UNIVERSITY OF AMSTERDAM

MASTER THESIS

Modelling Meta-Agreement through an Agent-Based Model

Author:
Amir Sahrani

Examiner:
Dr. Fernando P. Santos

Supervisor:
Prof. Dr. Ulle Endriss

Assessor:
Dr. Davide Grossi

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Computational Science

in the

Computational Science Lab
Informatics Institute

March 17, 2025

Declaration of Authorship

I, Amir Sahrani, declare that this thesis, entitled ‘Modelling Meta-Agreement through an Agent-Based Model’ and the work presented in it are my own. I confirm that:

- □ This work was done wholly or mainly while in candidature for a research degree at the University of Amsterdam.
- □ Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- □ Where I have consulted the published work of others, this is always clearly attributed.
- □ Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- □ I have acknowledged all main sources of help.
- □ Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Your signature

Date: March 17, 2025

“The majority, standing in for the people, wills everything and therefore wills nothing”

Joshua Cohen

Abstract

Include your abstract here Abstracts must include sufficient information for reviewers to judge the nature and significance of the topic, the adequacy of the investigative strategy, the nature of the results, and the conclusions. The abstract should summarize the substantive results of the work and not merely list topics to be discussed.

Length 200–400 words.

Acknowledgements

Thank the people that have helped, supervisors family etc.

Contents

List of Figures

List of Tables

List of Algorithms

ABBREVIATIONS

CSL	**C**omputational **S**cience **L**ab
UvA	**U**niversitiet **v**an **A**msterdam

SYMBOLS

N	The set of all voters
X	The set of all alternatives
$\succ$	A preference relationship
$\mathcal{D}$	A domain of possible profiles
$\triangleleft$	A geometric order over candidates

CHAPTER 1

INTRODUCTION

Though Black's result [1] is a famous positive result, it is far from the only positive result relating to domain restrictions. We first look into various domain restrictions and their properties. To fully understand why single-peakedness is specifically desirable we outline the political and philosophical reasons first, after which we elaborate on the mechanism through which deliberation should result in single-peakedness according to List [2]. Finally, for completeness' sake, we mention critiques of this theory.

CHAPTER 2

Preliminaries

We first proceed by giving a short introduction of social choice. We outline the basic model, and restate well known results relevant to the following chapters.

2.1 The basic model

To model elections, or more generally voting games, we represent voters by the set N consisting of n voters. The possible outcomes of an election, we represent with the set A consisting of $|A|$ possible outcomes, from now on we will refer to the outcomes of an election as alternatives. Each voter can represent their preference on alternatives through a preference relation $\succcurlyeq_i$, for example if voter 2 prefers outcome a to outcome b, we write $a \succcurlyeq_2 b$. If, however, this voter wants to make it clear a is strictly better than b, we instead write $a \succ_2 b$. When a voter specifies their preferences on the entire set of alternatives we call this a (weak) linear order. We call the set of possible linear orders over the alternatives $\mathcal{L}(A)$, the set of weak linear orders is denoted by $\hat{\mathcal{L}}(A)$. Thus, for an election, all voters report a (weak) linear order, the set of each voters preference is called a profile, denoted by $\boldsymbol{R}$. Finally, a rule f decides the outcome of the election based on the profile. We discuss the specifics of these rules in Section section 2.2.

The last general tool we will need is the *majority relation*. Given some profile $\boldsymbol{R}$ we can construct a majority relationship as follows: for each pair of alternatives x, y, we ask how many people prefer x to y; if the number of people who prefer x to y is greater than the other way around we write $x \succ_{\text{maj}} y$, if we have an even number of voters, these two number can be equal and this becomes a weak preference, we simply write $x \succcurlyeq_{\text{maj}} y$ (defaulting to lexicographical order). We proceed with an example.

EXAMPLE 1: *Majority relation*

1	2	3
a	b	a
b	c	c
c	a	b

Given the profile on the left, we first start by comparing a to b, both voters 1 and 3 prefer a to b, thus the majority has prefers a to b. Comparing b to c,the majority prefers b to c. Finally, comparing a to c, a is preferred again. Thus, the majority relation is $a \succ_{\text{maj}} b \succ_{\text{maj}} c$

A majority relation can be a-cyclic, and transitive, though neither are guaranteed. An a-cyclic majority profile is simply a majority relation without any cycles, meaning there does not exists a series of alternatives $a_1, \ldots a_n$ such that $a_1 \succ a_2 \succ \ldots \succ a_n \succcurlyeq a_1$. Transitivity is very similar, stating that the preferences between alternatives are transitive in that for any triplet of alternatives x, y, z if $x \succcurlyeq y$ and $y \succcurlyeq z$ then $x \succcurlyeq z$. These notions are similar, but transitivity is a stronger requirement, as it includes indifference.

2.2 Social Choice Functions

In order to decide the outcome of an election, we pick a social choice function f, this function should map all possible profiles to an outcome, thus $f : \boldsymbol{R} \to A$. A famous example of a SCF is the plurality rule, which simply elects the alternative voted into first place most often, though simple, it can also lead to a tie. Since the outcome of our SCF is only allowed to be a single alternative, the plurality rule needs to be equipped with a tie breaking mechanism in order to be a valid SCF, we require the tie-breaking to be deterministic.

2.2.1 Axioms

Though any voting rule that outputs one alternative for each profile is valid, for real elections organizers likely will want to ensure the rule has certain nice properties, such as not favoring some alternatives of others. In social choice these properties are called axioms, and the procedure of designing a rule based on axioms is called the axiomatic approach. The name of the property just described is the axiom of neutrality, stating that the voting rule should be neutral with respect to the alternatives. In this work three main axioms are of importance.

Axiom of Surjectivity. A SCF f is surjective, if for every alternative, there exists a profile R such that $f(R)$ elects it.

Axiom of Non-Dictatorship. A SCF f is non-dicatorial, if there does not exist a voter i such that $f(R) = \text{top}(i, R)$ for all profiles R, where $\text{top}(i, R)$ extracts voter i's most preferred alternative from profile R.

Axiom of Strategyproofness. A SCF f is strategy proof if, for any voter $i \in N$, i cannot report an untruthful preference, and thereby cause the outcome of the elective to improve for them.

