news is that we can achieve greater stability against strategic voting at a relatively low cost. In other words, this stability may come more easily if satisfaction and strategic voting are indeed linked.

Generally, I take the results of these simulations to highlight the importance of considering empirical evidence when studying manipulation in order to make descriptively adequate assumptions about voters. The results show that the widespread assumption that more votes lead to a higher frequency of the election result being overturned does not necessarily hold. To analyse the threat of manipulation, we need an accurate model of voter behaviour. Therefore, recommendations for voting rules should potentially vary on grounds of the election context, the type of voters. Only by addressing what constitutes a descriptively accurate model of voter behaviour can we formulate sound normative recommendations concerning manipulation. The previous chapter underscored the significance of normative assumptions in assessing the vulnerability to manipulation. This chapter aims to stress the importance of descriptive assumptions in this endeavour.

5
Conclusion

In conclusion, this thesis seeked to bridge critical gaps in the applications of the verdicts of social choice models. By examining the interplay between individual decision-making processes and the design of collective decision mechanisms, the thesis underscored the necessity of aligning social choice models more closely with the complexities of their target systems. The investigation reveals that abstracting away from individual behavioural does not always enhance the generality of models; instead, it can render them inapplicable or even misleading when applied to concrete social systems. Another key insight argued for across the chapters of this thesis is that the normative assumptions in social choice models require careful scrutiny, as they shape both the design and evaluation of collective decision-making mechanisms.

The thesis demonstrated the need for ongoing philosophical engagement with social choice models, not only to refine the models but also to ensure that they continue to serve as effective tools for guiding collective decision-making. By integrating insights from philosophy of science, economics, computer science, psychology, and political science, this thesis not only critiques existing models but also proposes pathways for developing more robust and applicable social choice mechanisms. It calls for a more nuanced approach that carefully considers the underlying assumptions about individual voters and the normative goals of collective decision-making processes.

Future research could explore the impact of new descriptive and normative assumptions employed in the various emerging fields that apply social choice models, such as computational social choice, artificial intelligence, and multi-agent systems. These disciplines bring fresh perspectives on individual behavioural, often involving algorithmic agents, and decentralised systems that challenge traditional assumptions of voter rationality. Additionally, extending the analysis of manipulability beyond voting rules to other collective decision mechanisms—such as matching markets, auctions, and cooperative games—offers a rich avenue for exploration.

References

Abdulkadiroglu, A., Pathak, P. A., Roth, A. E., & Sönmez, T. (2006). Changing the boston school choice mechanism..

Abizadeh, A. (2021). Counter-majoritarian democracy: Persistent minorities, federalism, and the power of numbers. *American Political Science Review*, *115*(3), 742–756. doi: 10.1017/S0003055421000375

Aldrich, J. H., Blais, A., & Stevenson, R. T. (2018). *Many faces of strategic voting: Tactical behavior in electoral systems around the world*. University of Michigan Press. doi: 10.3998/mpub.9914229

Aleskerov, F., & Kurbanov, E. (1999). Degree of manipulability of social choice procedures. In *Current trends in economics: Theory and applications* (pp. 13–27).

Alexandrova, A. (2018). Can the science of well-being be objective? *British Journal for the Philosophy of Science*, *69*(2), 421–445.

Allen, A. (2022). *Feminist perspectives on power.* The Stanford Encyclopedia of Philosophy (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.). Retrieved from `https://plato.stanford.edu/archives/fall2022/entries/feminist-power/`

Allen, M. R., Babiker, M., Chen, Y., de Coninck, H., Connors, S., van Diemen, R., ... others (2018). Summary for policymakers. In *Global warming of 1.5: An ipcc special report on the impacts of global warming of 1.5\c above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response* *to the threat of climate change, sustainable development, and efforts to eradicate poverty.* IPCC.

Alvarez, R. M., Boehmke, F. J., & Nagler, J. (2006). Strategic voting in british elections. *Electoral Studies*, *25*(1), 1–19. doi: 10.1016/j.electstud.2005.05.003

Anderson, E. S. (1999). What is the point of equality? *Ethics*, *109*(2), 287–337. doi: 10.1086/233897

Arrow, K. J. (1951/1963). *Social choice and individual values.* New York: Wiley.

Arvan, M., Bright, L. K., & Heesen, R. (2022). Jury theorems for peer review.

Augustin, T., Coolen, F. P., De Cooman, G., & Troffaes, M. C. (2014). *Introduction to imprecise probabilities.* John Wiley & Sons.

Aumann, R. J. (1997). Rationality and bounded rationality. In *Cooperation: Game-theoretic approaches* (pp. 219–231). Springer.

Austen-Smith, D., & Banks, J. S. (1996). Information aggregation, rationality, and the condorcet jury theorem. *American political science review*, 34–45.

Baldassarri, D., & Schadee, H. (2006). Voter heuristics and political cognition in italy: An empirical typology. *Electoral Studies*, *25*(3), 448–466.

