UNIVERSITY OF TURKU
Turku School of Economics

The Limits of Vulnerability to Manipulation as an Argument for Electoral Reform

Master's thesis

Author:
Luukas Laiho

Supervisor:
Professor Ville Korpela

19.4.2026
Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Master's thesis

Subject: Economics
Author: Luukas Laiho
Title: The Limits of Vulnerability to Manipulation as an Argument for Electoral Reform
Supervisor(s): Professor Ville Korpela
Number of pages: 70 pages
Date: 19.4.2026

The field of social choice theory has been shaped by the formative theorems of Arrow and Gibbard & Satterthwaite. Motivated by the Gibbard-Satterthwaite theorem effectively showing that all sensible voting rules are manipulable i.e., susceptible to strategic voting, metrics comparing the degree of manipulability of different voting rules have been developed. However, the concept of manipulation is often deemed undesirable without explicit justification. The objective of this thesis is to evaluate the ethical status and democratic implications of strategic voting, and what the significance of these factors are for applying the manipulability metrics. In particular, this thesis aims to assess the grounds on which concepts related to strategic voting can justifiably be used to argue for adopting one voting rule over another. A literature review was conducted on the metrics developed for measuring manipulability of voting rules and the welfare impacts of manipulation, as well as the ethical and democratic arguments for and against strategic voting. The findings of the literature review were applied to the discussions on electoral reform in the United States, focusing on the ongoing debate between Ranked-Choice Voting (RCV) and plurality voting. The findings of the literature review show that measuring manipulability is dependent on a variety of assumptions used in the different metrics, thus lacking unified consensus on the manipulability of different voting rules. The ethical status and democratic implications of strategic voting are found to be contestable as well. This implies that for a metric of manipulability to carry prescriptive weight in discussions of real-world election reform, the assumed normative wrong of strategic voting should be explicitly substantiated, and moreover, it should be justified how the used metrics tracks this particular concern.

Key words: strategic voting, manipulability of voting rules, Ranked-Choice Voting, social choice theory

Pro gradu -tutkielma

Oppiaine: Taloustiede
Tekijä: Luukas Laiho
Otsikko: The Limits of Vulnerability to Manipulation as an Argument for Electoral Reform
Ohjaaja: professori Ville Korpela
Sivumäärä: 70 sivua
Päivämäärä: 19.4.2026

Sosiaalisen valinnan teoria on muovautunut Arrowin sekä Gibbardin ja Satterthwaiten formatiivisten teoreemien myötä. Gibbard-Satterthwaiten teoreeman käytännössä osoittaa kaikkien mielekkäiden äänestyssääntöjen olevan manipuloitavissa, toisin sanoen alttiita strategiselle äänestämiselle, mikä on motivoinut eri menetelmiä mitata äänestyssääntöjen alttiuitta manipuloitavuudelle. Suurempaa alttiuitta strategiselle äänestämiselle pidetään tyypillisesti negatiivisena ominaisuutena, usein tätä näkemystä kuitenkaan tarkemmin perustelematta. Tämän tutkielman tavoitteena on arvioida strategisen äänestämisen eettistä asemaa sekä demokraattista merkitystä, sekä suhdetta näiden normatiivisten komponenttien ja äänestyssääntöjen manipuloitavuuden mittareiden hyödyntämisen välillä. Tutkielma arvioi miten strategiseen äänestämiseen liittyviä käsitteitä voi sosiaalisen valinnan teorian kirjallisuuden perusteella mielekkäästi käyttää äänestyssääntöjen puolesta tai vastaan argumentoidessa. Tutkielman kirjallisuuskatsaus kartoitti äänestyssääntöjen manipuloitavuuden mittareita sekä manipuloinnin hyvinvointivaikutusten mittaamiseen liittyvää kirjallisuutta, sekä eettisiä ja demokraattisia strategiseen äänestämiseen liittyviä argumentteja. Kirjallisuuskatsauksen havaintoja sovellettiin Amerikan Yhdysvalloissa käytäviin vaaliuudistuskeskusteluun, keskittyen siirtoäänivaalitavan implementointiin osavaltioissa. Kirjallisuuskatsauksen johtopäätökset ovat pitkälti hajanaisia. Äänestyssääntöjen alttiuiden manipulaatiolle sekä strategisen äänestämisen hyvinvointivaikutusten mittaaminen on perusteltua useilla eri mittareilla, malleilla ja taustaoletuksilla. Strategisen äänestämisen haitoista ei niinikään ole yksimielisyyttä, ja varteenotettavia vasta-argumentteja on olemassa. Täten, jotta manipuloitavuuden mittarilla olisi painoarvoa vaaliuudistuskeskusteluissa, strategisen äänestämisen oletetut haitat tulisi ilmaista tarkemmin, ja lisäksi osoitettava että kyseinen mittari mittaa kyseistä oletettua haittaa.

Avainsanat: strateginen äänestäminen. äänestyssääntöjen manipuloitavuus, siirtoäänivaalitapa, sosiaalisen valinnan teoria

TABLE OF CONTENTS

1 Introduction **8**

2 Voting Rules and Manipulation **10**

2.1 Social Choice Theory **10**

2.2 Preference Aggregation, Social Choice and Voting rules **11**

2.3 Arrow's Theorem & Properties of Voting Rules **14**

2.4 Gibbard-Satterthwaite Theorem & Manipulation **17**

3 Comparative Analysis of Voting Rule Manipulability **22**

3.1 Voting Rules **22**

3.1.1 Examples of Manipulation 24

3.2 Measuring Manipulability **26**

3.2.1 Compilation of Results 34

3.3 Welfare Impact of Manipulation **36**

4 Manipulation, Ethics and Democratic Ideals **43**

4.1 Ethics of Manipulation **43**

4.2 Democracy and Manipulation **46**

5 Role of Manipulability in the U.S. RCV Debate **50**

5.1 Ranked Choice Voting in the United States **50**

5.2 2022 Alaska Special Election for US House **52**

5.3 Lower manipulability as an argument for election reform **56**

6 Conclusions **60**

References **62**

List of tables

Table 3.1 Example voter preferences with 4 alternatives 24
Table 3.2 Example voter preferences with 3 alternatives (a) 24
Table 3.3 Example voter preferences with 3 alternatives (b) 25
Table 3.4 Example voter preferences with 3 alternatives (c) 25
Table 3.5 Number of distinct voting situations under IC and IAC 28
Table 5.1 Submitted ballots in the 2022 Alaska Special Election 52
Table 5.2 Alternative ballots in the 2022 Alaska Special Election 53

1 Introduction

Collective decision making and social choice theory are vital areas of microeconomic research, and the challenges involved in aggregating individual preferences to form a collective decision have been known ever since Kenneth Arrow's seminal impossibility theorem. But the challenges that collective preference aggregation faces extend beyond the conditions that Arrow imposed, when the possibility exists for insincere behaviour to be rewarded.

Voting is a common method through which many decisions of great societal importance are constantly made around the globe. These decisions range from political elections, referendums, and initiatives to decisions made by major organizations and institutions – such as corporations and universities. Some other decisions made through voting are not as grand as these, but still relevant to the everyday life of people, such as those made by smaller community and social groups.

The Gibbard-Satterthwaite theorem effectively shows that all voting rules i.e., procedures and mechanisms through which a collective decision is made, are necessarily either dictatorial or manipulable when the number of alternatives exceeds two options. Manipulability here refers to the idea that some individual decision maker, or a coalition formed by decision makers can change the outcome yielded by the voting rule into one they prefer by misrepresenting their preferences, as opposed to voting in accordance with their sincere preferences.

One way to interpret this is that all situations of collective decision making through voting can be seen as strategic interactions. Since it is impossible to craft a voting rule where no one would ever have an incentive to vote strategically, the question of interest often shifts onto how common these strategic incentives are. By developing measures of manipulability, different voting rules can be compared and contrasted on how susceptible to strategic voting they are.

But merely focusing on the potential prevalence of strategic voting seems to lack something. An equally important question is whether strategic voting in the first place is *bad* or something that ought to be mitigated – either by virtue of the outcomes it generates, or by some other factor altogether. In order to determine the normative weight of strategic voting i.e., how should different degrees of manipulability be interpreted, the welfare effects of manipulation should be considered – as well as the relationship between strategic voting and ethical and democratic ideals.

Figuring out the normative significance of strategic voting is important if manipulability as a concept is to be applied to matters in the real world – in particular, whether or not questions related to manipulation are worth considering in political reforms related to choosing an election system.

Thus, the primary research questions are as follows:

(1) To what extent does the theoretical literature on manipulability provide sufficient grounds for its use in electoral reform debates, considering

   a. Measures of voting rule manipulability

   b. Measures of welfare impacts of manipulation

   c. Ethical and democratic implications of strategic voting

The thesis proceeds as follows: the second chapter provides a brief overview of the field of social choice theory. The general framework is presented both formally and informally, and the theories of Arrow and Gibbard & Satterthwaite are presented more thoroughly. Additionally, different facets of strategic voting behaviour are explored. The third chapter provides a literature review on voting rule manipulability. It is explored how common it is for different voting mechanisms to be susceptible to strategic behaviour, the many ways this degree of manipulability can be measured, and what the welfare outcomes of strategic behaviour are. Chapter 4 is concerned with the ethical and democratic aspects of manipulation. The term manipulation has a clear negative connotation in everyday speech, but it is worth examining more explicitly what, if anything, actually makes manipulation something that a desirable voting rule ought to avoid. As voting and collective decision making is integral to democracy, the variety of questions voting rule manipulation presents for the validity and nature of democratic processes is explored. Chapter 5 focuses on applying the findings of the literature review on real-world political reform, questioning to what extent it is sensible to use a lower degree manipulability as an argument for a voting rule over another. The sixth chapter concludes the thesis.

2 Voting Rules and Manipulation

2.1 Social Choice Theory

When economists study voting, it is typically done through the lens of social choice theory. Social choice theory refers to a branch of economics, which focuses on studying several aspects of collective decision making. Social choice theory is tightly linked to – or alternatively can be considered a branch of welfare economics, since often the question of interest in the field is concerned with how to determine whether an alternative is collectively, or socially, better than another. In this sense the study of social choice theory is not merely positive but also normative, in that it is concerned with making value judgements about the different outcomes that different preference aggregation methods yield. (Feldman & Serrano, 2006)

Historically, the origins of social choice theory can be traced back to the late 18th century when two French mathematicians, Nicolas de Condorcet and Jean-Charles de Borda developed and advocated for two distinct methods of collective preference aggregation: the Condorcet method and the Borda rule. These methods, along with other findings made by the two Frenchmen are still central to the discussions and debates in social choice theory. (List, 2022)

In the mid-20th century Kenneth Arrow can be credited with formulating the basis for modern social choice theory, by 'axiomatizing' the study of collective preference aggregation methods, or social choice functions. Arrow studied whether any possible social choice functions could satisfy certain criteria or properties, certain axioms, which would be deemed as desirable for these functions to have. This study led to the formulation of his influential impossibility theorem, that is *Arrow's impossibility theorem*. (List, 2022)

While social choice theory does not necessarily explicitly study real-world voting – but rather the general methods and challenges associated with collective preference aggregation – voting is the most significant practical application of the field. The theorems and findings of the field have implications for the functioning of real-world voting procedures, which can give insights into the pitfalls and benefits of different voting mechanisms.

What motivates the study of social choice theory? As Arrow stated in the opening paragraph of his 1951 book *Social Choice and Individual Values*: "In a capitalist democracy there are essentially two methods by which social choices can be made: voting, typically used to make 'political' decisions, and the market mechanism, typically used to make 'economic' decisions" (Arrow, 1951/2012). In

modern democratic nations, many important social decisions are not resolved by mere market forces, and therefore the importance of studying and understanding voting mechanisms is highlighted. However, it is worthwhile to note that not all decisions made through voting are political in nature, in the sense that they are not concerned with national governance. Some collective decisions can be more mundane, such as voting for the time and place of a study group session, while other decisions societally extremely impactful, such as decisions made at large corporations by shareholder voting. Still, the significant welfare effects of important political decisions should not require much explaining and do function as motivation for the study of social choice theory.

One area of application in the field of politics for social choice theory are election reforms. In the U.S., a considerable majority of Americans consider their democracy to not be working well, and are in support of election reforms (Boatright et al., 2024). While the most common voting method is the plurality rule – also known as first-past-the-post – by February of 2026 an alternative voting method known as ranked-choice voting (RCV) has been implemented in nearly 50 American jurisdictions for elections (FairVote, 2026). Ranked choice voting has been argued to have the potential to revitalize the American democracy, grounded in the insights provided by social choice theory, which highlight its benefits over the plurality rule currently widely in use (Maskin, 2022).

2.2 Preference Aggregation, Social Choice and Voting rules

The terms preference aggregation rule, social choice rule, and voting rule can seem almost interchangeable and indistinguishable, but there are subtle differences in how these terms are used. Preference aggregation rules refer to functions, which generate collective or social preference relations based on individual preferences. Social choice rules differ in that they generate a single winning alternative based on reported individual preference profiles. These reported preference profiles must not necessarily match with the "true" preference profiles. A voting rule can be considered the real-world practical application of a social choice rule, where agents are voters who cast their votes by submitting ballots. (List, 2022)

To make the presentation more formal, there is a set of agents $N = \{1, 2, \dots, n\}$, (alternatively individuals, decision-making units or voters) and a set of alternatives A with m distinct alternatives e.g., $A = \{x, y, z\}$. Typically, in social choice theory both the set of agents and the set of alternatives are finite.

A subset R_i of $A \times A$ refers to a binary weak preference relation on A, where "xR_iy" means that $(x, y) \in R_i$. The binary relation is a weak ordering of A when it is transitive and complete. The set of all such weak orderings on A is denoted by $\mathcal{R}$. The preference relation for individual agents is denoted with subscripts $i \in N$, where xR_iy is read as "alternative x is at least as good as alternative y for agent i". Completeness means that the binary relation is defined for all pairs of alternatives in A. Transitivity means that for any three alternatives $\{x, y, z\} \in A$, if agent i prefers x to y and y to z, they will also prefer x to z. Formally presented:

Definition 2.2.1. Completeness: $\forall x, y \in A, xR_iy \vee yR_ix$

Definition 2.2.2. Transitivity: $\forall x, y, z \in A, (xR_iy \wedge yR_iz) \Rightarrow xR_iz$

The strict preference relation P_i and the indifference relation I_i can be derived from R_i:

Definition 2.2.3. Strict preference: $xP_iy \Leftrightarrow xR_iy \wedge \neg(yR_ix)$

Definition 2.2.4. Indifference: $xI_iy \Leftrightarrow xR_iy \wedge yR_ix$

When R_i is a weak preference ordering on A the strict preference relation P_i is complete and transitive on A as well.

A combination of weak preference orderings of all individuals $i \in N$, for some alternatives A is called a preference profile $R = (R_1, R_2, \dots, R_n)$. The set of all possible preference profiles is denoted as $\mathcal{R}^n$, and alternative profiles are distinguished by R, R', R''. A preference aggregation rule is a function $\mathcal{D} : \mathcal{R}^n \to \mathcal{R}$, that generates a social preference ordering R_S for a given preference profile R. An instance of the social preference relation xR_Sy can be read as 'x is weakly *socially* preferred to y". When considering an alternative profile R', the resulting social preference is denoted R'_S. Similarly to the individual relations, the social strict preference and social indifference relations derived from R_S are denoted by P_S and I_S respectively.

Notations $b_i(A, R)$ and $w_i(A, R)$ are used for the *best* and *worst* alternatives of A for agent i at some profile i.e., the most and least preferred alternatives. A lower contour set is denoted by $L_i(x, R) \equiv \{y \in A \mid xR_iy\}$, referring to the set of alternatives in A that agent i finds no better than x under preference profile R.

A social choice rule is a function $f : \mathcal{R}^n \to A$, which generates a "winning" alternative for a possible preference profile. For a preference profiler R, the outcome of the function $f(R) = x$, where $x \in A$, is the selected outcome. The difference between a preference aggregation function $\mathcal{D}$

and a social choice rule f is that the former produces a complete social preference ordering for the preference profile, whereas the latter only produces a single outcome.

An important aspect to note in the analytical framework presented above, which reflects the typical treatment of preferences in social choice theory, is that generally speaking preferences are considered to be *ordinal*, meaning that only the order of preferred alternatives matter, rather than the *intensity* of preferences. In other words, with ordinal preferences what is accounted for in preference aggregation is whether an agent prefers alternative x over y or not, instead of *how much* an agent prefers x over y.

Still, there are preference aggregation methods and voting rules that incorporate cardinal preferences and preference intensities, often by incorporating utility functions into the analysis. Here agents are assumed to have utility representations for their preferences over the set of alternatives i.e., $u : A \to \mathbb{R}, xR_iy \Leftrightarrow u_i(x) \geq u_i(y)$. A well-known example of such a preference aggregation rule would be the maximin rule – or the Rawlsian rule – in which the socially preferred alternative is the one that maximizes the utility of the worst-off person i.e., $\max_{x \in A} \min_{i \in N} u_i(x)$ (Sen, 1984). Another voting mechanism which incorporates cardinal preferences is quadratic voting, in which preference intensities are modelled by allowing voters to vote multiple times for a single cause, but with a quadratically increasing cost per vote. (Lalley & Weyl, 2018).