Axiom of Anonymity. A SCF f is anonimous if, when the labels of voters are shuffled, the winning alternative stays the same.

Axiom of Neutrality. A SCF f is neutral if, when the labels of the alternatives are shuffled, the winning alternative is the alternative who is ranked the same by each voter as the original winning alternative.

Another way to interpret strategyproofness is that the SCF should maximize the outcome for all voters, as such if a voter reports something which is not their true preference, the outcome will maximize the wrong preference and thus result in an outcome that is worse for you.

There are many more axioms on could reasonably argue for, however, these are enough to lead to the main impossibility result this works focuses on.

2.3 Negative results

Classic social choice theory has many negative results, one such example is the Condorcet cycle. This is a specific profile that results in a cycle in the majority relation, as shown in the following example.

EXAMPLE 2: *Condorcet cycle*

1	2	3
a	b	c
b	c	a
c	a	b

Voters 1 and 3 prefer a to b, forming a majority, next voters 1 and 2 prefer b to c, forming another majority. However, voters 2 and 3 prefer c to a forming a majority, and thus creating a cycle.

It is not hard to convinces oneself that under weak preferences the Condorcet cycle can occur anytime there are 3 or more alternatives and voters. While under strict preferences this can occur anytime the number of alternatives is odd and greater than 3, with the number of voters being a multiple of the number of alternatives. As we will show later, this profile can be the cause of some impossibility results.

One of the major negative results in social choice is that of the Gibbard Satherswaite theorem [3, 4].

Theorem 2.1. [Gibbard-Satherswaite] There exists no resolute Social Choice Function for elections with $|A| \geq 3$ that is surjective, strategyproof, and non-dictatorial.

Proof. ... □

2.4 Domain Restrictions

Many negative results are a consequence of a few ill-behaved profiles, if one can argue such profiles do not occur in the real election, there is some hope of constructing SCF's satisfying our axioms. To speak more formally about profiles "not occuring", we introduce Domain restrictions, for this we use the definition by Elkind et al. [5].

DEFINITION 1: *Domain*

Given a set of voters N, alternatives A, and conditions C, the domain $\mathcal{D}$ of an election is the set of all profiles R such that all conditions C are satisfied.

This definition is different from usual definitions in social choice in so far as it talks about allowed profiles instead of allowed votes.

As stated earlier, the Condorcet profile is one such ill-behaved profile, as each alternative, holds a majority preference over another alternative. Naturally one might consider if this profile might even come up in practice, since though conceivable it seems generally unlikely that there exists a perfect split in opinions. Quite naturally one of the first "solutions" one might consider is when the number of voters is not a multiple of the number of alternatives, though this is hardly a useful solution since it only prevents Condorcet cycles, it is the first example of a domain restriction, we define it as follows

DEFINITION 2: $\mathcal{D}_{\text{No-tie}}$

Let X be the set of alternatives and N be the set of voters, of size n such that $n \neq k \cdot |X|$ for any $k \in \mathbb{N}$. We call this domain $\mathcal{D}_{\text{No-tie}}$.

This allows us to state our first proposition.

Proposition 2.2. The plurality rule never returns a $|X|$-way tie between alternatives when applied to $\mathcal{D}_{\text{No-tie}}$

Proof. Assume, for the sake of contradiction, the plurality in fact does return a tie this must mean that all alternatives were ranked first an equal number of times, call this k, necessarily then, we have need exactly $k \cdot |X|$ voters, but this leads to a contradiction, as this would no longer be inside $\mathcal{D}_{\text{No-tie}}$. □

This is a simple result, but it serves as an example on how we can use the properties of the domain to prove things about the election. Gaertner [6] establishes 2 ways in which a domain can be restricted. Firstly we can restrict the domain to a number of voters or alternatives, which is what we did in $\mathcal{D}_{\text{No-tie}}$. Secondly, the domain can be restricted to have a certain structure, such as being single-peaked.

2.4.1 Single-Peaked profiles

In a election the alternatives might represent a axis, such that a voters is prefers an alternative more if they are closer to them on the axis. For example, if the alternatives represent free-trade vs regulation, we can imagine that a voter that is of the opinion that free trade is of ultimate importance will prefer alternatives more the more the are on the side of free trade. More generally, we call a profile single-peaked if there exists an axis on which we can place the alternative such that all voters' preferences have a single peak on this axis. Definition 3 makes this notion formal.

DEFINITION 3: *Single-Peaked Profiles*

A profile P is single-peaked, if given some ordering $\triangleleft$ over the alternatives, it holds that for all voters i, and all $a, b, c \in X$, if $a \triangleleft b \triangleleft c$, then either $a >_i b$ or $c >_i b$, but never both.

CHAPTER 3

LITERATURE REVIEW

In this chapter we explore previous results, as well as introducing relevant concepts.

3.1 Condorcet Domain

If our goal is to prevent Condorcet cycles, or in general have transitive majority relations, the best we could hope to do is to apply our domain restriction such that our domain contains all profiles P such that P has a (weak) Condorcet winner. We call this domain $\mathcal{D}_{\text{Condorcet}}$. Under this domain, let $f_{Condorcet}$ be the Condorcet Rule, which picks a Condorcet winner. Then $f_{Condorcet}$ is strategyproof over $\mathcal{D}_{\text{Condorcet}}$ [5].

Proof. (Elkind et al. [5]). Assume, for the sake of a contradiction, we have profiles $P = (\succ_1 \dots \succ_i \dots \succ_n)$ and $P' = (\succ_1 \dots \succ_{i'} \dots \succ_n)$ such that:

$$f_{Condorcet}(P) = a, \quad f_{Condorcet}(P') = b, \quad \text{and } a \neq b$$

Then under P a strict majority $N' \subseteq N$ have $a \succ b$, but $i \notin N'$, thus in P', N' is still a majority preferring a to b, but this is in contradiction to b winning in P'. $\square$

This result is strengthened by Campbell and Kelly [7, 8], showing that for an odd number of alternatives, $f_{\text{Condorcet}}$ is the only voting rule over $\mathcal{D}_{\text{Condorcet}}$ that is Strategyproof, Surjective and Non-dictatorial.