Barberà, S., Gul, F., & Stacchetti, E. (1993). Generalized median voter schemes and committees. *Journal of Economic Theory*, *61*(2), 262–289.

Barberà, S., & Jackson, M. O. (1992). Strategy-proof exchange..

Barberis, N., & Thaler, R. (2003). Chapter 18 a survey of behavioral finance. In *Financial markets and asset pricing* (Vol. 1, p. 1053-1128). Elsevier. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1574010203010276` doi: https://doi.org/10.1016/S1574-0102(03)01027-6

Barberà, S. (2011). Strategy-proof social choice theory. In *Handbook of* *social choice and welfare* (Vol. 2, pp. 731–831). Elsevier. doi: 10.1016/S0169-7218(10)00016-0

Barry, B. (1980). Is it better to be powerful or lucky? part 2. *Political Studies*, *28*(3), 338–352. doi: 10.1111/j.1467-9248.1980.tb01244.x

Bartholdi III, J. J., & Orlin, J. B. (1991). Single transferable vote resists strategic voting. *Social Choice and Welfare*, *8*(4), 341–354. doi: 10.1007/BF00754353

Bartholdi III, J. J., Tovey, C. A., & Trick, M. A. (1989). The computational difficulty of manipulating an election. *Social Choice and Welfare*, *6*(3), 227–241.

Beck, L., & Jahn, M. (2021). Normative models and their success. *Philosophy of the Social Sciences*, *51*(2), 123–150.

Berger, J. O. (1985). *Statistical decision theory and bayesian analysis* (Second Edition. ed.). New York, NY: Springer New York.

Bernheim, B. D. (2016). The good, the bad, and the ugly: A unified approach to behavioral welfare economics1. *Journal of Benefit-Cost Analysis*, *7*(1), 12–68.

Betzler, N., Slinko, A., & Uhlmann, J. (2013). On the computation of fully proportional representation. *Journal of Artificial Intelligence Research*, *47*, 475–519. doi: 10.1613/jair.3907

Black, D. (1958). *The theory of committees and elections.* London: Cambridge University Press.

Boland, P. J. (1989). Majority systems and the condorcet jury theorem. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *38*(3), 181–189.

Boutilier, C., Caragiannis, I., Haber, S., Lu, T., Procaccia, A. D., & Sheffet, O. (2012). Optimal social choice theory functions: A utilitarian view. In *Proceedings of the 13th acm conference on electronic commerce* (pp. 197–214). doi: 10.1145/2229012.2229033

Bradley, R. (2009). Revising incomplete attitudes. *Synthese*, *171*(2), 235–

256.

Bradley, R. (2017). *Decision theory with a human face*. Cambridge University Press.

Bradley, R., Helgeson, C., & Hill, B. (2017). Climate change assessments: confidence, probability, and decision. *Philosophy of Science*, *84*(3), 500–522.

Bradley, R., & Thompson, C. (2012). A (mainly epistemic) case for multiple-vote majority rule. *Episteme*, *9*(1), 63–79.

Bradley, S., Frigg, R., Du, H., & Smith, L. A. (2014). Model error and ensemble forecasting: A cautionary tale. *Scientific explanation and methodology of science*, *1*, 58–66.

Brams, S. J., & Fishburn, P. C. (1978). *Approval voting* (Vol. 72) (No. 3). American Political Science Review. doi: 10.1017/S0003055400269421

Briggs, R. A. (2023). Normative theories of rational choice: Expected utility. In E. N. Zalta & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy (winter 2023 edition).* Metaphysics Research Lab, Stanford University. Retrieved from `https://plato.stanford.edu/archives/win2023/entries/rationality-normative-utility/`

Bright, L. K. (2017). On fraud. *Philosophical Studies*, *174*(2), 291–310.

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. In *Psychology of learning and motivation* (Vol. 32, pp. 275–318). Elsevier.

Camerer, C. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of risk and uncertainty*, *5*(4), 325–370.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford University Press.

Cartwright, N. (2009). If no capacities then no credible worlds. but can models reveal capacities? *Erkenntnis*, *70*(1), 45–58.

Chakrabarty, D., & Swamy, C. (2014). Welfare maximization and truthfulness in mechanism design with ordinal preferences. In *Proceedings of the 5th conference on innovations in theoretical computer science* (pp. 105–120). doi: 10.1145/2554797.2554809

Chamberlin, J. R. (1985). An investigation into the relative manipulability of four voting systems. *Behavioral Science*, *30*(4), 195–203.

Christiano, T. (2008). *The constitution of equality: Democratic authority and its limits*. Oxford: Oxford University Press.

Christiano, T., & Bajaj, S. (2022). *Democracy.* The Stanford Encyclopedia of Philosophy (Spring 2022 Edition), Edward N. Zalta (ed.). Retrieved from `https://plato.stanford.edu/archives/spr2022/entries/democracy/`

Clark, N. (2019, Dec 9). *Red wall down 100 labour seats where leavers can tactically back the conservatives to get brexit done.* THE Sun. Retrieved from `https://www.thesun.co.uk/news/10512094/100-labour-seats-where-leavers-can-back-the-conservatives-to-get-brexit-done/`

Colyvan, M. (2013). Idealisations in normative models. *Synthese*, *190*(8), 1337–1350.