Arrow, who developed the core framework of modern social choice theory argued and decided against using cardinal preferences in preference aggregation. Arrow argued that utilities, or other methods of assigning preference intensities, are not observable or measurable. Therefore *a fortiori* there cannot be meaningful interpersonal comparability of utilities – if something is not observable, it cannot be comparable either (Arrow, 1951/2012). It seems impossible to analytically ground utility values into anything, without seeming arbitrary – any pattern of individual behaviour that would be consistent with a method of "cardinalization", would not only be consistent with that particular method, but plenty of other utility functions as well (Sen, 1984). What is observable, is how people behave, whether they choose alternative x over y, or y over z. The earlier framework of binary relations could be presented with utility functions instead, by substituting expressions such as xR_iy with $u_i(x) \geq u_i(y)$, but Arrow deemed that "if we are concerned with ordinal properties, it seems better to represent these directly" (Arrow, 1951/2012).

In the field of social choice theory there has been some disagreement about restricting the analysis to merely ordinal preferences. For example, Nobel laureate Amartya Sen has developed different

methods for interpersonal aggregation, interpersonal, and comparability of welfare in the context of collective choice (Sen, 1984). Additionally, Sen argued that if the purpose of a collective preference aggregation method is to establish what is socially desirable, instead of deciding what alternative to choose, then incorporating cardinality could be justifiable or even necessary (List, 2022).

Aki Lehtinen has argued for the importance of preference intensities, even when voting rules do not allow voters to explicitly report these intensities. Lehtinen claims that even in ordinal voting rules individuals can in a way express their preference intensities through strategic voting behaviour, thus critiquing the solely ordinal approach to studying preferences in social choice theory. (Lehtinen, 2011). Lehtinen's critique, and strategic behaviour more broadly, also challenges the observability of ordinal preferences. If more than two alternatives are allowed, and the possibility of strategic behaviour exists, observing ordinal preferences becomes ambiguous too.

Broader criticisms have been levied towards not only Arrow's framework of ordinal preferences, but the broader "welfarist-consequentialist" approach that is prominent in welfare economics and social choice theory e.g., for the lack of focus on interpersonal comparisons of preferences or on the concept of procedural fairness. (Suzumura, 2000)

2.3 Arrow's Theorem & Properties of Voting Rules

What should a good social choice rule look like i.e., what properties ought a social choice rule have, or what axioms should it be able to fulfil? In Arrow's analysis of preference aggregation (Arrow, 1951/2012), five criteria (or axioms) for a reasonable preference aggregation rule to satisfy were detailed:

> **Definition 2.3.1. Universal domain (UD)**: An aggregation rule satisfies universal domain if the set of admissible preference profiles is not in any way restricted i.e., for n voters the admissible domain is $\mathcal{R}^n$.
>
> **Definition 2.3.2. Weak Pareto principle (WP):** An aggregation rule F satisfies the weak Pareto principle if when all agents strictly prefer an alternative x over y, then x is socially strictly preferred to y. Formally, $\forall x, y \in A, [(\forall i \in N, xP_i y) \Rightarrow xP_S y]$.
>
> **Definition 2.3.3. Independence of irrelevant alternatives (IIA)**: An aggregation rule satisfies IIA if the social preference relation between pair of any two alternatives depends only on the individual preferences between these two alternatives. Formally, $\forall x, y \in A, \forall R, R' \in \mathcal{R}^n, [(\forall i \in N, xR_i y \Leftrightarrow xR'_i y) \Rightarrow (xR_S y \Leftrightarrow xR'_S y)$.

Definition 2.3.4. Transitivity (T): An aggregation rule satisfies transitivity if the social preference ordering is transitive i.e., if xR_Sy and yR_Sz, then xR_Sz.

Definition 2.3.5. Non-dictatorship (ND): An aggregation rule is non-dictatorial if there exists no agent whose preferences alone determine the outcome of the function. Formally, $\neg\exists i \in N : \forall R \in \mathcal{R}^n, \forall x, y \in A, xR_iy \Rightarrow xR_Sy$.

Arrow's impossibility theorem states that when the number of alternatives is over two, then no preference aggregation rule could satisfy all five axioms listed above.

Definition 2.3.6. Arrow's Impossibility Theorem: When $|A| \geq 3$, no preference aggregation rule can satisfy UD, WP, IIA, T and ND.

In other words, when the number of alternatives is over 2, any aggregation rule that satisfies UD, WP, IIA and T is necessarily dictatorial (Arrow, 1951/2012). Therefore, the theorem seems to show that aggregation of preferences in a sensible manner is impossible.

All these five axioms are quite intuitive, and it is easy to agree that violating them would imply a fault of some kind in the aggregation of preferences. If UD were to be violated, it would in practice mean that some particular preference ordering(s) would not be "allowed", either for some particular agent(s) or all agents. In order to violate the WP principle, all agents would have to strictly prefer x over y, but collectively y would be preferred over x based on the preference aggregation rule. Transitivity assumes the same consistency that individual preferences are assumed to have – violating it would imply a non-decisive, cyclical preference ordering. Violating ND would imply that the collective preference relation would be exclusively based on the preferences of a particular single agent. (Arrow, 1951/2012; Penn, 2015).

Possibly the least intuitive out of the axioms is IIA. In order to violate IIA, individual preferences between two alternatives x and y would stay the same for all agents, but the collective preference relation between x and y would change. In other words, the truth value for xR_iy would remain the same for all agents $i \in N$ – some could prefer x over y, some could prefer y over x – and still the collective preference ordering would change. In practice IIA can be violated by aggregation rules where an introduction of a new, "irrelevant", alternative z – or a change in the individual preferences between x and z or y and z – could change the collective preference relation between x and y. (Penn, 2015)

One approach to dealing with the implications of Arrow's theorem is to "circumvent" the result by relaxing some of the axioms e.g., by not allowing universal domain of preferences or by relaxing the requirements of transitivity. However, these approaches often raise their own issues (Morreau, 2019; Penn, 2015).

While Arrow's theorem is purely analytical in its structure, its implications extend well beyond the field of logic. The theorem has sparked discussions related to the legitimacy of voting and democracy, much of which has been conducted by philosophers, political scientists, and social scientists alike (Ingham, 2019). In the field of welfare economics and social choice theory, Arrow's new axiomatic approach to analysis of preference aggregation inspired many others, and "descendants" of Arrow's theorem have been created (Penn, 2015).

In addition to Arrow's axioms there are other properties that generally are perceived as good for preference aggregation and social choice rules to have (Brandt et al., 2016; Heckelman, 2015).

> **Definition 2.3.7. Anonymity:** A SCF satisfies anonymity when all agents' preferences are treated equally. In other words, if the same preference profiles were reported, or same voted casted by different people, this would not affect the outcome of the SCF.
>
> **Definition 2.3.8. Neutrality:** A SCF satisfies neutrality when all alternatives are treated equally. In other words, renaming or permutating the candidates does not change the outcome of the SCF.
>
> **Definition 2.3.9. Resoluteness:** A SCF satisfies resoluteness when it always produces one unique winner regardless of the (reported) preference profile, formally $|f(R)| = 1$.
>
> **Definition 2.3.10. Reinforcement:** A SCF satisfies reinforcement axiom if two distinct sets of agents select the same set of alternatives, then combining the sets of agents should not alter the outcome of the SCF. Formally, for two distinct sets of agents N, N' with preference profiles R, R': If $f(R) \cap f(R') \neq \emptyset \Rightarrow f(R) \cap f(R') = f(R + R')$.
>
> **Definition 2.3.11. Participation**: A SCF satisfies the participation axiom if an agent can never attain a worse outcome by casting a sincere vote instead of not voting. Formally, $f(R) R_i f(R')$, where $R' = R \setminus R_i$.

Participation is a unique axiom compared to the ones presented previously, as it assumes that voters can abstain from voting. Failure to meet the participation axiom leads to the so-called *no-show*

paradox, where an alternative would win if a voter, or a group of voters, who prefer that alternative would abstain from voting.

> **Definition 2.3.12. Condorcet criterion**: A SCF satisfies the Condorcet criterion if, whenever an alternative that is preferred over all other alternatives in pairwise comparisons exists, that alternative is the outcome of the function. Formally, $\forall R \in \mathcal{R}^n, (\exists x \in A\ , \forall y \in A \setminus \{x\}, |\{i \in N : xP_i y\}| > |\{i \in N : yP_i x\}|) \Rightarrow f(R) = x$

An alternative that is preferred over all other alternatives in pairwise comparisons is called the Condorcet winner. Condorcet paradoxes refer to instances where a Condorcet winner does not exist, as the collective preferences with regards to pairwise comparisons are cyclic.

> **Definition 2.3.13. Monotonicity:** A SCF satisfies monotonicity when a selected alternative stays the same, after its position has been raised in one or more preference orderings. Formally, monotonicity $\forall x \in A, \forall i \in N, \forall R, R' \in \mathcal{R}^n, \left(R_{-i} = R'_{-i} \wedge L_i(x, R) \subseteq L_i(x, R')\right) \Rightarrow (f(R) = x \Rightarrow f(R') = x)$.

Failures of monotonicity are situations where the initial outcome of a social choice function is no longer selected despite rising in the preference orderings of one or more individuals. This is referred to as an upwards monotonicity failure. In contrast, a downwards monotonicity failure refers to a situation where an alternative that was not initially selected becomes the outcome after being lowered in the preference orderings of some individuals.

As no voting rule can fulfil all the desirable criteria listed above, it is reasonable to state that the goal of a SCF is not necessarily to fulfil as many of them as possible. There are inherent trade-offs between fulfilling desirable axioms. For example, Moulin (1988) showed that for 4 or more alternatives, satisfying the Condorcet principle necessarily implies a violation of the participation condition, resulting in the no show paradox. Additionally, practical considerations are worth considering in the real-world, such as the simplicity of a choice rule i.e., how easy the choice rule is to understand and implement (Heckelman, 2015).

2.4 Gibbard-Satterthwaite Theorem & Manipulation

One property of social choice functions not mentioned in the previous chapter is *strategyproofness*.

> **Definition 2.4.1. Strategyproofness (SP):** A social choice function is strategyproof if no agent can obtain a more preferable outcome by misrepresenting their

preferences regardless of preferences reported by the other agents, meaning that reporting sincere preferences is a *weakly-dominant strategy* for every agent $i \in N$. Formally, $\forall i \in N, \forall R_i, R_i' \in \mathcal{R}, \forall \mathcal{R}_{-i} \in \mathcal{R}^{n-1}, f(R_i, R_{-i}) R_i f(R_i', R_{-i})$.

A strategyproof SCF is therefore one, where for all possible preference profiles, no agent can benefit from reporting insincere preferences. Similarly, a *manipulable* choice rule can be defined as one where there exists a sincere preference profile, where for some agent can benefit by not reporting their sincere preferences.

The Gibbard-Satterthwaite theorem, based on the works of Allan Gibbard (1973) and Mark Satterthwaite (1975) is a seminal theorem in the field of social choice theory, which forms the very core of studying voting rule manipulability.

Definition 2.4.2. Gibbard-Satterthwaite theorem: When $|A| \geq 3$, no resolute SCF can satisfy UD, SP and ND.

The theorem essentially states that when the number of alternatives exceeds two, any SCF that allows for universal domain of preferences, is either not strategyproof or dictatorial – alternatively, the only strategyproof choice functions are dictatorial.

It is clear that dictatorial rules are trivially SP. When one agent $i \in N$ always gets the alternative that they most prefer chosen, they have no incentive to misrepresent their preferences. For all other agents there are no incentives misrepresent their preferences either – regardless of what preferences they announce, whether sincere or not, it has no effect on the selected outcome of the choice rule.

Since a dictatorial rule can quite self-evidently be ruled out as not socially desirable, and limiting alternatives to 2 faces its own problems (what voting rule or other mechanism ought to be used when deciding which these 2 alternatives should be), it can be concluded that an implication of the Gibbard-Satterthwaite theorem is the inevitability of manipulability. This is a drastic result, in that just as with Arrow's theorem, it shows the impossibility of designing a "good" social choice rule – at least from the perspective where misrepresentation of preferences is deemed undesirable.

However, it is important to note that the theorem only shows a negative result – meaning, that the theorem only shows that the *possibility* for manipulation exists within all non-dictatorial choice rules when $|A| > 2$. In other words, there are preference profiles where no one has an incentive to manipulate e.g., if everyone prefers some alternative $a \in A$ over all other, that is $\forall i \in N, L_i(a) = A$, then, for all choice rules that satisfy WP, no one has an incentive for misrepresenting their

preferences. Therefore, as manipulability is not necessary for all preference profiles, another question altogether is how common this possibility for manipulability is. This topic of measuring and comparing voting rule manipulability is explored further in the third chapter of the thesis.

Similarly to Arrow's impossibility theorem, the question arises as to if it would be possible to "circumvent" the GS-theorem? As mentioned before, neither accepting a dictatorial rule nor limiting the number of alternatives seem too appealing. One approach that has been used is to question the other assumptions of the theorem, namely the assumption of universal domain (UD).

The most well-known restriction of preference domain is Duncan Black's (1948) single-peaked preferences (SPP). For a preference profile to be single-peaked, the set of alternatives is assumed to be ordered by some dimension, and all agents' preferences have a "peak" alternative $a \in A$, such that when alternatives get more "distant" from this peak, they become less preferred. Moulin (1980) showed that by restricting the set of possible preference profile to just single-peaked ones, a category of voting rules, so-called generalized median voter rules, can be shown to be strategyproof.

Since SPP clearly violates UD, these findings do not contradict GS-theorem. But when it comes to circumventing manipulability, SPP shows that forming a non-dictatorial strategyproof choice rule in restricted domain of preferences is possible. Another question altogether is how realistic an assumption SPP is in the real world, especially with cases that do not have a non-arbitrary way of ordering the alternatives. However, in political voting contexts SPP can be motivated by referring to a "left-right"-axis, where alternatives are ordered based on their political leaning, and all agents are assumed to have SPP over this ordered set of alternatives. (List, 2022)

In this thesis the term manipulation refers to voters misrepresenting their preferences, and more specifically, to voters gaining a more individually preferable outcome by this action of reporting *insincere* preferences. Manipulation in the context of sequential voting can also be used to refer to *agenda manipulation*, where the order in which different proposals are voted for can alter the outcome of a voting process (List, 2022). Strategic nomination is a term used to refer to adding or removing alternatives in order to alter the outcome of the choice rule (Green-Armytage, 2014). Additionally, in real-world elections the term manipulation could refer to a variety of phenomena e.g., voter fraud, voter intimidation, misinformation campaigns etc., which are outside the scope of this study.

So, in this thesis, a voting rule that is manipulable is one that is not strategyproof – and manipulation refers to the act where an agent misrepresents their preferences in order to change the outcome of the voting rule. Formally, a successful manipulation is attained by an agent $i \in N$ when with sincere preference the outcome of the choice rule would be $f(R) = a$, but by reporting an alternative preference profile R', where the only difference is in the reported preference of agent i, the agent can manipulate the outcome to a more personally preferable one $f(R) = b$, such that $bR_i a$.

Intuitively, the possibility for an individual to manipulate election results nearly vanishes as more and more voters are introduced, since a single agent's vote carries less impact. When the number of voters increases, the probability that a voter is the pivotal voter – one that can change the election outcome by altering their vote – decreases. (Slinko & White, 2013).

Therefore, the study of manipulability does not only focus on individual voters. The term *coalitional manipulability* is used to refer to a collection of agents reporting their preferences insincerely in order to change the outcome of the choice rule. Coalitional manipulability can be distinguished from *individual manipulability*. When analysing coalitional manipulation, the mechanism of coalition forming may or may not be explicitly detailed, i.e. how agents come together and form these strategies of manipulation (Slinko & White, 2013). However, in some situations it does make sense to assume that the process of coalition forming does not require explicit coordination, but that agents who manipulate assume that others with similar preferences could do the same (Peters & Veselova, 2023).

An aspect of manipulation that has been studied is related to the "safety" of misrepresenting preferences. In an attempt of manipulation via misrepresenting preferences, the goal of an agent is to change the outcome of a voting rule to a personally more favourable one. But it is conceivable that this attempt at manipulation could "backfire", and lead to an outcome that the agent prefers less, compared to the outcome that would have been attained via sincere voting behaviour.