When Surjectivity is strengthened to Neutrality, and Non-dictorship to Anonymity, $f_{\text{Condorcet}}$ is the only Strategyproof voting rule over $\mathcal{D}_{\text{Condorcet}}$ for an odd number of voters [9].

3.1.1 Hereditary Domains

Though this result is positive, we might wonder how stable it is, for this we need to define a notion of stability. On natural way to think about it is as follows: suppose one of the alternatives or voters drops out, do we keep the nice structure of the domain? If this is true we call a domain *Hereditary*.

DEFINITION 4: *Hereditary* (Elkind et al. [5])

A domain restriction onto $\mathcal{D}$ is *hereditary* if, for every profile $P \in \mathcal{D}$, and every profile P', that can be obtained by deleting voters and alternatives from P, P' is also in $\mathcal{D}$

$\mathcal{D}_{\text{Condorcet}}$ is not hereditary, this is easy to see through an example:

EXAMPLE 3: $\mathcal{D}_{\text{Condorcet}}$ *is not hereditary*

v_1	v_2	v_3	v_4
a	b	c	a
b	c	a	c
c	a	b	b

We can see that in this example, a is the weak Condorcet winner, as it beats b and is tied with c, however if we remove voter 4, we return to the original Condorcet cycle.

A domain not being hereditary means that the nice properties of the domain can be unstable, as the number of voters and alternatives might not be known or could be manipulated. Instead, we might want to look at hereditary domains. The first hereditary domain we present, will also be the main focus of this thesis. This is the domain of all single-peaked profiles.

Proposition 3.1. (Elkind et al. [5]). $\mathcal{D}_{\text{SP}}$ is hereditary.

Proof. (Voter Deletion). If we remove a voter, this does not affect the other voters, thus the profile is still single-peaked. ✓

(Alternative Deletion). Consider any voter i and their single-peaked vote, if we remove some alternative x, to this voter all alternatives which voter i preferred to x stay in the same position, while all other alternatives move up one rank, thus preserving the order, and single-peakedness. ✓ □

Instead of ordering the alternatives, we can imagine instead ordering the voters, such that we have a leftmost and rightmost voter, and all other voters can be placed between them based on their difference. In this case, a profile is single-crossing if, for any alternative a, its preference relation to another any alternative b flips at most once when traversing the voters in order $\lhd$.

DEFINITION 5: *Single-Crossing Profiles* (Elkind et al. [5])

A profile P is single-crossing w.r.t. some ordering $\lhd$, if for any $a, b \in X$, $\{i \in N : a > b\}$ and $\{i \in N : b >_i a\}$ are both intervals over $[n]$. A profile P is single crossing if the votes can be permuted such that it is single crossing w.r.t. some ordering.

Similar to single-peaked profiles, the domain of single-crossing profiles, $\mathcal{D}_{\text{SC}}$ is also hereditary

Proposition 3.2. $\mathcal{D}_{\text{SC}}$ is hereditary

Proof. (Voter Deletion). Deleting a voter preserves the ordering between voters, as such this cannot introduce a new crossing between alternatives. ✓

(Alternative Deletion). If we remove an alternative, the voters' rankings of the other alternatives does not change, thus preserving single-crossing. ✓ □

As to goal is to ensure we find ourselves in nicely structured domains, we need some mechanism through which we can ensure this is the case. Deliberation is the mechanism of choice, we will now provide a brief overview of the literature surrounding deliberation.

3.2 The History of Deliberation and Meta-Agreement

We have provided an overview of different domain restrictions and their properties, showing they avoid Condorcet cycles. Some argue however, that Condorcet cycles are empirically rare. The next section is dedicated to explaining why this is so through examining the historical ideas around deliberation and deliberative democracy, as well as that of Meta-Agreement.

3.2.1 Deliberation

Though deliberation is intuitively familiar, namely the process of multiple people talking through a problem with the goal of coming to an agreement, compromise or solution, providing a definition that is both clear and consistent with the literature in Political Science, Philosophy and Social choice is difficult. This intuition leaves some of the reasons for and goals of deliberation, as stated in the literature, unmentioned.

Freeman [10] gives an overview of deliberative democracy, in which he shares the intuitive idea that a deliberative democracy contains open discussion, open legislative deliberation and a pursuit of the common good. He also notes that there is no common agreement on the central features of a deliberative democracy, one account is that of deliberative democracy simply involving discussion among the public before voting. Another similar account is that this voting must not only be preceded by deliberation, but also general communication, all of which intended to change people's preferences. He further proceeds to give a more detailed conception of deliberative democracy, according to which a deliberative democracy is one in which political agents or their representatives

1. Aim to collect, deliberate and vote
2. Represent their sincere and informed judgements
3. Vote and deliberate on measures beneficial to the common good on the citizens
4. Are seen and see each other as political equals
5. Have Constitutional rights and their social means enable them to participate in public life
6. Are individually free, such that they have their own freely determined conceptions of the good
7. Have diverse and disagreeing conceptions of the good
8. Recognize and accept their duty as democratic citizens, and do not engage in public argument on the basis of their particular moral views incompatible with public reason.
9. Agree reason is public, in so much as it is related to and advances common interests of citizens
10. Agree that their common interest lies primarily in freedom, independence and equal status as citizens.

Firstly, why suddenly talk about deliberative democracy? how is this different from deliberation. Secondly, does this imply that this is already the case? Or should we aim to achieve a deliberative democracy?

These features allow us to be more precise when we talk about a deliberative democracy, and in turn be more careful about what deliberation must entail. Cohen [11] further argues that deliberation is needed for democratic legitimacy. By this he means that without deliberation, a democracy is simply the will of the majority, but since majority rule is unstable, it is simply a reflection of the particular institutional constrains at the time. He further goes on to describe the *ideal deliberative procedure* as follows

What does it mean to be unstable in this context? Elaborate on "particular institutional constraints"

1. Ideal deliberation is *free*, participants regard themselves as only bound by the results of the deliberation, and the preconditions thereof. Participants act in accordance with the decision made through deliberation, and it being agreed on is sufficient reason to do so.
2. Ideal deliberation is *reasoned*, parties are required to state their reasons for advancing proposals.
3. In ideal deliberation, parties are *equal*, both formally and substantively. There are no rules that single individuals out, and existing distributions of power to no lend a party the opportunity to contribute to deliberation.
4. Ideal deliberation aims to arrive at *consensus*, which can be rationally defended.