Conitzer, V., & Walsh, T. (2016). Barriers to manipulation in voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, & A. D. Procaccia (Eds.), *Handbook of computational social choice* (pp. 127–147). Cambridge University Press.

Conley, J. P., Toossi, A., & Wooders, M. H. (2006). Memetics and voting: How nature may make us public spirited. *International Journal of Game Theory*, *35*(1), 71–90.

Cordelli, C. (2020). *The privatized state*. Princeton University Press.

Dahl, R. A. (1957). The concept of power. *Behavioral Science*, *2*(3), 201–215. doi: 10.1002/bs.3830020303

Dancey, L., & Sheagley, G. (2013). Heuristics behaving badly: Party cues and voter knowledge. *American Journal of Political Science*, *57*(2), 312–325.

Dancy, J. (1995). In defence of thick concepts. *Midwest Studies in Philosophy*, *20*(1), 263–279.

Darcy, R., & Marsh, M. (1994). Decision heuristics: Ticket-splitting and the irish voter. *Electoral Studies*, *13*(1), 38–49.

Dasgupta, P., & Maskin, E. (2020). Strategy-proofness, independence of irrelevant alternatives, and majority rule. *American Economic Review: Insights*, *2*(4), 459–474.

Davies, D. (2019, Dec 11). *Voting tactically? here's how to do it right.* The Guardian. Retrieved from https://www.theguardian.com/commentisfree/2019/dec/11/voting-tactically-close-races-party-vote-tories-guidelines

Dethier, C. (2022). Interpreting the probabilistic language in ipcc reports.

Dietrich, F. (2008). The premises of condorcet's jury theorem are not simultaneously justified. *Episteme*, *5*(1), 56–73.

Dietrich, F., & Spiekermann, K. (2013). Epistemic democracy with defensible premises. *Economics and Philosophy*, *29*(1), 87–120.

Dowding, K. (1996). *Power.* Open University Press.

Dowding, K., & Van Hees, M. (2008). In praise of manipulation. *British Journal of Political Science*, *38*(1), 1–15.

Drewett, Z., & Richards, E. (2021, May 6). *What is tactical voting and why do people do it.* Metro. Retrieved from https://metro.co.uk/2021/05/06/what-is-tactical-voting-and-why-do-people-do-it-14529967/

Dubins, L. E., & Freedman, D. A. (1981). Machiavelli and the gale-shapley algorithm. *The American Mathematical Monthly*, *88*(7), 485–494.

Eggers, A. C., & Vivyan, N. (2020). Who votes more strategically? *American Political Science Review*, *114*(2), 470–485. doi: 10.1017/S0003055420000087

Elgot, J. (2022, Dec 8). *Campaign will target 62 seats to oust tories by tactical voting.* The Guardian. Retrieved from https://www.theguardian.com/politics/2022/dec/08/win-as-one-campaign-target-tories-seats-tactical-voting

Elkind, E., & Lipmaa, H. (2005). Hybrid voting protocols and hardness of manipulation. In *Algorithms and computation: 16th international symposium, isaac 2005, sanya, hainan, china, december 19-21, 2005. proceedings* (pp. 206–215). Springer Berlin Heidelberg.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, *75*(4), 643–669.

Elstein, D. Y., & Hurka, T. (2009). From thick to thin: Two moral reduction plans. *Canadian Journal of Philosophy*, *39*(4), 515–535.

Farquharson, R. (1969). *Theory of voting.* Oxford: Blackwell.

Favardin, P., & Lepelley, D. (2006). Some further results on the manipulability of social choice rules. *Social Choice and Welfare*, *26*(3), 485–509.

Favardin, P., Lepelley, D., & Serais, J.-P. (2002). Borda rule, copeland method and strategic manipulation. *Review of Economic Design*, *7*(3), 213–228.

Feddersen, T., & Pesendorfer, W. (1997). Voting behavior and information aggregation in elections with private information. *Econometrica: Journal of the Econometric Society*, 1029–1058.

Feigenbaum, J., & Shenker, S. (2004). Distributed algorithmic mechanism design: Recent results and future directions. In *Current trends in theoretical computer science: The challenge of the new century vol 1: Algorithms and complexity vol 2: Formal models and semantics* (pp. 403–434). World Scientific.

Felsenthal, D. S., & Machover, M. (1998). *The measurement of voting power: Theory and practice, problems and paradoxes.* Cheltenham: Edward Elgar.

Filos-Ratsikas, A., & Miltersen, P. B. (2014). Truthful approximations to range voting. In *Web and internet economics: 10th international conference, wine 2014, beijing, china, december 14-17, 2014. proceedings 10* (pp. 175–188).

Fiorina, M. P. (1981). *Retrospective voting in american national elections.* Yale University Press.