When casting a strategic vote could lead to a less favourable outcome, it is called *unsafe*. Strategic *overshooting* refers to an instance where too many individuals voted strategically, whereas strategic *undershooting* to one where too few individuals voted strategically, and as a result ended up with an alternative that they deem inferior. A *safe* strategic vote is a strategy, where an agent doesn't vote according to their true preferences, but this vote cannot lead to a worse outcome than a sincere vote. (Slinko & White, 2013)

The Gibbard-Satterthwaite theorem can be extended to show that any non-dictatorial social choice rule can have the possibility for a safe strategic vote (Slinko & White, 2013). The probabilistic (Wilson & Reyhani, 2010) and computational (Hazon & Elkind, 2010) aspects of safe strategic voting have been studied, as well as safe manipulation by coalitions (Peters & Veselova, 2023).

The notion of threats and counter-threats in the context of choice rule manipulability was introduced by Pattanaik (1976). A threat refers to a possibility of manipulation by an individual agent, and a counter-threat to the idea of other agents forming a coalition to counter this potential attempt at manipulation, by implementing a strategy which leads to a worse outcome for the initial threatener. These concepts form the basis of a less strict definition of stability compared to strategyproofness, where individuals who would have the incentive to manipulate under the typical framework, would refrain from doing so due to the existence of a counter-threat.

Broadly speaking manipulation via preference misrepresentation can be divided into two distinct categories. *Compromising* means that an agent ranks an alternative higher in their reported preferences compared to how they sincerely would, in order to get that alternative selected. Conversely, *burying* refers to the act of lowering an alternative in reported preferences, in order to get another alternative selected – or to lower the chances of the "buried" alternative getting selected. (Green-Armytage, 2014)

3 Comparative Analysis of Voting Rule Manipulability

This chapter centres on voting rule manipulability, and the consequences of manipulation. First, a number of different voting rules common in the literature will be introduced and formally defined. Then, examples of manipulation will be presented under three different voting rules. Afterwards, different measures of manipulability will be examined, and the assumptions and validity of these methods considered. Lastly, some studies focused primarily on the outcome analysis of voting rule manipulation – the welfare effects of manipulation – will be reviewed.

3.1 Voting Rules

There are many different voting rules that have been studied in the social choice literature, and that are used in different collective decision-making processes around the globe.

> **Definition 3.1.1. Plurality Rule:** The selected alternative is one that is ranked first by largest number of agents i.e., agents vote for one alternative, and the alternative receiving most votes gets selected.

> **Definition 3.1.2. Antiplurality (Veto) Rule:** Agents vote *against* one alternative, the alternative with the least vetoes wins. In other words, agents vote for all alternatives except one, and the alternative with most votes wins.

> **Definition 3.1.3. Borda's Rule:** Agents rank all alternatives, from which all alternatives get assigned a *Borda score* – first ranked alternative gets $m-1$ points, second $m-2$, and so on, until last alternative gets 0 points. The *Borda count* of an alternative is the sum of the points from all agents. Formally, when $r_i(x) = |\{y \in A \mid xP_iy\}|$. The Borda count of alternative a is $\sum_{i \in N} r_i(a)$.

Plurality, Veto, and Borda can all be classified into a group of voting rules called *scoring rules*, in which all alternatives are given a score determined by different *scoring vectors* and (reported) preference profiles. A scoring vector $s = (s_1, \dots, s_m)$ determines how many points the jth alternative in a voter's preferences receives. The plurality rule can be represented as the scoring vector $s = (1, 0, \dots, 0)$, the antiplurality (veto) as $s = (1, \dots, 1, 0)$, and Borda as $s = (m-1, m-2, \dots, 1, 0)$. (Y. A. Veselova, 2020)

> **Definition 3.1.4. Hare's Procedure:** Agents rank all alternatives. If an alternative receives the majority of first place rankings, this alternative wins. Otherwise, alternative with the

smallest number of first place rankings a gets removed, and the procedure is repeated without alternative a.

Hare's procedure is also known as instant-runoff voting (IRV) and often referred to as ranked choice voting (RCV). Single transferable vote (STV) refers to the multi-winner version of instant-runoff voting, which differs in that "surplus" votes from winning candidates may be "transferred" as well.

Definition 3.1.5. Coombs' Method: Otherwise, similar to Hare/IRV, but the alternative ranked last by the most voters is eliminated.

Definition 3.1.6. Nanson's Procedure: The Borda count of all alternatives is calculated. Alternatives that are below the mean Borda count are eliminated. The procedure is repeated until a single winner remains.

Definition 3.1.7. Black's Method: Agents rank all alternatives. The method picks the unique Condorcet winner if one exists, otherwise the Borda winner gets selected.

Definition 3.1.8. Copeland's method: Agents rank all alternatives. Pairwise comparisons of alternatives are conducted, and a winner of such a pairwise comparison i.e., the alternative ranked higher than the other by a larger number of agents receives one point. In case of a tie, both alternatives receive 0.5 points. The *Copeland score* of all alternatives is calculated, and the alternative with the highest score gets selected.

Different rules can be "extended" with an alternative tie-breaking method. For example, in the *second-order Copeland scheme*, the Copeland method is calculated as usual – but in the event of a tie, the sum of these tie-breaking alternatives' defeated competitors' Copeland scores is calculated, and the alternative whose defeated competitors have the highest total score wins (Bartholdi et al., 1989). In the *Condorcet-Hare* method, which selects the Condorcet winner – except in the case of a tie, where it selects the Hare winner (Green-Armytage et al., 2016).

Definition 3.1.9. Approval Voting: Agents can rate each alternative with either 0 or 1. The alternative with the most points get selected.

Approval vote differs from other rules in that the votes of individuals cannot be deduced from an ordinal preference profile alone. Two different agents can have the same preference orderings e.g., but a different "threshold" for approval.

3.1.1 Examples of Manipulation

The Borda rule can be used to provide a simple example of individual manipulation. Assuming the following preference profile with three voters, and four alternatives $A = \{a, b, c, d\}$:

	Voter 1	Voter 2	Voter 3
1st choice	a	a	b
2nd choice	b	b	a
3rd choice	c	c	c
4th choice	d	d	d

Table 3.1 Example voter preferences with four alternatives

With the preferences in Table 3.1, the Borda counts for alternatives are: $a = 8, b = 7, c = 3, d = 0$. But Voter 3 has a strategic incentive: by switching the placement of alternatives a and d in their rankings, the outcome of the election changes to b, which is personally more preferable for voter 3. This can be verified by calculating the new Borda counts: $a = 6$, $b = 7, c = 3, d = 2$.

For plurality voting, individual manipulation relies on the tie-breaking rule used. For a single voter to change the outcome of an election, there either needs to be a tie or a possibility for the voter to generate a tie. Assuming that the preference orderings are strict, consider a simple example with the following preference profile:

	Voter 1	Voter 2	Voter 3
1st choice	a	b	c
2nd choice	b	c	a
3rd choice	c	a	b

Table 3.2 Example voter preferences with three alternatives (a)

The election results in a three-way tie under sincere voting. If the tie-breaking rule is alphabetical, a is selected as the winner, and voter 1 has no incentive to strategize. However, voter 2 does have an incentive to cast their vote for c, resulting in c winning the election, therefore generating a more personally favorable outcome for voter 2, as cR_2a. But with other tie-breaking rules the incentives for strategizing could differ.

In the real world the strategic incentives under plurality voting are more common with coalitional manipulation, as the possibility for one voter being a pivotal voter that has a strategic incentive is minimal in larger electorates. Coalitional manipulability can be represented by depicting the distribution of preference orderings in the society.

	Type 1 voter (40%)	Type 2 voter (45%)	Type 3 voter (15%)
1st choice	a	b	c
2nd choice	b	c	a
3rd choice	c	a	b

Table 3.3 Example voter preferences with three alternatives (b)

In the sincere preferences presented in Table 3.3, 40% of the electorate have the preferences of a type 1 voter. In this sincere election b is the winner with a plurality of the vote. However, voters of type 3 have an incentive to engage in "lesser-evil" strategic voting, by switching their vote to a. If a coalition of more than $\frac{1}{3}$ of the type 3 voters (i.e., more than 5% of the total voting population) engage in such strategizing, a receives the plurality of the vote and wins the election, which for type 3 voters is a preferred outcome.

Under ranked-choice voting (RCV) the outcome of the above election in Table 3.3 with sincere preferences would be a. As no alternative receives the majority of the first-place votes, the alternative with the least first-place votes gets eliminated. Since c has the least first-place votes (15%), type 3 voters have their votes transferred to their second choice, a. In the second round a wins the election with 55% of the vote.

However, under RCV the Type 2 voters have an incentive to strategize in the election above. If more than 55.2% of the Type 2 voters (more than 16% of the total voting population) raised c in their rankings over b, the elimination order would change as b would receive least first place votes.

	Type 1 voter, 40%	Type 2 voter, 29% (sincere)	Type 2 voter, 16% (strategic)	Type 3 voter, 15%
1st choice	a	b	c	c
2nd choice	b	c	b	a
3rd choice	c	a	a	b

Table 3.4 Example voter preferences with three alternatives (c)

Here proportion of first-place votes is as follows: a: 40%, b: 29%, c: 16 + 15 = 31%. Thus, b gets eliminated first, and the sincere 2nd place votes for c from Type 2 voters transfer to alternative c. In the final round, alternative c wins the election with 60% of the vote.

The sincere preferences detailed in Table 3.3 showcase how, even in a simple example, the strategic incentives and outcomes can vary. Under plurality voting, b is the since outcome – whereas strategic incentives under plurality, can lead to a winning the election. Under RCV, a is the sincere election winner, but in the strategic example c is the winner. All three alternatives can win,

highlighting that justifying a particular election outcome as the legitimate winner can be difficult, especially in an instance like this that lacks a Condorcet winner.

Lastly, a common feature in the examples above is that they deal exclusively with ordinal preferences, as is usually the case in social choice theory. But when it comes to the question of assessing the impacts of strategic behaviour, incorporating cardinal preferences e.g., in the form of utilities, could alter how the manipulations presented above are perceived – and what the socially most desirable alternative would be. Some manipulations could substantially increase or decrease the aggregate social welfare, as individuals or voters of different types could have different utility profiles correspond to their preference orderings.

3.2 Measuring Manipulability

Different measures of voting rule manipulability have been developed within social choice theory. The aim of such a measure is to provide a numerical indicator for how likely it is for a choice rule to be manipulated. Often the question of interest is in comparing the manipulability values yielded for different voting rules, thus providing a way of comparing different voting rules on their manipulability.

Compared to the desirable properties listed in chapter 2 of this thesis, this method of numerical estimation offers an alternative approach to evaluating voting rules. Rather than merely listing out which axioms different voting rules satisfy; manipulability measures allow for a more proportional way of comparing choice rules. In fact, similar numerical measures have been used for other properties of voting rules e.g., the utilitarian efficiency of voting rules (Green-Armytage et al., 2016) and the frequency of electing Condorcet winners i.e., the Condorcet efficiency of a voting rule (Gehrlein et al., 2011).

An intuitive way to measure manipulability is to calculate the proportion of profiles that are susceptible to manipulation for a given voting rule. As the Gibbard-Satterthwaite theorem suggests, at least one profile within the set of all possible profiles is manipulable – but the theorem makes no statement on how many profiles are manipulable.

This measure of manipulability is often referred to as the Nitzan-Kelly Index (NKI) after the works of Nitzan (1985) and Kelly (1988, 1993), or as the *tainted measure* by Smith (1999). A *tainted* profile is one where at least one agent can benefit from misrepresenting their preferences. Denoting the set of all possible tainted profiles as K, the NKI for a given choice rule f, number of agents n,

and number of alternatives m is $|K|/(m!)^n$ where $|K|$ is the cardinality of the set K, and $(m!)^n$ is the number of possible preference profiles (Smith, 1999).

Even the simplest measures of manipulation, such as the NKI, rely on a number of core assumptions related to the behaviour and knowledge of the voters, distribution of preferences, as well as other factors.

Perhaps the most significant and discussed assumption of these is related to the preferences of agents, which are often modelled by a *statistical culture*, a probabilistic model of preference distribution. The distribution of preferences clearly affects the manipulability measures yielded by different indices – if a preference profile that is manipulable is deemed as more common than other profiles, this will impact the manipulability measure as a whole. The most common of these "cultures" is the *Impartial Culture* (IC), first developed by Guilbaud (1952). In IC preference profiles are used – it is explicitly modelled how each of n voters linearly rank each of m alternatives. Each of these possible $(m!)^n$ profiles are treated as equally likely – meaning that every possible combination of preferences is considered to have the same probability of being realized. (Eğecioğlu & Giritligil, 2013; Y. Veselova, 2012).

As under IC the number of profiles to be considered rapidly increases both with the number of alternatives and with the number of agents, this quickly creates some difficulties for larger values of n and m. One way of simplifying the analysis is to represent voter preferences anonymously, such that the names or identities of any individual voters are irrelevant. Then instead of focusing on preference profiles, where all individual preference orderings are known, only the preference profiles which differ in their permutations are considered. These anonymous preference profiles, also known as voting situations, are treated as equally likely under the *Impartial Anonymous Culture* (IAC), first described by Gehrlein & Fishburn (1976). The IAC method clearly only works for analysing anonymous voting rules, but as anonymity is often deemed to be a desirable property for a voting rule to have, this is not a significant restriction. The number of unique voting situations is given by the binomial coefficient $\binom{n+m!-1}{m!-1}$, which is smaller for any given n and m to the number of unique preference profiles in IC. (Eğecioğlu & Giritligil, 2013; Y. Veselova, 2012)

m	n	IC profiles	IAC situations
2	2	4	3
2	10	1 024	11
3	2	36	21

3	10	6 046 176	3 003
3	100	6,5332E+77	96 560 646

Table 3.5 Number of distinct voting situations under IC and IAC

Final variation of the impartial culture model is the *Impartial, Anonymous, and Neutral Culture* (IANC) model, developed by Eğecioğlu & Giritligil (2013), in which both the names of the voters as well as the names of the alternatives are immaterial i.e., not considered when distinguishing between preference profiles. In this model, preference profiles are partitioned into Anonymous and Neutral Equivalence Classes (ANECs). Similar to profiles under IC, and situations under IAC, each class is considered to be equally likely under IANC. Any preference profile within an ANEC can be considered as the representative profile, that is referred to as the *root* of the ANEC. (Eğecioğlu & Giritligil, 2013; Y. Veselova, 2012)

The Impartial Culture models are perhaps the most widely used ones but have also been heavily scrutinized in the past. The IC model is widely acknowledged to be unrealistic, as in the real world it would not be expected that all preference profiles are equally likely. However, the results of studies using IC are still treated as valid. Often the interpretation and justification provided for this ostensible discrepancy is that results using IC are treated as the theoretical worst-case scenario, in a wide variety of possible preference distributions for agents. (Tsetlin et al., 2003; Wollesen, 2025)

Often in the real world there are correlations between votes, and therefore some voting profiles might be more common than others. This can be referred to as *social homogeneity*, the tendency for voters to have similar preferences. The *Pólya-Eggenberger model*, is one method to model this phenomenon. This model is also known as the Pólya urn model, since its behaviour can be explained through an urn filled with all possible $m!$ preference profiles. In the model, a vote is randomly drawn out of the urn and put back along with α instances of the same profile. For $\alpha = 0$ the model resembles IC, and for $\alpha = 1$ it resembles IAC – whereas higher values of α are a way to model increasingly homogenous preferences. (Lepelley & Valognes, 2003; Walsh, 2010a)

Another way of modelling preferences with homogeneity is the Mallows model, first described by Mallows (1957). In the model, one "true" preference ordering σ is assumed. The probability of any other ranking r is determined via $P(r) = \frac{1}{Z}\phi^{d(r,\sigma)}$, where $\phi = (0, 1]$ is a dispersion parameter, Z is a normalizing constant and d is the Kendall's τ-distance between σ and r. When $\phi = 1$, this model is equivalent to IC. However, for other values of ϕ, the votes are typically highly correlated, and for a high number of candidates even unanimous. The Mixed Mallows model is a way to model

multiple different reference σ_i, suitable for modelling political polarization in an electorate. (Ianovski et al., 2022; Lu et al., 2012)

In spatial models, alternatives possess different attributes which form a multi-dimensional "attribute space". Different alternatives have specifiable quantities of each attribute and are located in the attribute space. Each voter has an "ideal point" in the attribute space, which is their most preferred amount for each attribute that an optimal alternative would possess – and an indifference map in the attribute space based on this ideal point. In a simple spatial model voters and alternatives are assumed to follow the same multivariate distribution on the attribute space. A voter's utility for a given alternative is the additive inverse of the Euclidean distance between the voter's ideal point and the alternative in this attribute space. (Green-Armytage et al., 2016; Tideman & Plassmann, 2014)

A more practically oriented method for producing preference distributions is to generate election data from real world election/survey data. This can require some interpretation, as the data is not necessarily equivalent to weak preference orderings. Green-Armytage et al. (2016) argue that such "synthetic election" construction is useful, and a way of bringing "quasi-empirical" models into the typically mathematically abstract analysis of voting rule manipulability.