3.2.2 Meta-Agreement

Consensus, sometimes referred to as substantive agreement, then seems like a natural goal for deliberation. Elster [12] argues that this is not only the goal, but through unanimous agreement this process completely replaces voting, thereby circumventing Arrow's impossibility theorem: "Or rather, there would not be any need for an aggregation mechanism, since a rational discussion would tend to produce unanimous preferences." (p. 112). Though it would be desirable to circumvent Arrow's impossibility theorem, in practice people, even after deliberation might not, indeed often do not, come to full substantive agreement. List [2] instead proposed another lens through which we can analyze deliberation and the type of agreement it induces.

Under *Meta-agreement* individuals do not need to agree on their most preferred outcome, instead they only need to agree on the dimensions of the problem. To contrast this with substantive agreement, under which individuals do not need to conceive of the problem in the same way, all they need is to agree on the same outcome. This means that under substantive agreement, voters can agree outcome $a > b$ for different reasons, while under meta-agreement, if voters disagree on $a > b$ it must be for the same reason.

According to List [2] there are three hypotheses that need to be satisfied for deliberation to induce meta-agreement:

- D1 Deliberation leads people to discover a single *issue*-dimension
- D2 Deliberation lets people place all possible alternatives in this *issue*-dimension

D3 After this deliberation, people update their preferences by picking a preferred outcome, and all other rankings are based on the distance to this outcome in the *issue*-dimension

All these are necessary conditions for *meta-agreement*, from this is it also clear to see that, given that there is exactly 1 *issue*-dimension, single-peaked profiles are, by definition, a direct consequence. This is the main reason meta-agreement is desirable, as it lets us circumvent the Gibbard-Satterthwaite theorem [3, 4] through restricting the domain of preference profiles to the single-peaked domain $\mathcal{D}_{\text{SP}}$

List et al. [13] provide empirical evidence for this theory of deliberation, showing deliberation increases proximity to single-peakedness (PtS), which they define as $S = \frac{m}{n}$ where $n = |N|$ and m is the largest subset of voters such that their profile is single-peaked. Furthermore, they also introduce the notion of salience, which represents to what extent a topic is salient in the voting population. In order to test whether deliberation increases single-peakedness *through* meta-agreement, they test the following four hypotheses: (H1) deliberation increases proximity to single-peakedness. (H2')[1] high salience issues show less increase in PtS than low salience issues. (H3) Effective deliberation, in the sense that more is learned during deliberation, results in bigger increases of PtS. (H4) All things equal, the increase is largest for issues with natural *issue*-dimensions. They find support for all these hypothesis, showing that on low-moderate salience issues PsT increases following deliberation. As well as showing that individuals learning most show the greatest movement towards single-peakedness.

It is important to note that these claims simply predict what will happen, there is not much explanatory power to these claims. Little is known about to process be which voters signal the issue dimensions, nor how they decide on which ones to present.

Furthermore, Ottonelli and Porello [14] show meta-agreement to be a stronger requirement than it may seem at a first glance. Firstly for (D1) to hold, the *issue*-dimension must hold some semantic meaning, as otherwise it is unclear how people can exchange conceptualization of the problem otherwise. Furthermore, the issues must consist of 2 semantic issues, otherwise with only 1 dimension voters simply reach substantive agreement. A further restriction on these two dimensions is that they need to be opposite, with opposite justifications. If this is not the case, a voter can agree with both justifications, and thereby introduce a new dimension “balance", which then violates the conditions under which single-peaked profiles guarantee the existence of fair, strategyproof voting rules. D2 requires that all voters share the exact same semantic understanding of the

[1] This is a test for a corollary. H2 states that the rate of increase of proximity to single-peakedness decreases. This is not experimentally testable, however since high salience means some sort of deliberation has happened before, they expect this to approximate this affect.

dimension, and the outcome associated with each alternative. Finally D3 requires D1 and D2 to have happened before in order. Clearly D3 is the weakest of the three.

Thus, meta-agreement as a means for single-peaked profiles is still quite restrictive, needing multiple forms of unanimity, and only applying to problems with certain properties. Nonetheless, meta-agreement could still provide explanatory power to deliberation.

3.3 Related Work

Rad and Roy [15] model deliberation and its effect on single-peakedness, though they argue single plateauedness is a more accurate term. To this end, they model each voter to have preferences order, and deliberation being the process of all voters announcing their preferences, after which all other voters update their current preference towards that of the announced ranking, in doing so they might have a bias towards their own preference, as such they try to minimize the distance between their current preference and the announced one. This process repeats until all voters have announced their opinion once, this happens for a number of rounds. The preference a voter adopts when updating must lie between their current profile and the announced profile, which profiles are considered to be "between" is defined by the distance metric used. They considered three metrics, the Kemeny-Snell (KS) [16], Duddy-Piggins (DP) [17], and Cook-Seiford (CS) [18]. Both KS and DP depend on the judgement set resulting from the voters preferences, the KS distance is then defined as the number of binary swaps a judgement set needs to undergo before it becomes the target judgement set, an example for such a swap would be going from $(a \succ b)$ to $\neg(a \succ b)$. The DP distance is defined on the graph of judgement sets, where 2 sets share an edge if there is no judgement set between them. Since KS and DP share their notion of betweenness, we introduce their betweenness as follows.

Definition 6: *J-Betweenness*

A judgement set J_i is between preferences J_j and J_k if for every proposition about $x, y \in A$, J_i either agrees with J_j or J_k.

Figure 3.1 shows a graph used for the DP distance in the case of 3 alternatives, for simplicity the associated preferences are used to label the judgement sets.

The CS distance is simpler and is simply defined as the number of positions two voters disagree on, and a preference is between two others if for each position it agrees with one of the two preferences.