Fisher, S. D. (2004). Definition and measurement of tactical voting: The role of rational choice. *British Journal of Political Science*, *34*(1), 152–166. doi: 10.1017/S0007123403000372

Flanigan, B., Liang, J., Procaccia, A. D., & Wang, S. (2024, March). Manipulation-robust selection of citizens' assemblies. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 38, pp. 9696–9703).

Foot, P. (1958). Moral arguments. *Mind*, *67*(268), 502–513.

Friedman, D. (1991). Evolutionary games in economics. *Econometrica: Journal of the Econometric Society*, 637–666.

Friedman, M. (1953). The methodology of positive economics. In *Essays in positive economics* (pp. 3–43). University of Chicago Press.

Frisch, D., & Baron, J. (1988). Ambiguity and rationality. *Journal of Behavioral Decision Making*, *1*(3), 149–157.

Garrett, G., & Tsebelis, G. (1999). Why resist the temptation to apply power indices to the european union? *Journal of Theoretical Politics*, *11*(3), 291–308. doi: 10.1177/0951692899011003004

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, *41*(4), 587–601. doi: 10.2307/1914083

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.

Gilboa, I., Postlewaite, A., & Schmeidler, D. (2009). Is it always rational to satisfy savage's axioms? *Economics & Philosophy*, *25*(3), 285–296.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of mathematical economics*, *18*(2), 141–153.

González-Ricoy, I., & Queralt, J. (2018). Political liberties and social equality. *Law and Philosophy*, *37*, 613–638. doi: 10.1007/s10982-018-9338-2

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B, Methodological*, *14*(1), 107–114.

Goodin, R. E., & Spiekermann, K. (2018). *An epistemic theory of democracy.* Oxford University Press.

Gould, C. C. (1990). *Rethinking democracy: Freedom and social co-operation in politics, economy, and society.* Cambridge University Press.

Grofman, B., Owen, G., & Feld, S. L. (1983). Thirteen theorems in search of the truth. *Theory and decision*, *15*(3), 261–278.

Harrison, G. W., & McDaniel, T. (2008). Voting games and computational complexity. *Oxford Economic Papers*, *60*(3), 546–565.

Hausman, D. M. (1994). Why look under the hood? In *The philosophy of economics: An anthology* (3rd ed., pp. 183–187). Cambridge University Press.

Hausman, D. M. (2000). Revealed preference, belief, and game theory. *Economics & Philosophy*, *16*(1), 99–115.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of risk and uncertainty*, *4*(1), 5–28.

Heesen, R., Bright, L. K., & Zucker, A. (2019). Vindicating methodological triangulation. *Synthese*, *196*(8), 3067–3081.

Herbert, T. (2019, Dec 12). *What is tactical voting in an election and does it really work in the uk?* The Evening Standard. Retrieved from `https://www.standard.co.uk/news/politics/what-is-tactical-voting-uk-does-it-work-a4310896.html`

Hofbauer, J., & Sigmund, K. (2003). Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, *40*(4), 479–519.

Holler, M. J., & Nurmi, H. (2014). Aspects of power overlooked by power indices. In *Voting power and procedures: Essays in honour of dan felsenthal and moshé machover* (pp. 205–219). Cham: Springer International Publishing.

Jeffrey, R. C. (1983). Bayesianism with a human face.

Joyce, J. M. (2005). How probabilities reflect evidence. *Philosophical perspectives*, *19*, 153–178.

Joyce, J. M. (2010). A defense of imprecise credences in inference and decision making. *Philosophical perspectives*, *24*, 281–323.

Kelly, J. S. (1993). Almost all social choice rules are highly manipulable, but a few aren't. *Social Choice and Welfare*, *10*(2), 161–175.

Kim, K. H., & Roush, F. W. (1996). Statistical manipulability of social choice functions. *Group Decision and Negotiation*, *5*(3), 263–282.

Kolodny, N. (2014a). Rule over none ii: Relational equality and the justification of democracy. *Philosophy & Public Affairs*, *42*(4), 287–336. doi: 10.1111/papa.12036

Kolodny, N. (2014b). Rule over none i: What justifies democracy? *Philosophy & Public Affairs*, *42*(3), 195–229. doi: 10.1111/papa.12034

Kube, S., & Puppe, C. (2009). (when and how) do voters try to manipulate? experimental evidence from borda elections. *Public Choice*, *139*(1), 39–52.

Kyburg Jr, H. E. (1983). *Rational belief.* (Tech. Rep.). ROCHESTER UNIV NY DEPT OF PHILOSOPHY.

Ladha, K. K. (1992). The condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science*, 617–634.

Ladha, K. K. (1993). Condorcet's jury theorem in light of de finetti's theorem. *Social Choice and Welfare*, *10*(1), 69–85.

Lau, R. R., & Redlawsk, D. P. (2006). *How voters decide: Information processing in election campaigns.* Cambridge University Press.

Lehtinen, A. (2008). The welfare consequences of strategic behaviour under approval and plurality voting. *European Journal of Political Economy*, *24*(3), 688–704.