Another important assumption of manipulability measures is related to the knowledge of other agents' behaviour. Models often assume perfect information of others' preferences and that these other agents are non-strategic in their behaviour – meaning that only a single agent will manipulate having full knowledge of the other agents' sincere preferences, and that these other agents will not alter their behaviour regardless of the consequences of this potential manipulation. When other agents are not strategic, in other words naïve, then knowledge of their preferences is knowledge of their voting behaviour. This approach differs starkly from a more game-theoretic perspective e.g., one where threats and counter threats are considered; where other agents would alter their behaviour if a potential attempt to manipulate could lead to a personally less favourable outcome (Favardin & Lepelley, 2006).

Since most of the literature deals with situations of perfect information, Eggers & Nowacki (2024) have argued that most studies have only measured *ex post* manipulability, as opposed to *ex ante* manipulation. Here *ex post* manipulation is defined as the "frequency with which voters would regret a sincere vote after the results are announced". Non-naïve strategic behaviour has received less attention in the literature. Most studies, even if they assume incomplete information, typically

only focus on if a manipulator or a manipulative coalition could exist, ignoring the potential game-playing by other voters (Favardin & Lepelley, 2006; Pritchard & Wilson, 2007).

Intuitively, as the number of voters increases, the probability that an individual voter can be a strategically pivotal voter decreases i.e., when the number of voters approaches infinity the individual manipulability of a voting rule approaches zero. This was first noted by Peleg (1979), and later substantiated by many others e.g., Nitzan (1985) and Favardin et al. (2002). The importance of studying *coalitional manipulability* stems from this finding, as well as from the intuition that people often in real-world elections would strategize as a group – either through an implicit or explicit process of coalition building. Therefore, another assumption tied to measuring manipulability is whether only individual or also coalitional manipulability is accounted for.

Measures of coalitional manipulability can be formed similarly to the NKI. For example, Lepelley & Mbih (1987) calculated the proportion of coalitionally unstable situations, where coalitional stability is defined as the non-existence of a *threat* to a preference profile. This measure is essentially the same as the NKI, just adjusted for coalitional manipulation. As singletons are often also considered to be coalitions, that is coalitions comprised of only one member, coalitional measures of manipulability often systematically exceed those of individual manipulability (Favardin et al., 2002). Lastly, the study of coalitional manipulability allows for a different type of measure for manipulability e.g., the minimum size of a manipulating coalition for a given voting situation (Pritchard & Wilson, 2007).

Differences in measures of manipulability might also result from the assumptions used in tie breaking. Commonly used assumption for the tie-breaking rule is the alphabetical tie-breaking rule, where in an event of a tie, the alternative that comes alphabetically first wins. Thus an alphabetical tie-breaking rule breaks the symmetry between alternatives (Aleskerov et al., 2011). (Pritchard & Wilson (2007) measured manipulability with a randomized tie-breaking rule and noted discrepancies in their results compared to previous studies that had used alphabetical tie-breaking. It is worth noting that randomized tie-breaking violates *resoluteness*, as the outcome of a voting rule might vary even if the votes are the same.

Alternatively, some studies have dealt with voting rules in the case of multiple choice, that is, where more than one alternative can be chosen. Here there is no need to choose a tie-breaking rule, but preference extension method(s) need to be used to distinguish between the personal preferability of different multi-valued outcomes. (Aleskerov et al., 2011, 2012).

Final common core assumption related to measuring manipulability are the number of voters and number of alternatives involved. These values can have various different effects on measures of interest e.g., as noted earlier, the probability for individual manipulability decreases as the number of voters increases – while increasing the number of alternatives rapidly increases the number of distinct profiles that need to be considered. In coalitional manipulability, the proportion of situations where an incentive to manipulate exists approaches one as the number of alternatives increases, for large numbers of voters (Lepelley & Mbih, 1987). Increasing the number of voters or alternatives might also give rise to computational issues, requiring either more efficient algorithms to calculate manipulable profiles, or having to rely on methods of statistical estimation.

Returning to the indices used in measuring manipulability, the NKI is not the only that has been developed. Smith (1999) presented some alternatives to the index. Particularly the *multiplicity weighted* measure is otherwise similar to the NKI, but all manipulable profiles are weighted by the number of agents that could manipulate the profile – placing a higher emphasis on profiles where more agents have an incentive to behave strategically. The intent of this measure is to counteract the imbalance that the NKI seemingly produces; by giving an equal amount of weight to those profiles where only one agent has an incentive to manipulate, having merely one agent with an incentive for strategizing "taints" the entire profile (Smith, 1999).

In part due to its wide use and popularity within the literature, NKI has received other criticism. Common critique is related to the fact that the index only measures the logical possibility of manipulation, rather than how common manipulation actually is. Without an account of how common different preference profiles are, and how often people in these instances recognize the potential to manipulate – and decide to act on this incentive – the results provided by the measure seem unrealistic. For example, Kube & Puppe (2009) found that in experimental situations, humans are relatively often found not to manipulate, even when presented with the simplest instances.

Therefore, it could be argued that either the NKI overestimates the prevalence of manipulation or alternatively does not say much about the actual prevalence of manipulation. Those in favour of the NKI may argue that the purpose of the index is only to provide a theoretical worst-case scenario of manipulation, rather than an accurate description of the prevalence of manipulation in real-world voting situations (Wollesen, 2025).

Early work in measuring coalitional manipulability comes from Chamberlin (1985). Chamberlin noted that the manipulability of voting rules can be divided into two separate questions: the logical possibility of manipulation, and the likelihood that manipulation would succeed – given it were

logically possible. Therefore, in the study Chamberlin developed four different measures for manipulability, the first of which measuring the logical possibility via calculating the percentage of election where manipulation is possible – essentially equivalent to the NKI, just for coalitional manipulation. The other three measures of the study are the average number of alternatives for which manipulation is possible, the average minimum coalition size required for manipulation, and the so-called average "margin-for-error" – measuring the possibility that a coalition would succeed in the intended manipulation, while failing to coordinate its ideal voting strategy.

Substantially different approaches from the NKI have been developed to measure manipulability. One such approach is to analyse the expected utility that can be gained by insincere behaviour; defining the susceptibility to manipulation as the maximum amount of expected utility that strategic behaviour can yield compared to sincere behaviour (Carroll, 2011). A similar measure was used by Eggers & Nowacki (2024), who show that this expected benefit measure can be decomposed into the amount of benefit that can be gained from insincere behaviour, as well as the proportion of agents that can benefit from misrepresenting preferences. Campbell & Kelly (2009) proposed a somewhat similar approach to measuring manipulation, involving a cost associated with manipulation – this cost could be viewed as the effort required to strategize or to gather information about other voters. If the potential gains of manipulation are small enough, then individuals would not make the effort to manipulate, and thus such a rule would be less manipulable. Similarly but without assumptions of utility values, Smith (1999) created the *improvement weighted* measure, where the amount of improvement an agent can gain by insincere behaviour is accounted for – this improvement being measured as the difference in the number of ranks within preferences between the insincere and sincere outcomes.

Other measures include the *expected measure* by Smith (1999), where manipulability is measured through calculating the proportion of insincere ballots which lead to a positive manipulation i.e., the probability that a randomly selected non-sincere vote leads to a personally more preferable outcome. Reyhani et al. (2009) criticized the more traditional measures of logical possibility of manipulation for failing to measure the effort required to form and coordinate a coalition – thus proposing different classes of measures based on the minimum size of a manipulating coalition, the number of manipulating coalitions per a given number of members, and the informational effort required to find a manipulative coalition.

A unique approach to measuring manipulability could have come from the field of matching theory, where strategyproofness is deemed a desirable property as well. Pathak & Sönmez (2013)

developed a measure where a mechanism, such as a voting rule, f is *at least as* manipulable as another mechanism g, if any profile that is vulnerable under f is also vulnerable under g. A mechanism f is *more* manipulable than mechanism g if f is at least as manipulable as g *and* there exists a profile where f is manipulable but g is not. Teplova & Ianovski (2022) attempted to utilize this measure in the field of voting but it failed – while this measure would sidestep some of the problems related to measuring manipulability e.g., accurate assessments of preference probability distributions, most voting rules were found to be incomparable under a measure like this.

One quite popular alternative approach to measuring voting rule manipulability comes from the field of computational social choice, where computational complexity is used as a proxy for manipulability. If finding a method to manipulate a particular voting rule requires extensive computational effort, such a rule could be deemed as less susceptible to manipulation. An advantage of this approach is that even when free and perfect information are assumed, a voting rule can still be deemed to be not manipulable *in practice* (Bartholdi et al., 1989). Compared to the NKI and other similar measures, with computational complexity the aim is not to show that manipulation is logically impossible. The study of manipulation forms an interesting dynamic in computer science, since contrary to most situations, high computational complexity is considered in its context desirable (Conitzer et al., 2007). However, as it turns out, many commonly used voting rules are easy to manipulate (Bartholdi et al., 1989). But the study of manipulability from the perspective of computational complexity can extend beyond this notion of manipulation being easy or hard e.g., Bachrach et al. (2011) study the computational complexity of forming coalitions from a game-theoretic perspective.

The usefulness of computational complexity as a measure for manipulation has been discussed. Similar to the NKI, it is questionable as to how well this approach of studying worst-case results transfer on to practice where manipulation might often be easier. Proponents of computational complexity as a reliable measure of manipulability argue that it better maps onto the reasoning capabilities of agents – at the very least better than the NKI. Harrison & McDaniel (2008) found through an experimental study in a controlled laboratory environment support for the idea that computational "difficulty" of manipulating translates to "cognitive" difficulty to manipulate, creating a sort of behavioural incentive compatibility.

It has been found based on empirical results that almost every election could either be easily manipulated or easily proven not to be manipulable (Walsh, 2010b). This should raise the concern as to how relevant computational complexity is when it comes to manipulation in practice.

Additionally, the view that computational difficulty translates to cognitive difficulty could be questioned. Even if computational complexity maps on better than the logical possibility of manipulation associated with NKI, perhaps a heuristic model of voting behaviour, such as one proposed by Wollesen (2025), would even better reflect the cognitive aspects of human voting behaviour.

A novel approach to measuring manipulability is the "machine learnability" of manipulation. Holliday et al. (2025) studied whether neural networks of varying sizes would be able to learn how to manipulate under various conditions related to information. This approach could provide new insights into manipulability, especially if learnability by neural networks maps onto human learning – or if machine learning models are used in elections with non-human agents.

Given all of these indices for measuring manipulability, along with having to choose assumptions related to voter behaviour and general electoral circumstances, it can certainly be stated that coming to a definitive conclusion as to which voting rule is least manipulable is unreasonable. As Saari (1990) noted, any rule can be shown to be the least manipulable by tweaking these aspects: choosing the "right" set of rules, imposing assumptions on voters' preferences and selecting a specific kind of measure.

Additionally, it is important to remember that many of these measures do not deal with the "actual" prevalence of manipulation, but rather something like the logical possibility of manipulation (Green-Armytage et al., 2016). This increases the difficulty of forming an argument that one choice rule is in the real world less manipulable than others, at the very least on only the basis of a single manipulability index.

3.2.1 Compilation of Results

As there are multiple different assumptions that go into measuring manipulability, it is challenging to form a clear, unified picture of manipulability across different voting rules. However, there are some general common results between different studies. With the NKI, Borda is often found to be more manipulable, plurality a less manipulable than Borda but still relatively manipulable, and Hare among the least manipulable rules (Aleskerov et al., 2011, 2012, 2015; Green-Armytage et al., 2016; Nitzan, 1985; Smith, 1999). Depending on the rules studied, other ones are found to be among the least manipulable such as Nanson's (Aleskerov et al., 2011, 2015; Favardin & Lepelley, 2006) and Condorcet-Hare (Green-Armytage et al., 2016).

Similar results extend to other measures such as those of coalitional manipulability studied by Chamberlin (1985), who found Hare to be generally speaking the least manipulable, Borda the most manipulable, with plurality and Coombs in the middle. In other studies of coalitional manipulability, Favardin et al. (2002) confirmed Borda to be more manipulable than Copeland.

Despite its poor reputation of high manipulability, the Borda rule has had some advocates. Saari (1990) has found the Borda count to be among the least manipulable rules to a set of manipulations called *micro manipulations*. Favardin & Lepelley (2006) found that once non-naïve behaviour from initial non-manipulators is assumed, the performance of Borda significantly increases with respect to susceptibility to manipulation.

The impact of selected statistical culture on the values of manipulability indices is notable, even in the simplest instances of NKI with Impartial Culture models. Comparing IC to IANC, for smaller values of $m \in \{3, 4\}$, Veselova (2012) established that the relative ranking of the studied voting rules varied between the two cultures i.e., the order of voting rules with respect to manipulability differed between IC and IAC. For larger values of m, this ranking of voting rules stayed quite consistent between cultures. Aleskerov et al. (2015) showed that in the case of $m = 3$, NKI indices for IAC are smaller than IC. In around 20% of the studied instances the least manipulable rule differed between IC and IAC.

Lepelley & Valognes (2003) illustrated that both for small and large electorates, the NKI across eight studied voting rules tends to decrease as social homogeneity increases. Across all degrees of social homogeneity, Hare tended to score the lowest in manipulability.

When it comes to computational complexity, Bartholdi et al. (1989) demonstrated that common voting rules e.g., plurality, Borda and Copeland, are not NP-complete via a simple Greedy-Manipulation algorithm, arguing for the so-called second order Copeland method, which is NP-complete to manipulate. Bartholdi & Orlin (1991) proved that STV is computationally resistant to manipulation, similarly Conitzer et al. (2007) confirmed that STV is among the hardest to manipulate, while plurality among the easiest. However, Walsh (2010a) demonstrated that while STV is in theory computationally hard to manipulate, from an empirical perspective almost all elections are either computationally easy to manipulate, or easy to prove not manipulable.

Under incomplete information Veselova (2020) found the NKI to lose some of its interpretability – less information leads to a higher manipulability index, though with incomplete information there is no longer guarantee that this manipulation would succeed. Eggers & Nowacki (2024) in their

conception of *ex ante* manipulability, find that there are more opportunities to manipulate in IRV, but that the expected benefit of manipulation is higher under plurality. In other measures incorporating expected utility, Carroll (2011) found the plurality rule to perform quite well, even better than rules previously shown to be resistant to manipulation, such as Copeland and STV.

Even with these various measures, it is worth keeping in mind the axiomatic approach explained in chapter 2 of this thesis. Sometimes achieving lower manipulability might come with a trade-off problem. For example, the Hare method has been shown to be generally good against manipulation, but it notoriously fails to meet the *monotonicity* criteria. Similarly, the Copeland method, as well as all other Condorcet rules, suffer from the *no-show paradox* mentioned in chapter 2.3., which all positional rules e.g., Borda, are immune to. Furthermore, as Favardin et al. (2002) note that while Borda is more prone to manipulation than the Copeland method under IAC, there are voting situations where Copeland is manipulable but Borda not (and clearly vice versa) – so the conclusion that Borda is more manipulable than Copeland relies on the IAC assumption that all voting situations have an equal probability of occurring.

3.3 Welfare Impact of Manipulation

In large parts of the literature on manipulation, manipulability is viewed as a negative, undesirable feature of a voting rule, that ought to be avoided and minimized – this view is often not substantiated, but rather just implicitly assumed to be the case. However, an alternative approach is to accept manipulation as an unavoidable reality of voting, directing the focus to the outcomes yielded via manipulation.

A logical implication of non-strategyproofness is that at least somebody can benefit from insincere behaviour – which leads to the question of how much they can benefit, and how much others could be harmed by this alternative outcome. If and when an outcome diverts from the sincere outcome as a result of manipulation, what is the impact of this deviation: which group loses and which wins, by how much, and how does the total welfare differ from the sincere outcome?

One approach is to analysing the welfare impacts of manipulation is to look at the bounds of gains and losses form manipulation. Campbell & Kelly (2009) studied the maximal gains from manipulation in terms of number of ranks between alternatives in personal preference orderings. Under their framework, a choice rule f allows a gain of t when there exists an agent h and two distinct profiles R and R' which differ only in their hth element, and h ranks $f(R)$ t positions higher than $f(R)$. The maximal gain of a rule $G(f) = k$ is the maximum value t that f allows.

Clearly, $G(f)$ is bounded by $m-1$, and $G(f)=0$ would imply a dictatorship. Furthermore, there are no ways to guarantee small maximal gains other than giving some individual a lot of power in determining the outcome of the choice rule i.e., making a rule "more" dictatorial or "less" anonymous.

Similarly Campbell & Kelly (2010) also studied the bounds of loss from manipulation. A choice rule allows a loss of t when there are two agents h and j, two profiles R, R' that differ only in their hth element, h ranks $f(R')$ higher than $f(R)$, and $f(R')$ ranks t positions lower for j than $f(R)$. The loss of a rule can be defined as a *weak* or *strong* loss, where a weak loss $L_W(f)$ is the maximum loss t that a choice rule allows, and a strong loss $L_S(f)$ is the maximum loss that an optimal manipulation can cause. Whereas $G(f)=0$ implies a dictatorship, $L_W(f)=0$ does not necessitate a dictatorship. However, a rule that satisfies the condition of *universally beneficial manipulation*, which requires $L_W(f)=0$, on a full domain necessarily violates the Pareto principle.