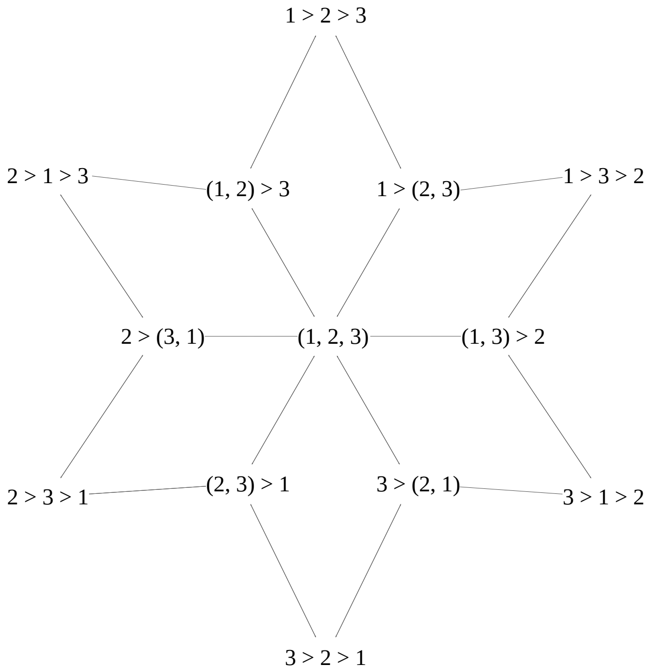

FIGURE 3.1: The graph of judgement sets for all preferences over three alternatives, brackets indicate ties.

Each distance has different trade-offs, CS is the simplest, but might exaggerate the distance when there are many alternatives, for example if 2 voters agree on the relative ranking of all but 1 alternative, which one voter happens to rank first, thereby shifting all other profiles right. The KS distance, using judgement sets instead of raw profiles captures this more effectively, while still being relatively easy to compute, but in cases of more disagreement, it is likely to over count the distance, since the binary changes to not capture logical necessities. For example, swapping $(a \succ b)$ to $\neg(a \succ b)$ must result in $(b \succ a)$ becoming true (in the case of strict preferences), thus one might reasonably conclude this should only count as 1 step. DP improves upon this, but in doing so becomes much harder to compute, mainly through the cost of constructing the full graph of judgement sets, which grows in $O(2^n)$ in the number of vertices, where n is the number of alternatives. This can easily be verified by noting that the number of judgements sets over n alternatives is $O(2^{n^2})$, where there is a proposition for each pair of alternatives, and a binary choice on each proposition.

Apart from these distances, they also define a voter as a tuple of a (weak) preference and a bias (towards their current position) $v = \langle r, b \rangle$, with $b \in \mathbb{R}_{[0,1]}$. Finally, a deliberation step $D_s : V \times r \to V$, with V being a set of voters and s being one of the spaces (KS, DP, CS). A round of deliberation consists of n deliberation steps, where each voter has announced their opinion once. We formulate this procedure in the following program:

input : Set of Voters V, metric space s
output: Updated set of Voters V

$V_{\mathrm{u}} \leftarrow V$ `// Set of unannounced voters (references to` V`)`
while $|V_u| > 0$ **do**
  Select a random $v \in V_{\mathrm{u}}$
  $V_{\mathrm{u}} \leftarrow V_{\mathrm{u}} \setminus \{v\}$
  $V \leftarrow D_s(V, v.r)$ `// Update voters based on` v`'s preference`

The deliberation step D_s then updates all voters such that their new preference minimize the following formula.

$$r = \sqrt{bd_s(r_i, r')^2 + (1-b)d_s(r_j, r')^2} \tag{3.1}$$

Where r_i, r_j are the voters and the announced preference, respectively, and r' is the voters new preference.

We present a replication and extension of their work chapter 6. Furthermore, we present novel (negative) results based on this model in chapter 4.

Though this model is simple and captures some communication of preferences, if we attempt to use it to model meta-agreement, it seems to be lacking in at least two important ways. Firstly, agents do not conceive of anything relating to the structure of the problem. They simply announce their preferences, and all other listen and update accordingly, thereby moving to some sort of substantive agreement. Secondly, the model presupposes that all opinions are equally defensible, and that each voter is equally able to formulate this defense. To address this we formulate a new model in Chapter 4.

CHAPTER 4

THEORETICAL RESULTS

In the model of deliberation of Rad and Roy [15], outlined in Section 3.3, aim to model deliberation and show that deliberation results in nicely structured profiles which allow for strategy proof voting rules. One important caveat, given by the authors as well, is all participants should honestly and truthfully participate in deliberation. We now provide a formal statement, showing deliberation does not prevent strategic behavior.

Proposition 4.1. The process of deliberation over $|A| \geq 3$ through deterministic deliberation procedure $D : \mathcal{L}(A)^n \to \mathcal{L}(A)^n$, followed by voting with voting rule f cannot be surjective, strategyproof and non-dictatorial.

Proof. Assume, towards a contradiction, such a pair of deliberative procedure (D) and voting rule (f) exists. Any deterministic deliberation procedure D could, in principle, be embedded into a voting rule $f'(\boldsymbol{R}) = f(D(\boldsymbol{R}))$, such that the voting rule simulates D before applying f, which would result in voting rule f' being surjective, strategyproof and non-dictatorial. This is a contradiction, by the Gibbard-Satterthwaite theorem [3, 4]. □

We extend upon this result, showing the inclusion of biases in voters does not mitigate the negative result. For this we define DB as follows:

DEFINITION 7: *Biased Deliberation*

A deliberative procedure with biases DB : $\mathcal{L}(A)^n \times \mathbb{R}^n_{[0,1]} \to \mathcal{L}(A)^n$ is an extension on a standard deliberative procedure. DB has access to the bias each voter has to their own opinion.

We now proceed with a corollary on Proposition 4.1. Towards this we assume biases are true, in the sense that a voter cannot help but be 'convinced' by the presented profiles as much as their bias allows for this. We think this assumption is a weak and natural in the light of the current model. Furthermore, a violation of this assumption would not imply the following corollary to be false, instead the bias itself becomes a point of strategy, allowing voters to pretend to be more hardheaded than they in fact are.