Lehtinen, A. (2015). A welfarist critique of social choice theory: Interpersonal comparisons in the theory of voting. *Erasmus Journal for Philosophy and Economics*, *8*(2), 34–83.

Lepelley, D., & Mbih, B. (1994). The vulnerability of four social choice functions to coalitional manipulation of preferences. *Social Choice and Welfare*, *11*(3), 253–275.

Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy*, *71*(13), 391–418.

Levi, I. (1985). Imprecision and indeterminacy in probability judgment. *Philosophy of Science*, *52*(3), 390–409.

List, C. (2022). Social choice theory. *The Stanford Encyclopedia of Philosophy*. Retrieved from `https://plato.stanford.edu/archives/win2022/entries/social-choice/` (Accessed: 2024-09-15)

List, C., & Goodin, R. E. (2001). Epistemic democracy: Generalizing the condorcet jury theorem.

List, C., & Spiekermann, K. (2016). The condorcet jury theorem and voter-specific truth.

Lukes, S. (2005). *Power: A radical view* (2nd ed.). Palgrave Macmillan.

Lupia, A. (1994). Shortcuts versus encyclopedias: Information and voting behavior in california insurance reform elections. *American Political Science Review*, *88*(1), 63–76.

Mahtani, A. (2019). Imprecise probabilities. In J. Pettigrew R. Weisberg (Ed.), *Open handbook of formal epistemology.*

Mäki, U., et al. (1992). On the method of isolation in economics. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, *26*(4), 317–351.

Mattei, N., Forshee, J., & Goldsmith, J. (2012). An empirical study of voting rules and manipulation with large datasets. In *Proceedings of comsoc* (p. 59).

McDermott, M. L. (2005). Candidate occupations and voter information shortcuts. *The Journal of Politics*, *67*(1), 201–219.

Meffert, M. F., & Gschwend, T. (2007). Strategic voting under proportional representation and coalition governments: A simulation and laboratory experiment. In *Rationalitätskonzepte, entscheidungsverhalten und ökonomische modellierung* (Vol. 7).

Mitchell, P., & Alexandrova, A. (2021). Well-being and pluralism. *Journal of Happiness Studies*, *22*(6), 2411–2433.

Monkovic, T. (2016, March 30). *Is it o.k. if your vote is a lie? ask a canadian.* The New York Times. Retrieved from `https://www.nytimes.com/2016/03/31/upshot/is-it-ok-if-your-vote-is-a-lie-ask-a-canadian.html`

Morgan, M. S. (2012). *The world in the model: How economists work and think*. Cambridge University Press.

Morreau, M. (2019). Arrow's theorem. *The Stanford Encyclopedia of Philosophy*. Retrieved from `https://plato.stanford.edu/archives/win2019/entries/arrows-theorem/` (Accessed: 2024-09-15)

Morriss, P. (2002). *Power: A philosophical analysis.* Manchester University Press.

Moulin, H. (1980). On strategy-proofness and single peakedness. *Public Choice*, *35*(4), 437–455.

Mueller, D. C. (2003). *Public choice iii.* Cambridge: Cambridge University Press.

Mueller, D. C. (2015). Public choice, social choice, and political economy. *Public Choice*, *163*(3), 379–387.

Myerson, R. B. (1979). Incentive compatibility and the bargaining problem. *Econometrica: Journal of the Econometric Society*, 61–73.

Myerson, R. B., & Weber, R. J. (1993). A theory of voting equilibria. *American Political Science Review*, *87*, 102–114. doi: 10.2307/2938962

Napel, S. (2018). Voting power. In *The oxford handbook of public choice* (Vol. 1, pp. 103–122). Oxford University Press.

Napel, S., & Widgrén, M. (2005). The possibility of a preference-based power

index. *Journal of Theoretical Politics*, *17*(3), 377–387.
Napel, S., & Widgrén, M. (2009). Strategic vs. non-strategic power in the eu council: The consultation procedure.
Nehring, K., & Puppe, C. (2007a). Efficient and strategy-proof voting rules: A characterization. *Games and Economic Behavior*, *59*(1), 132–183.
Nehring, K., & Puppe, C. (2007b). The structure of strategy-proof social choice part i: General characterization and possibility results on median spaces. *Journal of Economic Theory*, *135*(1), 269–305.
Nitzan, S. (1985). The vulnerability of point-voting schemes to preference variation and strategic manipulation. *Public Choice*, *47*(2), 349–379.
Nitzan, S., & Paroush, J. (1982). Optimal decision rules in uncertain dichotomous choice situations. *International Economic Review*, 289–297.
Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, *364*(6432), 56–58.
Otto, A. R., Taylor, E. G., & Markman, A. B. (2011). There are at least two kinds of probability matching: Evidence from a secondary task. *Cognition*, *118*(2), 274–279.
Pattanaik, P. K. (1978). *Strategy and group choice*. North-Holland.
Peleg, B. (1979). A note on manipulability of large voting schemes. *Theory and Decision*, *11*(4), 401–418.
Penrod, S., & Hastie, R. (1979). Models of jury decision making: A critical review. *Psychological Bulletin*, *86*(3), 462.
Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society*, *109*(1), 53–57. doi: 10.2307/2981392
Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Oxford: Oxford University Press.
Pincock, C. (2007). Mathematical idealization. *Philosophy of Science*, *74*(5), 957–967.
Plato. (2003). *The republic* (Revised Edition ed.; D. Lee, Ed.). Har-

mondsworth: Penguin Books.