Another similar approach is to calculate the proportion of "winners" and "losers" resulting from manipulation i.e., the maximum number of agents that either benefit or are harmed by manipulation. Campbell & Kelly (2014) study the maximum loss by defining their measure *the breadth of loss*, $Br(f)$, as the number of individuals who prefer the non-manipulated outcome to the manipulated one. Clearly the number of individuals who can be harmed ranges from $0 \leq Br(f) \leq n-1$. Campbell and Kelly decide to limit the number of agents by assuming that all agents are "essential", since adding "inessential" agents who never affect the outcome could alter the measure. In their results, they found that the breadth of loss in approval voting to be $Br(f)=n-1$, in the Borda rule with alphabetical tie-break $Br(f) \geq n-\lceil n/m \rceil$ - where $\lceil x \rceil$ is the ceiling function, and in Condorcet rules that satisfy the Pareto principle, when n is odd $Br(f) \geq (n-1)/2$, and when n is even $Br(f) \geq n/2$. However, the authors are not aware of any rules that would satisfy Pareto and Condorcet criteria, with a $Br(f)$ close to $n/2$, so these are just the theoretical bounds.

Demeze et al. (2016) likewise studied the proportions of those who benefit and those who are harmed by manipulation, but differing from Campbell & Kelly (2014), they treat voting as a game, where initial non-manipulators may response by manipulating as well. In their findings, under the plurality and Borda rules, manipulation can benefit $1/2$ to $2/3$ of voters and hurt $1/3$ to $1/2$ of voters. Under antiplurality (i.e., veto), manipulation can benefit $\frac{1}{3}$ to 100% of the population and hurt 0% to $2/3$ of the population. Therefore, interestingly antiplurality is the only rule of the three studied where a pareto improvement could be achieved via manipulation.

However, these measures do not give much information on *how* the population is affected via manipulation. Instead of focusing on the bounds of manipulation, an alternative approach is to construct a social welfare function in order to rank and compare sincere and manipulated outcomes from a social perspective.

Lu et al. (2012) use the score produced by a scoring rule as the social welfare function – the social welfare of alternative a for rule r with preferences R is determined via $SW(a, R)$. This method is argued to be a proxy for representing the utility values that different alternatives would obtain. In particular, the study is focused on the loss of social welfare that manipulation creates for the sincere voters, what is referred to as *regret*. In the paper, coalitional manipulators are assumed to have partial, probabilistic knowledge of other voters' preferences. Preference distributions are generated with (mixed) Mallows models utilizing real world electoral data and different values for ϕ. The empirical results of the study are limited to the Borda rule. The findings state that while the logical possibility of manipulation might be significant, in general the expected regret rate is fairly low – in terms of normalized percentages at most a 4% loss in social welfare. The authors of the study explain this by the notion that if manipulation were to succeed, the alternative generally speaking will be relatively high up in collective rankings, therefore not resulting in significant social loss.

Ianovski et al. (2022) conducted a study with a similar approach, just with a wider range of considerations – however the study is limited to only individual manipulation. The most well-known of the 15 voting rules studied are plurality and Borda rules, and the preference distributions studied include IC, spatial and (mixed) Mallows models as well as cultures based on empirical preference data. The three social welfare functions used are all based on the Borda score, first one being the sum of Borda scores, which approximates a utilitarian approach to measuring social welfare. The second measure is the Rawlsian welfare i.e., the Borda score of the voter who ranks the alternative the lowest, reflecting more egalitarian principles. Third measure is the Nash welfare function, determined by the product of the Borda scores, normalized by taking the nth root i.e., the geometric mean of the Borda scores – intended to balance egalitarian and utilitarian considerations.

The results of the paper by Ianovski et al. (2022) show that manipulation has an effect on the social welfare of the outcome, and that the existence of a manipulator will often change which rule is the most "optimal". In general, the study finds the impact of manipulation to be negative on social welfare.

A similar approach to using a social welfare function is to measure the so-called *Price of Anarchy* (PoA) for manipulation. The purpose of the measure is to compare the worst possible equilibrium

outcome to the optimal equilibrium outcome, but in the context of voting and manipulation a more reasonable measure is given by comparing the sincere outcome to the equilibrium winner resulted from strategic behaviour. This comparison is done by giving a score to these two outcomes and calculating the ratio between the two scores, thus giving the measures a similar structure to that of analysing social welfare functions. More specifically, this is referred to as the *dynamic price of anarchy* (DPoA). (Branzei et al., 2013)

Branzei et al. (2013) study three common *positional scoring rules*: plurality, veto (i.e., antiplurality), and Borda. Strategic voting is modelled as a process of *iterative voting*, where voters start from the sincere preference profile, and one by one get the chance to potentially alter their vote via their *best response strategy*. The category of Nash equilibria studied in the papers are ones to which such an iterative process of best responses would converge to. Most questionable methodological decision in the study is to use the score determined by the voting rule as a proxy for their quality – the score of an alternative is determined by the number of points it would receive under a given scoring rule with an assumed sincere preference profile.

The results of Branzei et al. (2013) show that under their framework, manipulated outcomes under plurality are the closest to sincere outcomes, meaning that the potential negative impacts of strategic behaviour for plurality voting are small – that DPoA is close to 1. Similarly, for veto with $m = 3$, DPoA is close to 1, but for $m \geq 4$ it is increasing with the number of alternatives. For Borda DPoA is linearly growing with the number of agents. Thus, by their estimates, strategic voting is the least harmful in plurality, more harmful in veto, and the worst under the Borda rule.

In addition to DPoA, Branzei et al. (2013) introduced the *additive dynamic price of anarchy* (ADPoA), which measures the difference – instead of the ratio – of the two outcomes. The ADPoA of plurality rule is deemed to be 1, meaning that the worst-case difference between the two scores compared is one. Kavner & Xia (2021) attempt to extend this measure to a different context: one where instead of looking at the voting rule scores of different outcomes, the outcomes are evaluated based on their social welfare. This is modelled by giving agents rank-based utility vectors over the alternatives. The findings of Kavner & Xia differ from Branzei et al., finding that with utility vectors differing from their respective scoring rule vectors, ADPoA of plurality is linearly increasing in the number of agents.

Following this result, Kavner & Xia (2021) study the *expected additive dynamic price of anarchy* (EADPoA) – where instead of merely focusing on the worst-case scenario of manipulation, they examine the average-case effect of manipulation. They find that under Impartial Culture

manipulation has an expected positive impact for the total welfare of a society with the plurality rule.

Clearly there is a challenge when it comes to extending the results of these studies focused iterative voting and best-response mechanics to more typically considered elections, where all voters vote at the same time, and only get to cast their vote once. One way to reconcile these two approaches is to consider the process of iterative voting and best responses to reflect something like opinion polls being broadcasted and voters updating their vote based on those results. Additionally, there exists voting mechanisms where people can immediately see the distribution of previous votes, and even alter their decision e.g., online platforms like Facebook and Doodle (Meir, 2017). However, there is no guarantee that strategic behaviour from some or all agents would necessarily lead to a Nash equilibrium.

One last method of measuring the welfare impacts of manipulation comes from a utilitarian framework, where the *utilitarian efficiency* and the difference in utilities between a sincere and a manipulated outcome of a voting rule are measured. Lehtinen (2007a, 2007b, 2008) studied the impacts of strategic voting with simulated voting games, where incomplete information of voters is modelled through a signal-extraction model and voters are assumed to behave in expected-utility maximization. Utilitarian efficiency is defined as the percentage of voting games in which the alternative with the highest sum of utilities i.e., the *utilitarian winner*, is selected by the rule (Lehtinen, 2007a).

Employing this approach, Lehtinen studied strategic voting in two parliamentary voting rules – so-called amendment and elimination agendas (Lehtinen, 2007b), the Borda rule (Lehtinen, 2007a) as well as approval and plurality voting (Lehtinen, 2008).

Lehtinen finds that strategic voting across multiple different voting rules and parameters increases utilitarian efficiency and yields a higher average utility than sincere voting behaviour. The extent of the differences varies depending on a variety of factors e.g., the information accessible to voters, but the general direction of strategic voting under Lehtinen's framework is that strategic voting has a positive impact on welfare. Lehtinen finds that generally speaking approval voting performs better than plurality voting. Additionally, the commonly perceived negative trait of high manipulability associated with the Borda rule seems to turn on its head, as strategic voting has higher utilitarian efficiency within the rule across different scenarios.

Lehtinen argues for the use of preference intensities; in fact, the assumption of preference intensities is what in his view allows for strategic voting to be beneficial. Strategic voting provides a method for voters to express their preference intensities, even when the voting rule in question does not explicitly ask for this information. Lehtinen argues that, under incomplete information, an alternative that is more intensively supported by voters receives more strategic votes than a less intensively supported alternative, which can lead into an outcome where a more intensively supported alternative defeats a less intensively supported alternative that, in terms of ordinal preferences, had a larger voter base.

However, questions of preference intensities require interpersonal comparison of utilities, which as expressed in chapter 2.2 of this thesis, is a practice not generally accepted in the field of social choice theory. Lehtinen acknowledges the epistemological impossibility associated with determining the exact utilities of individual agents or somehow comparing these values between agents. However, Lehtinen argues in favour of using preference intensities on the intuitive basis that different voters care more or less about different topics – and if preference intensities have a normative weight for individuals, should they not have such a weight for collectives as a whole as well?

Lehtinen generates different methods of interpersonal comparison of utilities. While utilitarian efficiency might slightly vary based on the selected method, the welfare effects of strategic voting are generally speaking positive, regardless of the selected method of interpersonal comparison. Thus, Lehtinen argues for the robustness of the interpersonal comparison of utilities.

Eggers & Nowacki (2024) primarily studied the susceptibility to manipulation via a model of expected utilities gained from manipulation under incomplete information. In addition to this, the study looked at the welfare impacts of strategic voting in plurality voting and IRV. In their findings, strategic voting tends to have a larger effect in the plurality rule compared to IRV in terms of change in the chosen alternative. When it comes to welfare effects of manipulation, strategic behaviour leads to slightly better outcomes under plurality, and slightly worse outcomes in IRV – but these impacts are quite small. This study challenges the somewhat common perception, where IRV is argued for by referring to its low manipulability as a benefit compared to widely used rules such as plurality. This can be the case, but if the outcomes of manipulation are generally negative under IRV, as opposed to generally positive under plurality, the argument becomes less compelling.

Some trade-offs have been associated with rules gaining a higher utilitarian efficiency. Lehtinen (2007a) under his framework showed that strategic voting, while increasing utilitarian efficiency,

unambiguously decreases the Condorcet efficiency of a rule. Green-Armytage et al. (2016) found a connection between increasing utilitarian efficiency and decreasing resistance to manipulation i.e., higher susceptibility to manipulation. These trade-offs are worth considering when evaluating which elements of a voting rule ought to be maximized or minimized.

There seems to be no widely accepted methodology for measuring the welfare impacts of strategic voting, and no general consensus as to whether these welfare impacts are positive or negative. Issues arise when determining the measure for social welfare – utilitarian considerations almost by their very nature necessitate incorporating cardinality and preference intensities, which brings about its own difficulties. Scoring rules, such as the Borda count, can be used as a proxy for utility based social welfare functions. Additionally, arguments can be made against the very idea of utility maximization as the basis of evaluating the value collective decisions, instead promoting alternative ideas such as Rawlsian egalitarianism.

4 Manipulation, Ethics and Democratic Ideals

This chapter will reflect on some of the broader implications of manipulation – in particular, the ethical considerations related to manipulation of voting rules, and the tensions between democratic ideals and strategic voting. Even though delineating clearly between squarely ethical and democratic questions might just be impossible, this distinction provides a useful way of differentiating between types of harm associated with manipulation.

The core question of this chapter is to examine why manipulation is so often considered both unethical and problematic for democracy. Arguments presented in the literature are covered. Additionally, some counter-arguments for these claims as well as arguments in favour of manipulation are explored. However, these are rarer in the study of manipulation, as the conventional view is that strategyproofness is desirable, and given its general impossibility within voting, minimal manipulation is sought after.

4.1 Ethics of Manipulation

The distinction between positive and normative economics is fundamental to the methodology of economics. Bluntly characterized, positive economics deals with questions on how things are, whereas normative economics with how things ought to be. Often in economics, normative considerations are characterized by evaluations of welfare outcomes, mainly because the core methods used in economics are quite suitable for evaluating welfare, rather than other concepts often deemed of moral worth e.g., freedom or justice. Therefore, normative economics is often referred to as welfare economics. (Hausman, 2024)

Social choice theory as a field of study – whether it is a branch of welfare economics or not – has normative elements in its core assumptions. Arrow's theorem is not considered impactful just because it describes a necessary connection between a set of axioms, but rather because these axioms are considered to be desirable i.e., reflective of what a *good* social choice rule *ought* to look like. Similarly, most studies in social choice theory are not concerned with empirical questions on how people in the real world make collective decisions, but rather on how to make *good* collective decisions. For example, the question of which voting rule is more Condorcet consistent is studied because it is deemed to be of social value.

It is worth noting that normativity does not necessarily have to be about ethics or morality. For example, rational choice theory can be viewed as a normative theory of rational behaviour: rather

than being concerned with how people actually act in the world – that is, how things are – it is concerned with how rational agents *ought* to act (Hands, 2012). Nevertheless, this subchapter is concerned with the ethically normative components of manipulation.

Broadly speaking the most prevalent subsets of moral theories are consequentialist (Sinnott-Armstrong, 2023) and deontological theories (Alexander & Moore, 2024). From both of these perspectives the morality of insincere voting can be evaluated. Within consequentialism, the consequences of actions determine their moral status: whether manipulation is acceptable or not depends on the outcomes resulted by manipulation. Conversely, most deontological theories can be contrasted by the fact that outcomes of actions are ignored. What makes something right or wrong is independent of its consequences; manipulation could be considered unethical due to some morally undesirable aspect inherent in it – or similarly morally justified regardless of its consequences within some other deontological theory. While both these concepts are broad and contain variety within themselves, they provide a framework for distinguishing between different methods of evaluating the morality of manipulation.

A straightforward consequentialist argument employed against manipulation is that it can lead to worse welfare outcomes. While a successful attempt at manipulation necessitates an improvement in welfare for at least one agent, the manipulator, there is no guarantee that this outcome will increase the aggregate social welfare. Additionally, since there is a kind of randomness associated with strategic behaviour, in particular within instances where all agents are not perfectly informed on others' preferences, there is no guarantee that an attempt at manipulation would work as intended (Satterthwaite, 1973). This makes it possible for manipulation to "backfire" and lead to a worse outcome, not just for the non-manipulators, but for the whole electorate: an information environment can be conceived where a majority of people would prefer a specific alternative but wouldn't consider this alternative to be popular, therefore strategically voting and electing another less preferred alternative (Conitzer & Walsh, 2016). Even Dowding & Hees, 2008 (2008), who argue in favour of manipulation, consider this randomness objection against manipulation to be among the stronger arguments against it.

Manipulation can equally well be justified from a consequentialist perspective, by showing that manipulation leads to better outcomes. In many instances of collective decision making, it seems clear that some people are more affected than others by which outcome gets selected; this can be used to argue that preference intensities should be accounted for within collective decision making. If Lehtinen (2007b, 2007a, 2008) is correct in that strategic voting allows voters to express

preference intensities within ordinal voting rules, insincere voting behaviour can be justified from this welfare-consequentialist perspective. In fact, this would lead to a situation where instead of strategyproofness being a desirable property, as it commonly is considered within social choice theory and in mechanism design more broadly, it would instead be normatively unacceptable within voting. Furthermore, Lehtinen (2011, 2015) argues against the normative weight of Arrow's theorem on these grounds – if IIA is justified primarily on the basis that it limits strategic voting, the positive welfare impacts of strategic voting would undermine the normative acceptability of this axiom.

However, even if the previously raised issues associated with interpersonal comparability were sidestepped, and it could be unequivocally stated that strategic voting generates better outcomes on the aggregate level, a problem remains as to who the groups "suffering" from manipulation are. In a world where strategic voting clearly benefits someone, this could mean that sincerity could be punished. This leads to quite uncomfortable conclusions, as voting mechanisms would in effect penalize honesty.

From a deontological perspective, dishonesty is considered the central problem associated with manipulation. Misrepresentation of preferences can be likened with lying, which under many deontological ethical theories is deemed as a moral wrong. So, under this account, manipulation is not wrong because of the outcomes it may or may not produce, but because lying in and of itself is wrong. This can be further motivated by referencing the idea of the "general will" or the "will of the people". When voters misrepresent their preferences, the true will of the people cannot be determined, and the purpose of an election is undermined (Lagerspetz, 2016).

Gschwend & Meffert (2017) argue against this conception of insincere voting as lying by noting that a voting scenario typically does not ask for the true preference ordering of voters. A typical election ballot does not instruct to vote for the alternative that a voter would want most to win, but rather the alternative that a voter wants to vote for. Therefore, equating a vote that does not match with a voter's preference to lying seems to lose much of its force.