Corollary 4.2. A deliberative procedure with biases, followed by voting with any voting rule f, cannot be surjective, strategyproof and non-dictatorial

The proof of this follows from a reduction of the biased Deliberation DB to general deliberation D.

Proof. Take any election consisting of biased deliberation DB and voting rule f, since biases $\boldsymbol{b}$ are true by assumption, they must be fixed, meaning that $\boldsymbol{b}$ is not reported but some fact of the matter. If this election was immune to strategic manipulation, then a deliberative procedure D could embed this b, and simulate biased deliberation DB, resulting in $D'(\boldsymbol{R}) = \mathrm{DB}(\boldsymbol{R}, \boldsymbol{b})$. As a direct corollary to Proposition 4.1, such a D' cannot be surjective, strategyproof and non-dictatorial, showing a contradiction. □

This result is independent of the metric space chosen. From here we now show that even if we take the deliberation procedures on its own, it still not immune to strategic manipulation. For this we restate strategyproofness as follows:

DEFINITION 8: *Strategyproofness of Deliberation*

A deliberation procedure is strategyproof if there exists no voter i such that there is a profile $\boldsymbol{R}$, in which i misreporting their preference R_i as R_i' results in the profile after deliberation $D(\boldsymbol{R})$ is further from the i's original preference than if they had reported R_i'. This distance is measured as

$$\mathrm{Dist}(R_i, D(\boldsymbol{R})) \geq \mathrm{Dist}(R_i, D(\boldsymbol{R}')).$$

Where the Dist function is simply the sum of all distances between R_i and all preferences in $\boldsymbol{R}$.

One important note is that in the final profile, the preferences of voter i might not be the same as it was before the deliberation. That is why the distance is calculated w.r.t. i's original preference. Intuitively this could be read as i misreporting their preference to prevent even their own mind from being changed. Using this definition, we show that the deliberative procedures, under the metric spaces KS, DP, CS are not strategyproof. Stated as follows:

Proposition 4.3. Deliberation under distance measures *KS*, *DP*, *CS* is not strategyproof, for $n \geq 2$ and $m \geq 3$.

We provide a proof by construction, we show how to do this for KS and DP, as they share the same profiles for this proof. The proof for CS is laid out in Appendix A

Proof. Assume the following population: we have voter 1 whose bias is 1, and all other voters $j \neq 1$ have bias 0.5. Furthermore, we have $\text{Dist}(R_1, R_j) = 2$ for all j. Voter 1 now has the option to report R'_1 instead, which has $\text{Dist}(R'_1, R_j) = 4$ and $\text{Dist}(R'_1, R_1) = 2$. If voter 1 reports R'_1, then all j will update towards 1's true preference, as using equation (3.1) we get $r(R_j, R'_1, R_1) = 4$, while $r(R_j, R'_1, R_j) = r(R_j, R'_1, R'_1) = 16$.

Resulting in $\text{Dist}(R_1, D(R_1, \boldsymbol{R}_{-1})) = 2(n-1) > \text{Dist}(R_1, D(R'_1, \boldsymbol{R}_{-1})) = 0$.

Since 1 has a bias of 1, the order of the deliberation has no effect.

We now show that for distance measures KS and DP, there exists these 3 preference orderings such that the necessary profile can be constructed. We use the following profiles:

$$R'_1 = a \succ c \succ b \succ \cdots \succ m,$$
$$R_1 = a \succ b \succ c \succ \cdots \succ m,$$
$$R_j = b \succ a \succ c \succ \cdots \succ m.$$

As we are only allowing strict preferences, both distance metrics behave locally the same, with the distance of two profiles being 2 whenever one is 1 swap of alternatives away from the other. This means that R_i and R_j have a distance of 2, as well as R'_1 and R_1 having a distance of 2. In this case the total distance from R'_1 to R_j is simply the sum of the local distances for both distance metrics, thus satisfying our requirements. □

These results show it is frivolous to attempt to design a strategy proof deliberation procedure of the likes shown. Instead, focus is now brought to modeling 'ideal' deliberation, as laid out in Section 3.2.2. We provide the following mathematical formulations to the four tenants laid out. *Freedom*: voters can report any preference, *Reason*: voters are rational, *Equality*: no voter has special rights *Consensus*: voters deliberate aim to reach consensus. Which we extend with *Honesty*: Voters represent their true beliefs and preferences only.

4.1 Our model

In an attempt to model meta-agreement through deliberation, our model needs to make a proper distinction between the 'level' and the meta 'level'. In order to do so, we propose the following, let $E = \{e_1, \cdots e_k\}$ denote the set of events that could occur. A voter $i \in N$, at the base level has a preference over these events. At a meta level, however, a voter has a probability distribution over the outcomes, conditional on the alternative $a \in A$ elected. Now deliberation can be modelled as a deliberation on probability distributions.

More specifically, we model deliberation as a DeGroot learning model. In this model, a voter is a node in a graph, and deliberation can be modeled as a Markov chain. A voter has probability matrix P_i, defined as follows:

$$P_i = \begin{bmatrix} p_i(e_1|a) & \dots & p_i(e_1|m) \\ \vdots & \ddots & \vdots \\ p_i(e_k|a) & \dots & p_i(e_k|m) \end{bmatrix}$$

Under the constraint that each column must sum to one, representing that an alternative must have at least one outcome. Note that this does not mean that all outcomes have to be equally likely, nor that an alternative can only represent one outcome. Let $\boldsymbol{P} = [P_1, \dots, P_n]^T$ denote the population opinion, which has shape $|N| \times |E| \times |A|$.

In order to extract a ballot from this matrix, we assume Borda scores for each event, using this we can model the most preferred alternative as the one that maximizes the expected utility with respect to the Borda scores, starting from 1 instead of 0, and voter i's subjective probability.

Then a deliberative step can be modelled using a transition matrix T, defined as follows:

$$T = \begin{bmatrix} t_{11} & \dots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nn} \end{bmatrix}$$

Here each t_{ij} represents how much voter i trusts the opinion of voter j, in order for this to be a proper stochastic matrix, all rows must sum to one, and have non-negative entries. Although this last requirement could be seen as unrealistic, as a voter might actively distrust another voter and update away from their opinion.