Popkin, S. L. (1995). Information shortcuts and the reasoning voter. In *Information, participation and choice: An economic theory of democracy in perspective* (pp. 17–35).

Potthoff, R. F., & Brams, S. J. (1998). Proportional representation: Broadening the options. *Journal of Theoretical Politics*, *10*(2), 147–178. doi: 10.1177/0951692898010002002

Pritchard, G., & Wilson, M. C. (2007). Exact results on manipulability of positional voting rules. *Social Choice and Welfare*, *29*(3), 487–519.

Putnam, H. (2002). *The collapse of the fact-value dichotomy and other essays*. Harvard University Press.

Risse, M. (2009). On the philosophy of group decision methods i: The nonobviousness of majority rule. *Philosophy Compass*, *4*(5), 793–802.

Roberts, D. J., & Postlewaite, A. (1976). The incentives for price-taking behavior in large exchange economies. *Econometrica: journal of the Econometric Society*, 115–127.

Romeijn, J.-W., & Atkinson, D. (2011). Learning juror competence: A generalized condorcet jury theorem. *Politics, Philosophy & Economics*, *10*(3), 237–262.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research*, *7*(4), 617–642.

Roussos, J. (2022). Modelling in normative ethics. *Ethical Theory and Moral Practice*, *25*(5), 865–889.

Saari, D. G. (1990). Susceptibility to manipulation. *Public Choice*, *64*(1), 21–41. doi: 10.1007/BF00184002

Saari, D. G. (1992). Millions of election outcomes from a single profile. *Social Choice and Welfare*, *9*(4), 277–306.

Satterthwaite, M. A. (1973). Existence of a strategy-proof voting procedure: A topic in social choice theory. *The University of Wisconsin-Madison*.

Satterthwaite, M. A. (1975). 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. doi: 10.1016/0022-0531(75)90050-2

Scheffler, S. (2015). The practice of equality. In C. Fourie, F. Schuppert, & I. Wallimann-Helmer (Eds.), *Social equality: On what it means to be equal* (pp. 21–44). Oxford: Oxford University Press.

Schmidtchen, D., & Steunenberg, B. (2014). On the possibility of a preference-based power index: The strategic power index revisited. In *Voting power and procedures: Essays in honour of dan felsenthal and moshé machover* (pp. 259–286). Cham: Springer International Publishing.

Scott, J. (2018). *Power.* John Wiley & Sons.

Seidenfeld, T. (2004). A contrast between two decision rules for use with (convex) sets of probabilities: γ-maximin versus e-admissibility. *Synthese*, *140*(1/2), 69–88.

Sen, A. (1977). Social choice theory: A re-examination. *Econometrica: Journal of the Econometric Society*, 53–89.

Sengupta, M. (1980). Monotonicity, independence of irrelevant alternatives and strategy-proofness of social decision functions. *The Review of Economic Studies*, *47*(2), 393–407.

Shrestha, Y. R., & Yang, Y. (2019). Fairness in algorithmic decision-making: Applications in multi-winner voting, machine learning, and recommender systems. *Algorithms*, *12*(9), 199. doi: 10.3390/a12090199

Sieg, G., & Schulz, C. (1995). Evolutionary dynamics in the voting game. *Public Choice*, *85*(1-2), 157–172.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Singer, P. (1973). Altruism and commerce: A defense of titmuss against arrow. *Philosophy & Public Affairs*, 312–350.

Slinko, A. (2004). How large should a coalition be to manipulate an election? *Mathematical Social Sciences*, *47*(3), 289–293. doi: 10.1016/j.mathsocsci.2004.06.002

Smith, D. A. (1999). Manipulability measures of common social choice functions. *Social Choice and Welfare*, *16*(4), 639–661.

Sniderman, P. M., Brody, R. A., & Tetlock, P. E. (1993). *Reasoning and choice: Explorations in political psychology.* Cambridge University Press.

Steele, K. (2012). The scientist qua policy advisor makes value judgments. *Philosophy of Science*, *79*(5), 893–904.

Steunenberg, B., Schmidtchen, D., & Koboldt, C. (1999). Strategic power in the european union: evaluating the distribution of power in policy games. *Journal of Theoretical Politics*, *11*(3), 339–366.

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, *87*(3), 449–508.

Tal, M., Meir, R., & Gal, Y. A. (2015, May). A study of human behavior in online voting. In *Proceedings of the 2015 international conference on autonomous agents and multiagent systems* (pp. 665–673).

Thoma, J. (2024). Social science, policy and democracy. *Philosophy & Public Affairs*, *52*(1), 5–41.