Dowding & Hees (2008) approach this "insincerity" argument against manipulation by creating the seemingly oxymoronic concept of "sincere manipulation". Under sincere manipulation, an agent votes for the alternative that they prefer the most among the *feasible* outcomes. For example, in a voting situation under the plurality rule with three alternatives $A = \{x, y, z\}$, if x and y are neck and neck in popularity, while the popularity of z is extremely low, it seems unreasonable to expect all agents who rank z as their most preferred alternative to be morally obligated to vote for it. Dowding

& van Hees claim that it is not insincere to "abandon" a most preferred alternative for reasons of feasibility. Generalizing, the expected actions of other people can deem an individual's most preferred course of action unfeasible – adjusting behaviour based on this hardly seems insincere, but rather just an attempt of trying to best accomplish their objective given the circumstances. Therefore, Dowding & van Hees claim that the concept of insincerity is stretched too far without accounting for this feasibility constraint.

Wollesen (2025) argues that the normative grounds for why manipulation is wrong need to be developed in order to measure manipulability, and in order to evaluate the results of these measures. This argument relies on the idea of ethically "thick concepts", concepts that involve both a normative evaluation and a non-evaluative description (Väyrynen, 2021). Manipulability is a thick concept; it incorporates both a descriptive as well as a normative component. Wollesen claims that a measure for manipulation should reflect the values which guide why manipulation is bad. If a measure is supposed to provide normative guidance e.g., providing reason as to why a voting rule ought to be chosen over another, there needs to be clarity in what the normative concern is i.e., what it is that makes manipulation wrong. Wollesen suggests a "value pluralistic" approach, where different measures would reflect different normative judgements related to manipulation.

4.2 Democracy and Manipulation

The core results of social choice theory i.e., Arrow's and Gibbard-Satterthwaite's theorems, have been incorporated into the normative study of democracy. Perhaps the most influential attempt at this was done by a prominent political scientist William H. Riker. Riker's central work, *Liberalism against Populism*, covers the challenges that these negative results of social choice theory provide for democratic ideals (Riker, 1988).

One core question relates to the interpretation and existence of a general will, term first coined by the 18th century philosopher Jean-Jacques Rousseau (Bertram, 2024). The general will, or the will of the people, while somewhat ambiguous as a concept, typically refers to the idea of a single sensible objectively correct collective interest. The later developed concept of preference aggregation via social choice functions somewhat resembles this idea.

Riker (1988) argues against the populist interpretation of voting, which in his use of the term is largely based on the existence of the general will. Riker believes that there are two distinct interpretations of voting: the liberalist and the populist interpretation. Under the liberalist interpretation, voting is merely a method to control elected officials and nothing else. There is no

assumption that the electorate is or has to be correct. Conversely, under the populist interpretation voting is a way to discover and transfer the general will into elected officials. This clearly requires both the existence of a unique general will, and a way to access that general will via a fair collective preference aggregation method.

Riker's thesis is that the results of the seminal theorems of social choice theory undermine the vary basis of the populist interpretation of voting. As Arrow's theorem questions the very idea of a fair and logical preference aggregation method, and the Gibbard-Satterthwaite theorem shows that manipulability is effectively unavoidable, Riker argues that the very notion of the popular will is unclear. When a choice rule is manipulable it can never be known whether or not manipulation took place – Riker claims that through this "the truth and meaning of all outcomes is rendered dubious".

Riker's conclusions are very sceptical; even though the liberalist interpretation of voting is not rejected outright such as the populist one, as the liberalist view of voting can be achieved, still the remains of such a liberal democracy is built on seemingly arbitrary and uninterpretable voting procedures. Riker's arguments have received a lot of criticism, for example related to his strict distinction between liberalism and populism, somewhat imprecise definition of populism, and overtly sceptical practical conclusions based on theoretical assumptions (Lagerspetz, 2016).

An important and increasingly influential strand of democratic theory in recent decades is deliberative democracy. Within deliberative democracy, the goal is not to merely aggregate preferences or to implement the general will, but to reach consensus through communication and reflection of values and interests regarding issues of collective importance. Good deliberation is based on values of reciprocity, respect, and equality. These deliberative ideals are not meant to be understood as necessarily practically reachable, but something that deliberative democratic institutions should strive towards. Deliberative democracy can be contrasted with aggregative democracy, in which counting and aggregating votes is central (Bächtiger et al., 2018).

Even though social choice theory and deliberative democracy exemplify substantially different approaches towards analysing democratic decision-making, some proponents of deliberative democracy have argued that it could remedy the core problems of social choice theory. Dryzek & List (2003) argued that group deliberation could incentivize sincere preference revelation, for example by adding "costs" to insincere behaviour. This cost is based on the idea that especially in recurrent interactions, being exposed as a "liar" would be socially harmful. Were this true, the threat of strategic manipulation would decrease under deliberative processes. Similarly, Dryzek and List argued that deliberation could narrow the domain of actual preference profiles, thus relaxing the

universal domain assumptions and providing “escape-routes” both from Arrow’s and Gibbard-Satterthwaite’s theorems.

Nurmi (2023) has argued against the views of (Dryzek & List, 2003) and others who have suggested that these central negative results of social choice theory could be avoided via deliberative democratic institutions. Nurmi questions the quite idealized assumptions associated with the deliberative processes as well as the relationship between amount of information accessible and manipulability. If deliberation induces sincere preference revelation, this might provide opportunities for manipulation that would not be accessible under more restricted information environments, creating exactly the opposite effect that an advocate for deliberative democracy would hope for.

Within the social choice theory literature, a variety of problems on how manipulation impedes the fulfilment of different democratic ideals have been presented. One of these problems pertains to the difference between voters’ ability to strategize, highlighted first by Satterthwaite (1973). Different voters might have access to more complete or reliable information, or the resources to make better informed decisions, leading to disenfranchisement of less resourceful or capable voters (Conitzer & Walsh, 2016). Eggers & Vivyan (2020) find that typically richer and older voters are at a greater advantage when it comes to strategic voting.

This problem relates to the democratic ideal of “one person one vote” – manipulation allowing voting to become a game of skill can undermine the idea of equal power between voters (Wollesen, 2025). Typically, winning elections due to greater strategic aptitude is not deemed desirable for a democracy. Furthermore, within democratic parties and coalitions, larger, more established organizations are better equipped to compose more complex strategies, thus further undermining democratic equality (Lagerspetz, 2016).

A potential counterargument comes from the existence of preference intensities. If strategic voting provides an avenue for expressing these, would it not make sense to allow more invested voters to wield more power by becoming better informed. Additionally, Dowding & Hees (2008) argue that the possibility for greater benefit via manipulation can aid democracy, by providing an incentive for individuals to learn more about the functioning of their electoral systems and the preferences or voting habits of other citizens. Though this seems to just push back the problem one step, as those with more time and resources to learn are advantaged.

Another democratic problem commonly associated with manipulation is related to the lack of transparency. This could be characterized as the lack of transparency related to the preferences of voters or the preferences of elected representatives. With elected representatives in particular, the problem is that there is no guarantee that their casted vote would match their true opinion – making it difficult for voters to assess their representative (Lagerspetz, 2016; Satterthwaite, 1973). More generally, a problem with lack of transparency is the muddying of interpreting election results, as the ”true” popularity of alternatives remains unknown one strategic voting takes place (Conitzer & Walsh, 2016).

Dowding & Hees (2008) argue against the class of transparency arguments, by claiming that this just further encourages people to participate with the democracy around them in order to better find out the true opinions of other voters. When it comes to the non-transparency of representatives' preferences, Dowding & van Hees deem the true preferences of representatives irrelevant for voters; arguing that what should matter to voters is the actual voting record of their representative, not what their true underlying preferences might look like.

A third, more general problem related to manipulability is the potential inefficiency and “wasted” effort associated with it. Satterthwaite (1973) noted that manipulability can create an incentive for committee members to not reveal their true preferences, as others could utilize this in their strategizing – leading to an inefficient flow of information. Similarly, Conitzer & Walsh (2016) highlight the amount of effort required in order to successfully manipulate. This can include gathering information, communication, and computational resources – and it can remain unclear whether or not collectively speaking these costs associated with manipulation outweigh the benefits. This inefficiency resembles the tragedy of the commons: while manipulation may be individually rational, everyone collectively bearing the cost of strategizing could lead to a net loss.

5 Role of Manipulability in the U.S. RCV Debate

Discussions of election reform have been topical in the United States of America throughout the early 21st century. One polarizing issue has been ranked-choice voting (RCV). RCV could be defined as an umbrella term for all voting methods where voters rank candidates in order of their (reported) preferences, but in public discussions it is in single-winner elections most often used to refer to instant-runoff voting (IRV).

Advocates of RCV over plurality often cite many different arguments for adopting it as a new voting method. One of these arguments is that RCV is less susceptible to strategic voting i.e., less manipulable. However, as discussed in the previous chapter the property of being less manipulable is often deemed to be desirable without much explicit substantiation. The aim of this chapter is to provide a brief overview on the status of RCV in the U.S. in 2026, assessing some of the arguments for and against RCV. Primary focus will be on evaluating whether lower manipulability as such is something that can *prima facie* serve as an argument for adopting a voting rule, given the findings of the literature review performed in chapters 3 and 4 of this thesis.

This chapter does not seek to argue against ranked-choice voting, nor defend plurality voting or other alternative voting electoral systems. Rather its purpose is to evaluate how arguments grounded in social choice theory can be utilized within contemporary discussions of election reform.

5.1 Ranked Choice Voting in the United States

RCV has gained growing attention in public discourse and has been implemented in some U.S. jurisdictions in recent years. While technically the history of ranked-choice voting in the U.S. dates back to early 20th century with the adoption of STV in some multi-winner elections for proportional representation, instant-runoff voting is a more contemporary topic, with the earliest adoption of it transpiring in the early 21st century.

In the United States, as of February 2026, RCV has been used statewide in Alaska and Maine for federal elections, such as the presidential election, since 2022 and 2016 respectively. In total, 48 jurisdictions in the U.S. use ranked choice voting in public elections. Additionally, six U.S. states use RCV for military and overseas voters for their elections. (FairVote, 2026)

RCV has also faced significant opposition. According to (Ballotpedia, 2026), as of February 2026 the use of RCV has been restricted or banned in 18 individual states in the U.S., first in February of 2022 by the state of Tennessee (Tenn. SB 1820, 2022). In April 2025, Donald Trump, the president

of the United States stated that RCV is “one of the greatest threats to democracy” (Trump, 2025). In January of 2026, the “Make Elections Great Again” Act was proposed by the U.S. House Committee on House Administration chair Bryan Steil, which would in addition to other election reforms, prohibit the use of RCV in all federal elections (Make Elections Great Again Act, 2026). The recency of these bans, comments, and initiatives highlights the topical and polarizing status of RCV in political discussions in the U.S.

When arguing for or against implementing RCV, it is most often pitted against plurality voting, as it is the most common method of voting used in the United States. RCV is often advocated for promoting a healthier and less polarizing campaigning environment, promoting diversity, allowing more voter choice, and perhaps most importantly, for choosing a more popular winner i.e., better representing the “will of the people”. Arguments against RCV are often concerned with its increased complexity – related either to administration or voter understanding, or notions that it favours one political party over the other.

Addressing and evaluating the aforementioned arguments falls outside the scope of this chapter and thesis. The relevant argument is that RCV reduces incentives and opportunities for strategic voting, allowing for voters to express their preferences more sincerely. This has been invoked by proponents of RCV ranging from academics to political advocacy groups as a reason to prefer RCV over plurality voting. However, given the complexities related to measuring the prevalence and welfare impacts of manipulation, as well as the contestability of the ethics of insincere voting, it is worth examining to what extent the concept of lower manipulability – along with other concepts of social choice theory – can and should be meaningfully applied to arguments of election reform.

Real-world discussions rarely use the formal language of social choice theory. Rather than referring to specific measures of manipulability, terms like “strategic voting” and “honesty” are used by political advocacy groups. For example, *FairVote* (FairVote, 2026) states that “our ‘choose-one’ elections incentivize voters to think strategically, rather than honestly” and “RCV incentivizes sincere voting while other methods create strong opportunities for strategic exploitation”. In such arguments, the underlying ethical or democratic harm of manipulation is often left implicit.

Some academic arguments make this connection more explicitly. For instance, Dasgupta & Maskin (2019) suggest that strategizing places a burden on the voter and that strategic voting behaviour might lead to unpredictable outcomes, offering a clearer account on what would make strategyproofness or lower manipulability desirable within the context of voting. Yet even in academic work, the negative character of strategic behaviour is still frequently implicitly assumed

rather than explicitly explained. Since terms such as ‘manipulation’, ‘strategizing’ and ‘insincerity’ are normatively charged, their use can provide a rhetorical force to an argument even without a clear account of what makes such behaviour undesirable within voting.

5.2 2022 Alaska Special Election for US House

Manipulability is certainly not the only link between theoretical findings of social choice theory and real-world elections – and advocating for or against election reform on the basis of these findings. The concept of a Condorcet winner (Definition 2.3.13.) and the property of monotonicity (Definition 2.3.14.) are related to the discussions concerning RCV. The 2022 Alaskan Special Election for US House provides an interesting case study to examine this link between theory and practice.

As noted earlier in this thesis, RCV can notoriously fail monotonicity. Failures (or paradoxes) of monotonicity can be distinguished upward monotonicity failures and downward monotonicity failures. In an upward monotonicity failure, an initially winning candidate could end up losing the election by being moved up in some ballots i.e., by receiving more first place votes. Conversely, in downward monotonicity paradox, a losing candidate can become the winning candidate by being moved down in some rankings. For clarity, it is not the case that some election results are non-monotonic, but rather that RCV is non-monotonic, and that some preference profiles might be vulnerable to failures of monotonicity.

In the 2022 Alaskan election, there were three candidates: Nick Begich, Sarah Palin, and Mary Peltola. The submitted ballots are reflected in the following table (Graham-Squire & McCune, 2022):

Number of voters	27053	15467	11290	34049	3652	21272	47407	4645	23747
1st choice	Begich	Begich	Begich	Palin	Palin	Palin	Peltola	Peltola	Peltola
2nd choice	Palin	Peltola		Begich	Peltola		Begich	Palin	
3rd choice	Peltola	Palin		Peltola	Begich		Palin	Begich	

Table 5.1 Submitted ballots in the 2022 Alaska Special Election

Peltola received the most first-place votes, tallying 75 799 (40.2%) of the vote. Second came Palin (58 973; 31.3 %), and the least first-place votes received Begich (53 810; 28.5%) who then by the mechanism of RCV got eliminated from the election, as no candidate had a majority (> 50%) of the vote. Voters who ranked Begich in the first place had their votes transferred to their second choice, with Peltola “receiving” 15 467 votes, and Palin 27 053 votes respectively. The 11 290 voters who

only ranked Begich had their votes "exhausted" i.e., removed from consideration in the election. In the second round Peltola (91 266; 51.5%) beat Palin (86 026; 48.5%).

This election is interesting as it highlights how RCV violates both the monotonicity principle and the Condorcet winner criteria. Adjusting the submitted ballots in a way where six thousand voters who only ranked Palin first had instead submitted the ballot Peltola ≻ Palin ≻ Begich, the election result would have looked like the following:

Number of voters	27053	15467	11290	34049	3652	**15272**	47407	**10645**	23747
1st choice	Begich	Begich	Begich	Palin	Palin	Palin	Peltola	Peltola	Peltola
2nd choice	Palin	Peltola		Begich	Peltola		Begich	Palin	
3rd choice	Peltola	Palin		Peltola	Begich		Palin	Begich	

Table 5.2 Alternative ballots in the 2022 Alaska Special Election

Here the first-place votes would have been as follows: Begich 53 810 (28.5%), Palin 52 973 (28.1%) and Peltola 81 799 (43.4%). Instead of Begich being eliminated in the first round, Palin gets eliminated. From the first-place Palin voters, 34 049 votes get transferred to Begich, 3 652 votes to Peltola, and 15 272 votes get exhausted. In the final tally, Begich wins this election with 87 859 (50.7%) votes – causing Peltola to lose the election despite her first-place votes increasing within the voting population – and thus violating monotonicity.

Returning to the Table 5.1, interestingly, the outcome would have been different had the elimination method been changed from the one receiving *least first place* votes being eliminated, to the one receiving *most last place* votes being eliminated – i.e., changing from IRV to Coombs' method (Definition 3.1.5.). Begich received the smallest amount of last place votes by a wide margin, 8 297 – compared to Palin's 62 874 and Peltola's 61 102. Therefore, with the Coombs' method of elimination, Palin gets eliminated in the first round, and Begich defeats Peltola in the second round.

The monotonicity paradox observed in the Alaskan election is a rare occurrence. There are only a handful of other well-known instances of such an event transpiring in U.S. election history – a 2009 Burlington mayoral election and a 2021 Minneapolis city council election being the most studied (McCune & McCune, 2021). Across 2000 RCV elections in the U.S., the upward monotonicity paradox has been observed in around 1.0% of them (Institute for Mathematics and Democracy, 2025).