Using this, we can now model the opinions of voters after a deliberative step as a matrix multiplication:

$$P^{(1)} = TP^{(0)} \tag{4.1}$$

The resulting probability distribution is then simply a weighted linear combination of all probability distributions.

Finally, we provide an example of the first deliberation round in example 4.1

EXAMPLE 4: *DeGroot deliberation*

We have voters $N = \{1, 2\}$, events $E = \{e_1, e_2\}$, and candidates $A = \{a, b\}$. The voters both think that $e_1 \succ e_2$, and this these are the probabilities:

$$P_1 = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 1 \end{bmatrix} \quad P_2 = \begin{bmatrix} 1 & 0.9 \\ 0 & 0.1 \end{bmatrix}$$

This results in voter 1 preferring candidate b over candidate a, while voter 2, prefers a. Intuitively, since voter 1 thinks e_1 is equally likely for each alternative, while e_2 will not happen under a, it makes sense for them to prefer candidate b.
Now deliberating with the following trust matrix:

$$T = \begin{bmatrix} t_{11} & \dots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nn} \end{bmatrix}$$

We get the following updated opinions:

$$\begin{aligned} \boldsymbol{P}^{(1)} &= T\boldsymbol{P}^{(0)} \\ &= T\begin{bmatrix} P_1 P_2 \end{bmatrix}^T \\ &= \begin{bmatrix} (0.3P_1 + 0.7P_2) & (0.2P_1 + 0.8P_2) \end{bmatrix}^T \\ &= \begin{bmatrix} \begin{bmatrix} 0.35 & 0.63 \\ 0.15 & 0.37 \end{bmatrix} & \begin{bmatrix} 0.9 & 0.72 \\ 0.1 & 0.18 \end{bmatrix} \end{bmatrix}^T \end{aligned}$$

These new probabilities are not yet in full consensus, however, looking at their corresponding ballots there is consensus on their most preferred alternative.

4.1.1 Consensus

Using this model of deliberation, meta-agreement can be seen as some common probability distribution over all events. If the goal of deliberation is meta-agreement, then the study of interest becomes the dynamics of convergence towards a unified probability distribution.

We present a summary of results relating to strongly connected graphs, as well as graphs for which there exists only closed and strongly connected subsets of nodes. For other results we refer to Golub and Jackson [19]. Firstly we focus on the strongly connected graphs.

Proposition 4.4. (Golub and Jackson [19]). For a strongly connected matrix T, the following properties are equivalent:

- T is Convergent
- T is Aperiodic
- There exists a left eigenvector $\boldsymbol{s}$ for matrix T, with corresponding eigenvalue 1, whose entries sum to one, such that for every P_i, we have

$$\left(\lim_{t\to\infty} T^t \boldsymbol{P}\right)_i = \boldsymbol{s}\boldsymbol{P}$$

This result is very positive from a convergence dynamics point of view, as no knowledge of the initial distribution is needed to determine convergence, it allows us to simply verify one of these three properties on the network. Though strongly connected graph might be a strong requirement, in the case of small scale (in person) deliberation, this might not be infeasible. Fortunately, even outside this setting it might be possible to reach convergence. For this we first define what a closed set of nodes is.

DEFINITION 9: *Closed set of Nodes*

A set of Nodes $C = \{1, \ldots, n\}$ is closed if for each $i, j \in C$ we have $T_{ij} \geq 0$ and for each $i \in C, j \notin C$ we have $T_{ij} = 0$

Using this definition, if each node is part of a closed set, we can form the following proposition

Proposition 4.5. (Golub and Jackson [19]). If for each $i \in N$, i is a member of a closed set in the graph, and each closed set is strongly connected, T is convergent.

CHAPTER 5

METHODS

We determine whether our profiles are single-peaked using the algorithm by ..., implemented in `Ocaml` as follows

```
let agent = 1 in
let text = "Testing" in
(* This is a comment *)
print_int agent
```

CHAPTER 6

Experimental Results

We first present a full replication and extension of the work by Rad and Roy [15]. Then we present the simulations based on our model of meta-deliberation, as well as the results of the sensitivity analysis on both models. All code for the replication, main experiment and visualization can be found in this Repository.

6.1 Replication

We are able to fully replicate the results found by Rad and Roy [15], in Figure 6.1 we see that while the bias is less than 0.73, all metric results in a-cyclic preferences. We also replicate the behavior of the KS metric, where biases in the range of 0.73-0.85, show even some initial a-cyclic profiles can become cyclic. Figure 6.2 Further explains this by showing that within this range we always observe 3 unique profile for the KS metric, while DP and CS have already settled on 6 profiles, thereby representing all possible preferences. Figure 6.3 shows KS introduces ambiguity in the case that there was a Condorcet winner, resulting in losing the original nice profile. Finally, the proximity to single-peakedness shows a slightly more positive note for the KS metric, showing that while the DP and CS bottom out to the minimum proximity to single-peakedness, KS stays relatively close. Though this should be taken with a grain of salt, as it is likely a consequence of the unique preferences being smaller.

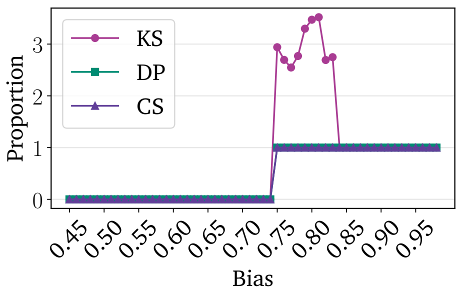

FIGURE 6.1: The proportion of cyclic profiles remaining, 0 indicating that no cyclic profiles were present after deliberation.

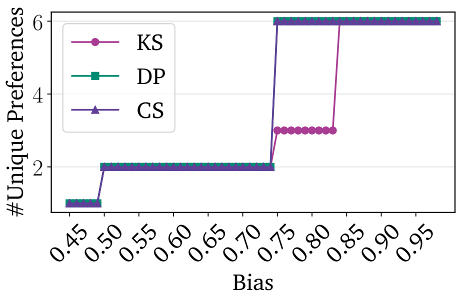

FIGURE 6.2: Number of unique preferences at the final step of deliberation.