Thomson, W. (2001). On the axiomatic method and its recent applications to game theory and resource allocation. *Social Choice and Welfare*, *18*(2), 327–386.

Times, T. F. (2015, April 30). *General election: The compelling case for continuity in britain.* The Financial Times. Retrieved from `https://www.ft.com/content/e61ce174-ea94-11e4-96ec-00144feab7de`

Troffaes, M. C. (2007). Decision making under uncertainty using imprecise probabilities. *International journal of approximate reasoning*, *45*(1), 17–29.

Tsetlin, I., Regenwetter, M., & Grofman, B. (2003). The impartial culture maximizes the probability of majority cycles. *Social Choice and*

Welfare, *21*(3), 387–418.

Urken, A. B., & Traflet, S. (1983). Optimal jury design. *Jurimetrics J.*, *24*, 218.

Väyrynen, P. (2021). Thick ethical concepts. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy (spring 2021 edition).* (Retrieved from https://plato.stanford.edu/archives/spr2021/entries/thick-ethical-concepts/)

Viehoff, D. (2014). Democratic equality and political authority. *Philosophy & Public Affairs*, *42*(4), 337–375. doi: 10.1111/papa.12035

Viehoff, D. (2019). Power and equality. *Oxford Studies in Political Philosophy*, *5*, 1–38.

Waldron, J. (1999). *The dignity of legislation* (Vol. 2). Cambridge University Press.

Wang, Z., Xu, B., & Zhou, H.-J. (2014). Social cycling and conditional responses in the rock-paper-scissors game. *Scientific Reports*, *4*, 5830.

Wheeler, G. (2020). Bounded rationality. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy (spring 2020 edition).* Metaphysics Research Lab, Stanford University. Retrieved from `https://plato.stanford.edu/archives/spr2020/entries/bounded-rationality/`

Williams, B. (1985). *Ethics and the limits of philosophy.* Harvard University Press.

Wilson, J. L. (2019). *Democratic equality.* Princeton University Press.

Worthy, D. A., Hawthorne, M. J., & Otto, A. R. (2013). Heterogeneity of strategy use in the iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychonomic Bulletin & Review*, *20*(2), 364–371.

Zuckerman, M., Procaccia, A. D., & Rosenschein, J. S. (2009). Algorithms for the coalitional manipulation problem. *Artificial Intelligence*, *173*(2), 392–412. doi: 10.1016/j.artint.2008.11.003

Appendix

A.1 MFC, Honesty, and EU-maximisation Guarantee Voter Competence

Proof. Assume, without loss of generality, that 1 is the true state of the world. We know that if $p_i(T) = 0.5$ then $EU_i(V_i = T) = 0$ and $EU_i(V_i = F) = F$. Hence if $p_i(T) > 0.5$ then $EU_i(V_i = T) > EU_i(V_i = F)$. We know from MFC that if $X = T$, $Pr(p_i(T) > 0.5|X = T) > 0.5$. Thus, this simple model guarantees *voter competence*. □

A.2 Under SMFSC, Γ-Maximin Guarantees Voter Competence

Proof. Without loss of generality, assume that $X = T$. If any voter i uses Γ-Maximin we know that for $D = \{T, F\}$, for any voter i, whenever $\underline{EU_i}_{\mathcal{P}_i}(T) > \underline{EU_i}_{\mathcal{P}_i}(F)$, then $V_i = T$. Since $\mathcal{P}(T)$ is a convex set of probabilities, we only need to use the extreme points $\underline{P_i}(T)$ and $\overline{P_i}(T)$ to find $\underline{EU_i}_{\mathcal{P}_i}(T)$ and $\underline{EU_i}_{\mathcal{P}_i}(F)$. In combination with the utility assumption that for any voter i, $U_i(T, T) = U_i(F, F) = 1$, and $U_i(T, F) = U_i(F, T) = -1$, we have:

$$\underline{EU_i}_{\mathcal{P}_i}(T) = \min\{\underline{P_i}(T) - (1 - \underline{P_i}(T)), \overline{P_i}(T) - (1 - \overline{P_i}(T))\}$$

$$\underline{EU}_{i_{\mathcal{P}_i}}(F) = \min\{-\underline{P_i}(T) + (1 - \underline{P_i}(T)), -\overline{P_i}(T) + (1 - \overline{P_i}(T))\}$$

Because $\underline{P_i}(T) \leq \overline{P_i}(T)$, we can write:

$$\underline{EU}_{i_{\mathcal{P}_i}}(T) = \underline{P_i}(T) - (1 - \underline{P_i}(T))$$

$$\underline{EU}_{i_{\mathcal{P}_i}}(F) = -\overline{P_i}(T) + (1 - \overline{P_i}(T))$$