Despite the rarity of the paradox, these individual instances can have major consequences in real-world election reforms – for instance, the city of Burlington abandoned the use of RCV after the

2009 election (Graham-Squire & McCune, 2023). Similarly, following the 2022 election in Alaska a ballot initiative was voted on in 2024, on whether or not RCV would be repealed. The ballot did not pass i.e., RCV was not repealed, but the margin was very narrow – 50.11% voting "No" and 49.89% "Yes" (Alaska Division of Elections, 2024a, 2024b).

Advocates of RCV often see the property of non-monotonicity as a mere technical or hypothetical concern, something not worth being concerned about in the real world. *FairVote* (2022), for instance claims that non-monotonicity "is irrelevant in practice, as it does not result in a 'wrong' candidate winning, and it will not affect voter strategy".

More theoretical approaches have also been taken to establish the frequency of monotonicity failures under RCV. Lepelley et al. (1996) find that under IAC upward monotonicity failures exist in 4.51% of situations and downward monotonicity failures in 1.97%, whereas Miller (2012) finds the prevalence of upward monotonicity failures to be 12.0% and of downward monotonicity failure 4.8% under IC. Ornstein & Norman (2014) use a spatial model of voter behaviour, finding that the frequency of upward monotonicity failures increases as the competitiveness of the election increases. They determine a lower bound of 15% for competitive elections with three candidates, with an upper bound of 51% - finding most of the simulated monotonicity failures to have occurred due to Condorcet inefficiencies in competitive elections. Similar results were later confirmed by Miller (2017).

Restricting the study of monotonicity failure to competitive elections is sensible, as for clear-cut elections most reasonable voting rules should produce the same result. In the context of a competitive election, like the 2022 Alaska special election, monotonicity violations may serve as a meaningful objection for advocates of RCV – or, at the very least as a powerful rhetorical tool.

This ties back into the question as to what extent should theoretical results, such as monotonicity violations or incentives for strategic voting, weigh in the evaluations of election methods? If low probability of occurrence weakens concerns of monotonicity failures, should it not equally weaken concerns about strategic voting under other voting rules?

In principle, the preference profiles vulnerable to monotonicity failures under RCV can introduce a type of strategic reasoning that differs from those possible under plurality voting. While strategic voting under plurality is typically concerned with not "wasting" a vote or voting for a perceived lesser of two evils, under RCV potential strategic incentives would be less transparent.

While the cases of monotonicity violations are uncommon, the theoretical possibility of non-monotonicity introduces a layer of strategic uncertainty. Voters may be unsure whether ranking their sincere first choice could paradoxically contribute to its defeat. In addition to this, *in theory*, a strategizing coalition could "sabotage" a candidate's success in a RCV election by increasing their support – by exploiting a perceived potential failure of monotonicity. This, however, seems like only a theoretical possibility, rather than a real worry – given how large the coalition would have to be, how accurate their information of others' behaviour would have to be, and how this strategizing could likely "backfire". Still, even if in practice incredibly unlikely, this kind of strategic incentive seems intuitively more ethically and democratically questionable than the "lesser evil" voting strategies of plurality voting.

A further interesting aspect of the Alaskan case for the purposes of this thesis is to question whether strategizing would have been morally wrong or democratically problematic. Assuming that the ballots presented in Table 5.1 reflect the sincere preferences of the voting population, then Peltola was not the Condorcet winner of the election – but Begich instead. In head-to-head matchups, Begich would have defeated both Palin and Peltola, as being preferred by more people over the other individual candidates.

This ties back into the ambiguous normative status of strategic voting. Suppose around 2 600 voters with the preferences Palin $\succ$ Begich $\succ$ Peltola inferred that Begich had a better chance to win against Peltola in the second round. By insincerely ranking Begich first (Begich $\succ$ Palin $\succ$ Peltola), they could have changed the outcome from Peltola to Begich, thereby achieving a more favourable outcome for themselves i.e., a successful coalitional manipulation. In addition to the benefit for these individuals, the outcome would also be socially preferable from a Condorcet-efficiency perspective. This raises familiar questions: on what basis would such strategizing be morally problematic, and how should its democratic implications be evaluated?

Of course, this entire discussion relies on an assumption that is impossible to verify – that the submitted ballots were reflective of the sincere preferences of the electorate. Most likely it is the case that some voters were not sincere: hypothetically, Begich's status as an apparent Condorcet winner may itself be the result of strategic behaviour. This highlights the epistemic challenge of observing true preferences, which complicates any normative assessment of strategic voting in real elections.

5.3 Lower manipulability as an argument for election reform

When RCV is said to be less susceptible to strategic voting, what this usually means is that a voter does not have to vote in a particular strategic way e.g., for the lesser of two evils. But as the theoretical literature demonstrates, and as the Alaskan election illustrates, RCV – much like any other voting rule – does not eliminate the possibility for strategic voting behaviour. In light of the literature covered in chapters 3 and 4 of this thesis, the legitimacy of using theoretical findings related to voting rule manipulability as an argument for a voting rule ought to be questioned.

To begin with, there are multiple methods of measuring manipulability, and even within specific numerical measures of manipulability, such as the Nitzan-Kelly Index, there are a wide variety of background assumptions that can alter the "ranking" of two different voting rules. These assumptions include preference probability distribution i.e. the statistical culture, agent behaviour e.g., the rate and extent of strategic behaviour, knowledge of other agents' behaviour, number of voters and candidates. As Saari (1990) noted "with the appropriate assumptions, with a correctly constructed scenario, any system can be justified as being strategically the best".

If the object of interest for a measure of manipulability is merely the logical possibility of manipulation, then unrealistic assumptions such as preference distributions like Impartial Culture pose no issue. But for a measure of manipulability to have predictive power, the parameters of the model should more closely reflect the voting population and their behaviour.

Furthermore, there is no clear consensus on the welfare impacts of strategic voting, as different results have been found depending on the methodologies employed. Directly related to the debate between plurality voting and RCV, Eggers & Nowacki (2024) found that strategic voting generates slightly better welfare outcomes under plurality, and slightly worse outcomes under RCV. Strategic incentives depend on the expectation of others' behaviour: under IRV they are stronger when others are expected to vote sincerely, whereas under plurality they are stronger when others are expected to vote strategically. Overall, they argue that IRV is less susceptible to strategic voting because the expected gains from strategizing fall as strategic behaviour becomes more widespread – generating a contrast between the "bandwagon effect", where under plurality voting a preferred candidate getting abandoned by others incentivizes strategizing. Kavner & Xia (2021), in their framework for studying welfare impacts, claim that manipulation under the plurality rule has an expected positive impact on the total welfare of a society.

From a welfare-consequentialist perspective, manipulation is not intrinsically problematic – its ethical status depends on its effects on aggregate welfare. If strategic voting behaviour increases welfare, it is not objectionable on these grounds. However, the literature on welfare effects does not provide clear support for making lower manipulability a decisive reform criterion. Measures of manipulability and welfare vary, and the welfare effects of strategic behaviour are ambiguous and context dependent.

At the same time, even if the outcomes might sometimes be better after strategizing, there are other welfare-consequentialist reasons to favour lower manipulability. Because voters in the real-world operate under incomplete information, the aggregate effects of strategic voting are difficult to predict. Even if manipulation can increase welfare in some settings, it may also distort the "intended" inputs of the voting rule, leading to potentially unpredictable outcomes, which can be worse for a voting population's welfare. From this perspective, reducing susceptibility to manipulation can be defended as a way of limiting systemic risk, even if full strategyproofness remains unattainable within voting.

But as expressed in chapter 4, a welfare-consequentialist view is not the only normative concern associated with manipulability. A strong objection against strategic voting stems from the challenge that incentives to strategize pose for democratic ideals. When voting turns into a strategic game, those with greater informational or organizational resources may be better equipped to exploit opportunities for manipulation, in a way undermining the egalitarian principle of "one person one vote".

The divergence in normative concerns is reflected in different metrics of manipulability. Measures such as the Nitzan-Kelly index (NKI) focus on the logical possibility of manipulations and their normative force is typically grounded in welfare-consequentialist concerns – if manipulation is assumed to be harmful more often than not, voting rules that have fewer manipulable profiles are preferable. By contrast computational complexity reflects the egalitarian concerns more – as finding a successful manipulation becomes increasingly difficult, the access to manipulation gets restricted to a smaller portion of the voter base with more resources to find these manipulations. Somewhat unexpectedly, from an egalitarian perspective an argument can thus be made that a lower difficulty of finding successful manipulations is preferable. (Wollesen, 2025)

Following from this, there is a challenge in that each metric embeds particular normative commitments. Wollesen (2025) argues for a plurality of metrics; rather than searching for a single

decisive measure, different metrics could track different normative concerns. On this view, lower manipulability is not inherently good, but dependent on the metric used.

In the context of the U.S. debate over electoral reform, concepts related to manipulability surface when evaluating RCV against plurality voting. While the structure of RCV eliminates certain forms of strategic voting, it does not get rid of strategic incentives completely. RCV's more complex ballots and elimination procedures can make these strategic incentives less transparent and more difficult for voters to identify or act upon, for example in the case of monotonicity failures. These counterintuitive incentives for strategizing illustrate a limitation of arguments that present RCV as a method reducing the need for strategic thinking. This highlights that concerns about manipulation are not merely about frequency, but also about who has access to exploiting these strategic opportunities, how predictable the outcomes are, and how strategic behaviour intersects with welfare considerations and democratic ideals.

By contrast, strategic incentives under plurality voting are generally more straightforward and easier for voters to detect, making manipulation more transparent. While this transparency may reduce complexity, the ethical significance of such strategic behaviour still depends on which normative concerns are prioritised.

The literature covered, arguments highlighted, and practical case from Alaska show that technical properties of voting rules matter to an extent for normative evaluation, and perhaps even more so public perception, but cannot alone determine which voting method is most preferable to adopt in the real-world. Susceptibility to strategic voting can be used as a powerful rhetorical tool, but its influence relies at times on an ambiguous and inconclusive analytical and ethical basis.

Again, it is worth emphasizing that although often invoked in discussions related to RCV, properties of manipulability or monotonicity are not in general the central focus of voting system reform discussions. There are a multitude of other arguments for or against adopting or comparing voting rules. Some may be grounded in social choice theory, and others more social e.g., RCV incentivizes candidates to get along better with each other (Compton, 2025), as well as practical arguments related to implementing a new voting system e.g., some ballots being more complicated to fill, thus potentially confusing to voters. The conclusion of this chapter is not that the pursuit of RCV reform ought to be abandoned, or that there are no good arguments in support of it.

Instead, the conclusion of this chapter aligns more so with the view expressed by Wollesen (2025), namely, that for a measure of manipulability to have meaningful weight in real-world discussions

related to election reform, there needs to be clarity on what – if anything – makes manipulation wrong, something that ought to be minimized and avoided. This holds for appeals to manipulability in the reform advocating for any voting rule, not just RCV. Appealing to lower manipulability as a reason for favouring a voting rule should not collapse to the claim, “It is less susceptible to strategic voting, therefore, it is better”. Instead, first the particular moral wrong that strategic voting is thought to involve should be identified – and then shown how this particular voting rule is less susceptible to that specific concern.

.

6 Conclusions

This thesis provided a literature review on metrics of manipulability, as well as the welfare impacts and ethical and democratic implications of strategic voting, with the aim of evaluating whether this body of work provides sufficient grounds for using manipulability as a compelling argument in electoral reform debates. These findings were applied to the real-world case of ranked-choice voting (RCV) advocacy in the United States.

The main findings of the literature review reflect a broad inconclusiveness in the topics covered. There is no clear consensus on how the manipulability of a voting rule should be measured, and many of the findings are dependent on a variety of assumptions. Likewise, the welfare impacts of manipulability, while less studied, are still framework dependent and often ambiguous. The normative status of strategic voting remains contested as well. While there are some compelling arguments towards limiting strategic incentives for manipulation, such as the potential unpredictability of manipulation, a case can equally well be made for the acceptance or even reverence of strategic voting. This ambiguity is reflected in the metrics of manipulations themselves – different normative concerns can justify the use of entirely different metrics of manipulation. Egalitarian concerns may point to the use of measures based on computational complexity, while welfare-consequentialist concerns may point towards frequency-based measures such as the Nitzan-Kelly index. Yet even within the welfare-consequentialist framework, there is no agreement on whether strategic voting is harmful in the first place, further undermining the normative grounding of any single measure.

Therefore, the conclusion of this thesis is that the literature does not currently provide sufficient grounds for straightforward use of concepts related to voting rule manipulability within debates of election reform. It should, however, be noted that this does not deem studying manipulability frivolous. Rather this conclusion highlights that these theoretical measures should be readjusted to more explicitly acknowledge their normative assumptions, if they are intended to have greater argumentative weight in discussions of real-world election reform.

The conclusions drawn here are naturally bounded by the scope of the literature reviewed. The thesis primarily focused single-winner voting rules, leaving aside multi-winner rules as well as the broader range of collective decision-making procedures where questions of strategic incentives can arise differently. The literature reviewed is mostly theoretical, and the actual strategic behaviour of voters in actual elections was less explored – which is worth exploring when examining the

relationship between applying theoretical measures to practice. Additionally, the thesis focused specifically on strategic behaviour of voters, rather than other strategic phenomena such as strategic nomination of candidates. For future research, the most pressing question is related to the normative grounding itself: clarifying what, if anything, makes strategic voting undesirable, and what metrics best reflect these specific concerns. Similarly to the already developed measures of manipulability, new approaches to measuring manipulability, such as the machine learnability of manipulation, should be considered with this normative grounding in mind – especially if their results are intended to meaningfully guide real-world election reform.

References

Alaska Division of Elections. (2024a). *Ballot Measure No. 2: An Act Restoring Political Party Primaries and Single-Choice General Elections (22AKHE)*. https://www.elections.alaska.gov/doc/oep/2024/Ballot%20Measure%202_Eng.pdf

Alaska Division of Elections. (2024b). *State of Alaska 2024 General Election: Election Summary Report*. https://www.elections.alaska.gov/results/24GENR/ElectionSummaryReport.pdf

Aleskerov, F., Ivanov, A., Karabekyan, D., & Yakuba, V. (2015). Manipulability of aggregation procedures in impartial anonymous culture. *Procedia Computer Science*, *55*, 1250–1257. https://doi.org/10.1016/j.procs.2015.07.133

Aleskerov, F., Karabekyan, D., Sanver, M. R., & Yakuba, V. (2011). An individual manipulability of positional voting rules. *SERIEs*, *2*(4), 431–446. https://doi.org/10.1007/s13209-011-0050-y

Aleskerov, F., Karabekyan, D., Sanver, M. R., & Yakuba, V. (2012). On the manipulability of voting rules: The case of 4 and 5 alternatives. *Mathematical Social Sciences*, *64*(1), Article 1.

Alexander, L., & Moore, M. (2024). Deontological ethics. In E. N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Winter 2024). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2024/entries/ethics-deontological/

An Act to Amend Tennessee Code Annotated, Title 2, Relative to Instant Runoff Voting, SB 1820 (Public Chapter 621), Tennessee General Assembly, 112th General Assembly (2022). https://publications.tnsosfiles.com/acts/112/pub/pc0621.pdf

Arrow, K. J. (2012). *Social Choice and Individual Values*. Yale University Press. https://www.jstor.org/stable/j.ctt1nqb90 (Original work published 1951)

Bachrach, Y., Elkind, E., & Faliszewski, P. (2011). Coalitional Voting Manipulation: A Game-Theoretic Perspective. *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence*. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-020

Bächtiger, A., Dryzek, J. S., Mansbridge, J., & Warren, M. E. (2018). *The Oxford Handbook of Deliberative Democracy*. Oxford University Press.

Ballotpedia. (2026). *Ranked-choice voting (RCV)*. Ballotpedia. https://ballotpedia.org/Ranked-choice_voting_(RCV)

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

Bartholdi, J. J., Tovey, C. A., & Trick, M. A. (1989). The computational difficulty of manipulating an election. *Social Choice and Welfare*, *6*(3), 227–241. https://doi.org/10.1007/BF00295861

Bertram, C. (2024). Jean Jacques Rousseau. In E. N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Summer 2024). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2024/entries/rousseau/

Black, D. (1948). On the Rationale of Group Decision-making. *Journal of Political Economy*, *56*(1), 23–34. https://doi.org/10.1086/256633

Boatright, R. G., Tolbert, C. J., & Micatka, N. K. (2024). Public opinion on reforming U.S. primaries. *Social Science Quarterly*, *105*(3), 876–893. https://doi.org/10.1111/ssqu.13370

Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (2016). *Handbook of computational social choice*. Cambridge University Press.