Thus, for any voter i, if $\underline{P_i}(T) > 1 - \overline{P_i}(T)$, it follows $\underline{EU}_{i_{\mathcal{P}_i}}(T) > \underline{EU}_{i_{\mathcal{P}_i}}(F)$ and it follows that $V_i = T$. By SMFSC, for any voter i, $Pr(\underline{P_i}(T) > 1 - \overline{P_i}(T)) > 0.5$, thus for any voter i, $Pr(\underline{EU}_{i_{\mathcal{P}_i}}(T) > \underline{EU}_{i_{\mathcal{P}_i}}(F)) > 0.5$ and it follows that $Pr(V_i = T) > 0.5$. Therefore, voters are guaranteed to be competent. □

A.3 Under SMFSC, E-admissibility Does Not Guarantee Voter Competence

Proof. Without loss of generality, assume that $X = T$. If every voter i uses E-admissibility, we know that for $D = \{T, F\}$, for any voter i whenever for every $p_i(T) \in \mathcal{P}_i(T)$, $EU_{p_i(T)}(T) > EU_{p_i(T)}(F)$, then $V_i = T$.

Moreover, for $p_i(T) = 0.5$, $EU_{p_i(T)}(T) = EU_{p_i(T)}(F) = 0$, and thus $V_i = F$ or $V_i = T$. Also, for $p_i(T) < 0.5$, $EU_{p_i(T)}(T) < EU_{p_i(T)}(F)$, and thus $V_i = F$. Furthermore, for $p_i(T) > 0.5$, $EU_{p_i(T)}(T) > EU_{p_i(T)}(F)$, and thus $V_i = T$. Since for every $p_i(T) \in \mathcal{P}_i(T)$, it is the case that $p_i(T) \geq \underline{P_i}(T)$, and it follows that if $EU_{\underline{P_i}(T)} > 0$ for every voter i, then $V_i = T$. However, if $EU_{\underline{P_i}(T)} \leq 0$, it may be that $V_i = F$. Thus, if and only if $\underline{P_i}(T) > 0.5$, it is guaranteed that $V_i = T$. It follows that to guarantee $Pr(V_i = T) > 0.5$, $Pr(\underline{P_i}(T) > 0.5) > 0.5$. Yet, SMFSC only guarantees $Pr(\underline{P_i}(T) > 1 - \overline{P_i}(T)) > 0.5$, which is compatible with $Pr(\underline{P_i}(T) > 0.5) = 0$. Thus, voter competence under SMFSC is not guaranteed. □

A.4 Under MFSC, if Voters Can Abstain, Γ-Maximin Can Retrieve Epistemic Benefits of Voting

Proof. Assume without loss of generality that $X = T$. Then, if every voter i uses Γ-Maximin for $D = \{T, F, A\}$, then $V_i = T$, whenever $\underline{EU_{i}}_{\mathcal{P}_i}(T) > \underline{EU_{i}}_{\mathcal{P}_i}(F)$ and $\underline{EU_{i}}_{\mathcal{P}_i}(T) > \underline{EU_{i}}_{\mathcal{P}_i}(A)$. Similarly, if $\underline{EU_{i}}_{\mathcal{P}_{i(T)}}(F) > \underline{EU_{i}}_{\mathcal{P}_i}(T)$ and $\underline{EU_{i}}_{\mathcal{P}_i}(F) > \underline{EU_{i}}_{\mathcal{P}_i}(A)$, then for any voter i, $V_i = F$.

We already know that

$$\underline{EU_{i}}_{\mathcal{P}_i}(T) = \underline{P_i}(T) - (1 - \underline{P_i}(T))$$

$$\underline{EU_{i}}_{\mathcal{P}_i}(F) = -\overline{P_i}(T) + (1 - \overline{P_i}(T))$$

$$\underline{EU_{i}}_{\mathcal{P}_i}(A) = 0$$

Now consider that $\underline{P_i}(T) - (1 - \underline{P_i}(T)) > 0$ only if $\underline{P_i}(T) > 0.5$. Similarly, $-\overline{P_i}(T) + (1 - \overline{P_i}(T)) > 0$ only if $\overline{P_i}(T) < 0.5$. Thus the following holds:

$$opt_{\underline{EU_{i}}_{\mathcal{P}_i}}(D) = \begin{cases} \{T\} & \text{if } \underline{P_i}(T) > 0.5 \\ \{F\} & \text{if } \overline{P_i}(T) < 0.5 \\ \{F, T, A\} & \text{if } \underline{P_i}(T) = \overline{P_i}(T) = 0.5 \\ \{A\} & \text{otherwise} \end{cases}$$

MFSC guarantees that $Pr(\underline{P_i}(T) > 0.5) > Pr(\overline{P_i}(T) < 0.5)$, and it follows that $Pr(V_i = T) > Pr(V_i = F)$, and thus generalized voter competence is guaranteed (i.e., $c > 0.5$). Abstaining does not increase the votes for verdict T and also not for verdict F. Thus, as long as $Pr(V_i = T) > Pr(V_i = F)$, it will be the case that by the law of large numbers, as the number of voters approaches infinity, almost surely $\sum V_i = T > \sum V_i = F$, and thus the election outcome is almost surely T. $\square$