Branzei, S., Caragiannis, I., Morgenstern, J., & Procaccia, A. (2013). How Bad Is Selfish Voting? *Proceedings of the AAAI Conference on Artificial Intelligence*, *27*(1), Article 1. https://doi.org/10.1609/aaai.v27i1.8667

Campbell, D. E., & Kelly, J. S. (2009). Gains from manipulating social choice rules. *Economic Theory*, *40*(3), 349–371. https://doi.org/10.1007/s00199-008-0380-6

Campbell, D. E., & Kelly, J. S. (2010). Losses due to manipulation of social choice rules. *Economic Theory*, *45*(3), 453–467.

Campbell, D. E., & Kelly, J. S. (2014). Breadth of loss due to manipulation. *Economic Theory*, *55*(2), 393–414. https://doi.org/10.1007/s00199-013-0752-4

Carroll, G. (2011). *A Quantitative Approach to Incentives: Application to Voting Rules* [Unpublished manuscript]. Massachusetts Institute of Technology.

Chamberlin, J. R. (1985). An investigation into the relative manipulability of four voting systems. *Behavioral Science*, *30*(4), 195–203. https://doi.org/10.1002/bs.3830300404

Compton, K. (2025). Depolarizing America with Ranked-Choice Voting. *University of Pittsburgh Law Review*, *86*(4). https://doi.org/10.5195/lawreview.2025.1096

Conitzer, V., Sandholm, T., & Lang, J. (2007). When are elections with few candidates hard to manipulate? *Journal of the ACM*, *54*(3), 14. https://doi.org/10.1145/1236457.1236461

Conitzer, V., & Walsh, T. (2016). Barriers to Manipulation in Voting. In H. Moulin, *Handbook of Computational Social Choice* (1st edn, pp. 127–145). Cambridge University Press. https://doi.org/10.1017/CBO9781107446984.007

Dasgupta, P., & Maskin, E. (2019). *Elections and Strategic Voting: Condorcet and Borda* [Working Paper]. Toulouse School of Economics.

Demeze, H., Moyouwou, I., & Pongou, R. (2016). *The Welfare Economics of Tactical Voting in Democracies: A Partial Identification Equilibrium Analysis* (MPRA Paper No. 70607). University Library of Munich. https://uni-muenchen.de

Dowding, K., & Hees, M. V. (2008). In Praise of Manipulation. *British Journal of Political Science*, *38*(1), 1–15. https://doi.org/10.1017/S000712340800001X

Dryzek, J. S., & List, C. (2003). Social Choice Theory and Deliberative Democracy: A Reconciliation. *British Journal of Political Science*, *33*(01), Article 01. https://doi.org/10.1017/S0007123403000012

Eğecioğlu, Ö., & Giritligil, A. E. (2013). The Impartial, Anonymous, and Neutral Culture Model: A Probability Model for Sampling Public Preference Structures. *The Journal of Mathematical Sociology*, *37*(4), 203–222. https://doi.org/10.1080/0022250X.2011.597012

Eggers, A. C., & Nowacki, T. (2024). Susceptibility to Strategic Voting: A Comparison of Plurality and Instant-Runoff Elections. *The Journal of Politics*, *86*(2), 521–534. https://doi.org/10.1086/726943

Eggers, A. C., & Vivyan, N. (2020). Who Votes More Strategically? *American Political Science Review*, *114*(2), 470–485. https://doi.org/10.1017/S0003055419000820

FairVote. (2022). *Alternatives to RCV*. FairVote. https://fairvote.org/archives/alternatives-to-rcv/

FairVote. (2026). *Ranked Choice Voting Information*. FairVote. https://fairvote.org/our-reforms/ranked-choice-voting-information/

Favardin, P., & Lepelley, D. (2006). Some Further Results on the Manipulability of Social Choice Rules. *Social Choice and Welfare*, *26*(3), Article 3. https://doi.org/10.1007/s00355-006-0106-2

Favardin, P., Lepelley, D., & Serais, J. (2002). Borda rule, Copeland method and strategic manipulation. *Review of Economic Design*, *7*(2), 213–228. https://doi.org/10.1007/s100580200073

Feldman, A. M., & Serrano, R. (Eds). (2006). Introduction. In *Welfare Economics and Social Choice Theory, 2nd Edition* (pp. 1–9). Springer US. https://doi.org/10.1007/0-387-29368-X_1

Gehrlein, W. V., & Fishburn, P. C. (1976). Condorcet's Paradox and Anonymous Preference Profiles. *Public Choice*, *26*, 1–18. https://doi.org/10.1007/BF01725789

Gehrlein, W. V., Lepelley, D., & Smaoui, H. (2011). The Condorcet Efficiency of Voting Rules with Mutually Coherent Voter Preferences: A Borda Compromise. *Annals of Economics and Statistics*, (101/102), 107–125. https://doi.org/10.2307/41615476

Gibbard, A. (1973). Manipulation of Voting Schemes: A General Result. *Econometrica*, *41*(4), 587–601. https://doi.org/10.2307/1914083

Graham-Squire, A., & McCune, D. (2022). *A Mathematical Analysis of the 2022 Alaska Special Election for US House* (arXiv:2209.04764). arXiv. https://doi.org/10.48550/arXiv.2209.04764

Graham-Squire, A., & McCune, D. (2023). *An Examination of Ranked Choice Voting in the United States, 2004-2022* (arXiv:2301.12075). arXiv. https://doi.org/10.48550/arXiv.2301.12075

Green-Armytage, J. (2014). Strategic voting and nomination. *Social Choice and Welfare*, *42*(1), 111–138. https://doi.org/10.1007/s00355-013-0725-3

Green-Armytage, J., Tideman, T. N., & Cosman, R. (2016). Statistical evaluation of voting rules. *Social Choice and Welfare*, *46*(1), Article 1. https://doi.org/10.1007/s00355-015-0909-0

Gschwend, T., & Meffert, M. F. (2017). Strategic Voting. In *The SAGE Handbook of Electoral Behaviour*.

Guilbaud, G. T. (1952). *Les théories de l'intérêt général et le problème logique de l'agrégation*. https://doi.org/10.3406/ecoap.1952.3831

Hands, D. W. (2012). The Positive-Normative Dichotomy and Economics. In *Philosophy of Economics* (pp. 219–239). Elsevier. https://doi.org/10.1016/B978-0-444-51676-3.50009-9

Harrison, G. W., & McDaniel, T. (2008). Voting games and computational complexity. *Oxford Economic Papers*, *60*(3), 546–565. https://doi.org/10.1093/oep/gpm045

Hausman, D. M. (2024). Philosophy of Economics. In E. N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Fall 2024). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2024/entries/economics/

Heckelman, J. C. (2015). *Chapter 15: Properties and paradoxes of common voting rules*. https://www.elgaronline.com/edcollchap/edcoll/9781783470723/9781783470723.00023.xml

Holliday, W. H., Kristoffersen, A., & Pacuit, E. (2025). *Learning to Manipulate under Limited Information* (arXiv:2401.16412). arXiv. https://doi.org/10.48550/arXiv.2401.16412

Ianovski, E., Teplova, D., & Kuka, V. (2022). Welfare Effects of Strategic Voting Under Scoring Rules. In D. Baumeister & J. Rothe (Eds), *Multi-Agent Systems* (Vol. 13442, pp. 207–220). Springer International Publishing. https://doi.org/10.1007/978-3-031-20614-6_12

Ingham, S. (2019). Why Arrow's theorem matters for political theory even if preference cycles never occur. *Public Choice*, *179*(1–2), 97–111. https://doi.org/10.1007/s11127-018-0521-9

Institute for Mathematics and Democracy. (2025). *Empirical analysis of ranked choice voting methods*. Institute for Mathematics and Democracy. https://mathematics-democracy-institute.org/empirical-analysis-of-ranked-choice-voting-methods/

Kavner, J., & Xia, L. (2021). Strategic Behavior is Bliss: Iterative Voting Improves Social Welfare. *Advances in Neural Information Processing Systems*, *34*, 19021–19032. https://proceedings.neurips.cc/paper/2021/hash/9edcc1391c208ba0b503fe9a22574251-Abstract.html

Kelly, J. S. (1988). 4. Minimal Manipulability and Local Strategy-Proofness. *Social Choice and Welfare*, *5*(1), 81–85.

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.

Kube, S., & Puppe, C. (2009). (When and how) do voters try to manipulate?: Experimental evidence from Borda elections. *Public Choice*, *139*(1–2), 39–52. https://doi.org/10.1007/s11127-008-9376-9

Lagerspetz, E. (2016). *Social Choice and Democratic Values*. Springer International Publishing. https://doi.org/10.1007/978-3-319-23261-4

Lalley, S. P., & Weyl, E. G. (2018). Quadratic Voting: How Mechanism Design Can Radicalize Democracy. *AEA Papers & Proceedings*, *108*, 33–37. https://doi.org/10.1257/pandp.20181002

Lehtinen, A. (2007a). The Borda rule is also intended for dishonest men. *Public Choice*, *133*(1/2), 73–90. https://doi.org/10.1007/s11127-007-9178-5

Lehtinen, A. (2007b). The Welfare Consequences of Strategic Voting in Two Commonly Used Parliamentary Agendas. *Theory and Decision*, *63*(1), Article 1. https://doi.org/10.1007/s11238-007-9028-4

Lehtinen, A. (2008). The welfare consequences of strategic behaviour under approval and plurality voting. *European Journal of Political Economy*, *24*(3), Article 3. https://doi.org/10.1016/j.ejpoleco.2008.03.002

Lehtinen, A. (2011). A welfarist critique of social choice theory. *Journal of Theoretical Politics*, *23*(3), 359–381. https://doi.org/10.1177/0951629811411753

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), Article 2. https://doi.org/10.23941/ejpe.v8i2.200

Lepelley, D., Chantreuil, F., & Berg, S. (1996). The likelihood of monotonicity paradoxes in run-off elections. *Mathematical Social Sciences*, *31*(3), 133–146. https://doi.org/10.1016/0165-4896(95)00804-7

Lepelley, D., & Mbih, B. (1987). The proportion of coalitionally unstable situations under the plurality rule. *Economics Letters*, *24*(4), 311–315. https://doi.org/10.1016/0165-1765(87)90062-0

Lepelley, D., & Valognes, F. (2003). Voting Rules, Manipulability and Social Homogeneity. *Public Choice*, *116*(1/2), Article 1/2. https://doi.org/10.1023/A:1024221816507

List, C. (2022). Social Choice Theory. In E. N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Winter 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2022/entries/social-choice/

Lu, T., Tang, P., Procaccia, A. D., & Boutilier, C. (2012). *Bayesian Vote Manipulation: Optimal Strategies and Impact on Welfare* (arXiv:1210.4895). arXiv. https://doi.org/10.48550/arXiv.1210.4895

Make Elections Great Again Act, H.R. 7300, 119th Cong. (2026). https://www.congress.gov/bill/119th-congress/house-bill/7300

Mallows, C. L. (1957). Non-Null Ranking Models. I. *Biometrika*, *44*(1/2), 114–130. https://doi.org/10.2307/2333244

Maskin, E. (2022). How to Improve Ranked-Choice Voting and Democracy. *Capitalism and Society*, *16*(1).

McCune, D., & McCune, L. (2021). *The Curious Case of the 2021 Minneapolis Ward 2 City Council Election* (arXiv:2111.09846). arXiv. https://doi.org/10.48550/arXiv.2111.09846

Meir, R. (2017). Iterative voting. *Trends in Computational Social Choice*, *4*, 69–86.

Miller, N. R. (2012). Monotonicity failure in irv elections with three candidates. *Second World Congress of the Public Choice Societies*. https://rangevoting.org/MFandIRV.pdf

Miller, N. R. (2017). Closeness matters: Monotonicity failure in IRV elections with three candidates. *Public Choice*, *173*(1/2), 91–108.

Morreau, M. (2019). Arrow's Theorem. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/arrows-theorem/

Moulin, H. (1980). On strategy-proofness and single peakedness. *Public Choice*, *35*(4), 437–455. https://doi.org/10.1007/BF00128122

Moulin, H. (1988). Condorcet's principle implies the no show paradox. *Journal of Economic Theory*, *45*(1), 53–64. https://doi.org/10.1016/0022-0531(88)90253-0

Nitzan, S. (1985). The vulnerability of point-voting schemes to preference variation and strategic manipulation. *Public Choice*, *47*(2), 349–370. https://doi.org/10.1007/BF00127531

Nurmi, H. (2023). Deliberative Democracy and Incompatibilities of Choice Norms. *Behavioral Sciences*, *13*(12), 985.

Ornstein, J. T., & Norman, R. Z. (2014). Frequency of monotonicity failure under Instant Runoff Voting: Estimates based on a spatial model of elections. *Public Choice*, *161*(1/2), 1–9.

Pathak, P. A., & Sönmez, T. (2013). School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation. *The American Economic Review*, *103*(1), 80–106. https://doi.org/10.1257/aer.103.1.80

Pattanaik, P. K. (1976). Threats, Counter-Threats, and Strategic Voting. *Econometrica*, *44*(1), 91–103. https://doi.org/10.2307/1911383

Peleg, B. (1979). A note on manipulability of large voting schemes. *Theory and Decision*, *11*(4), 401–412. https://doi.org/10.1007/BF00139450

Penn, E. M. (2015). Arrow's theorem and its descendants. In *Handbook of Social Choice and Voting* (pp. 237–262). Edward Elgar Publishing. https://doi.org/https://doi.org/10.4337/9781783470730.00022

Peters, H., & Veselova, Y. (2023). On the safety of group manipulation. *Social Choice and Welfare*, *61*(3), 713–732. https://doi.org/10.1007/s00355-023-01469-z

Pritchard, G., & Wilson, M. C. (2007). Exact results on manipulability of positional voting rules. *Social Choice and Welfare*, *29*(3), 487–513. https://doi.org/10.1007/s00355-007-0216-5

Reyhani, R., Pritchard, G., & Wilson, M. (2009). *New Measures of the Diffculty of Manipulation of Voting Rules* (CDMTCS Research Reports CDMTCS-357). Department of Computer Science, The University of Auckland, New Zealand. https://hdl.handle.net/2292/3864

Riker, W. H. (1988). *Liberalism against Populism: A Confrontation between the Theory of Democracy and the Theory of Social Choice*. Waveland Press.

Saari, D. G. (1990). Susceptibility to manipulation. *Public Choice*, *64*(1), 21–41. https://doi.org/10.1007/BF00125915

Satterthwaite, M. A. (1973). *Existence of a Strategy Proof Procedure: A Topic in Social Choice Theory* [Doctoral thesis, University of Wisconsin-Madison]. ProQuest Dissertations & Theses (7321018).

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. https://doi.org/10.1016/0022-0531(75)90050-2

Sen, A. K. (1984). *Collective Choice and Social Welfare*. Elsevier Science & Technology. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=1877078

Sinnott-Armstrong, W. (2023). Consequentialism. In E. N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2023/entries/consequentialism/

Slinko, A., & White, S. (2013). *Is it ever safe to vote strategically?* (arXiv:1301.1420). arXiv. https://doi.org/10.48550/arXiv.1301.1420

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

Suzumura, K. (2000). Welfare Economics Beyond Welfarist-Consequentialism. *The Japanese Economic Review*, *51*(1), 1–32. https://doi.org/10.1111/1468-5876.t01-1-00135

Teplova, D., & Ianovski, E. (2022). *Comparing the Manipulability of Approval Voting and Borda* (arXiv:2203.15494). arXiv. https://doi.org/10.48550/arXiv.2203.15494

Tideman, T. N., & Plassmann, F. (2014). Which voting rule is most likely to choose the "best" candidate? *Public Choice*, *158*(3–4), Article 3–4. https://doi.org/10.1007/s11127-012-9935-y

Trump, D. J. (2025, April 20). [Online post]. Truth Social. https://truthsocial.com/@realDonaldTrump/posts/114367252307184336

Tsetlin, I., Regenwetter, M., & Grofman, B. (2003). The impartial culture maximizes the probability of majority cycles. *Social Choice and Welfare*, *21*(3), 387–398.

Väyrynen, P. (2021). Thick Ethical Concepts. In E. N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Spring 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entrieshick-ethical-concepts/

Veselova, Y. (2012). The Difference between Manipulability Indexes in IC and IANC Models. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2151263

Veselova, Y. A. (2020). Does Incomplete Information Reduce Manipulability? *Group Decision and Negotiation*, *29*(3), Article 3. https://doi.org/10.1007/s10726-020-09670-6

Walsh, T. (2010a). *An Empirical Study of the Manipulability of Single Transferable Voting* (arXiv:1005.5268). arXiv. https://doi.org/10.48550/arXiv.1005.5268

Walsh, T. (2010b). *Is Computational Complexity a Barrier to Manipulation?* (arXiv:1007.0776). arXiv. https://doi.org/10.48550/arXiv.1007.0776

Wollesen, B. (2025). *Descriptive Assumptions and Normative Justifications in Social Choice Theory: Ambiguity, Strategic Voting and Measurement* [Doctoral thesis, London School of Economics and Political Science]. https://doi.org/10.21953/lse.00004834