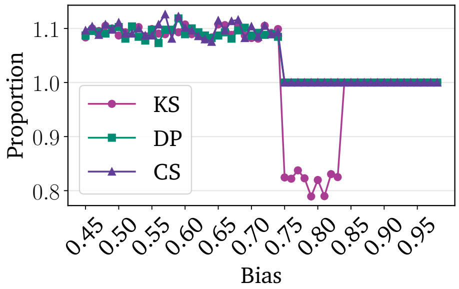

FIGURE 6.3: The proportion of Condorcet winners left after deliberation, value above one indicate Condorcet winners emerging during deliberation

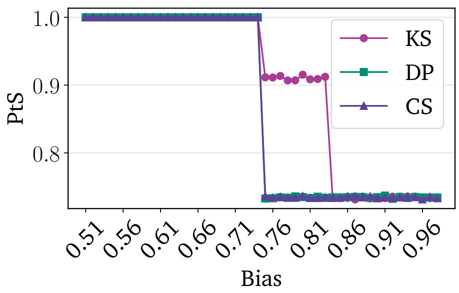

FIGURE 6.4: Proximity to single-peakedness after deliberation. Proximity to single-peakedness as defined in Section 3.3.

CHAPTER 7

DISCUSSION

CHAPTER 8

Conclusion

CHAPTER 9

Ethics and Data Management

A new requirement for the thesis is that there must be a short section in which you reflect on the ethical aspects of your project. This requirement is related to one of the final objectives that a graduated student of the Master of Computational Science must meet: “The graduate of the program has insight into the social significance of Computational Science and the responsibilities of experts in this field within science and in society". You don’t need to devote an entire chapter to this; a short section or paragraph is sufficient.

I acknowledge that the thesis adheres to the ethical code (https://student.uva.nl/en/topics/ethics-in-research) and research data management policies (https://rdm.uva.nl/en) of UvA and IvI.

The following table lists the data used in this thesis (including source codes). I confirm that the list is complete and the listed data are sufficient to reproduce the results of the thesis. If a prohibitive non-disclosure agreement is in effect at the time of submission “NDA” is written under “Availability” and “License” for the concerned data items.

Short description (max. 10 words)	**Availability (e.g., URL, DOI)**	**License (e.g., MIT, GPL, CC)**
Example dataset 1	<github url>or Figshare	GPL
Example source code	DOI (from Zenodo)	MIT
Example sensitive data	NDA	NDA

APPENDIX A

EXTENDED PROOFS

Finally, for CS, R_1 and R_j stay the same, while $R_1^{'} = c \succ a \succ b \succ \cdots \succ m$, resulting in $\mathrm{Dist}_{\mathrm{CS}}(R_1^{'}, R_j) = |2-2| + |1-3| + |3-1| = 4$.

BIBLIOGRAPHY

[1] Duncan Black. On the Rationale of Group Decision-making. *Journal of Political Economy*, 56(1):23–34, February 1948. ISSN 0022-3808. doi: 10.1086/256633.

[2] Christian List. Two Concepts of Agreement. *The Good Society*, 11(1):72–79, 2002. ISSN 1538-9731.

[3] Allan Gibbard. Manipulation of Voting Schemes: A General Result. *Econometrica*, 41(4):587–601, 1973. ISSN 0012-9682. doi: 10.2307/1914083.

[4] Mark Allen Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, April 1975. ISSN 0022-0531. doi: 10.1016/0022-0531(75)90050-2.

[5] Edith Elkind, Martin Lackner, and Dominik Peters. Preference Restrictions in Computational Social Choice: A Survey, May 2022.

[6] Wulf Gaertner. Domain restrictions. In *Handbook of Social Choice and Welfare*, volume 1 of *Handbook of Social Choice and Welfare*, pages 131–170. Elsevier, January 2002. doi: 10.1016/S1574-0110(02)80007-8.

[7] Donald E. Campbell and Jerry S. Kelly. Non-monotonicity does not imply the no-show paradox. *Social Choice and Welfare*, 19(3):513–515, 2002. ISSN 0176-1714.

[8] Donald E. Campbell and Jerry S. Kelly. Correction to "A Strategy-proofness Characterization of Majority Rule". *Economic Theory Bulletin*, 4(1):121–124, April 2016. ISSN 2196-1093. doi: 10.1007/s40505-015-0066-8.

[9] Donald E. Campbell and Jerry S. Kelly. Anonymous, neutral, and strategy-proof rules on the Condorcet domain. *Economics Letters*, 128:79–82, March 2015. ISSN 0165-1765. doi: 10.1016/j.econlet.2015.01.009.

[10] Samuel Freeman. Deliberative Democracy: A Sympathetic Comment. *Philosophy & Public Affairs*, 29(4):371–418, 2000. ISSN 1088-4963. doi: 10.1111/j.1088-4963.2000.00371.x.

[11] Joshua Cohen. Deliberation and Democratic Legimitimacy. In *Debates in Contemporary Political Philosophy*. Routledge, 2002. ISBN 978-0-203-98682-0.

[12] Jon Elster. The market and the forum: Three varieties of political theory. In *Debates in Contemporary Political Philosophy*. Routledge, 2002. ISBN 978-0-203-98682-0.

[13] Christian List, Robert C. Luskin, James S. Fishkin, and Iain McLean. Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls. *The Journal of Politics*, 75(1):80–95, January 2013. ISSN 0022-3816, 1468-2508. doi: 10.1017/S0022381612000886.

[14] Valeria Ottonelli and Daniele Porello. On the elusive notion of meta-agreement. *Politics, Philosophy & Economics*, 12(1):68–92, February 2013. ISSN 1470-594X. doi: 10.1177/1470594X11433742.

[15] Soroush Rafiee Rad and Olivier Roy. Deliberation, Single-Peakedness, and Coherent Aggregation. *American Political Science Review*, 115(2):629–648, May 2021. ISSN 0003-0554, 1537-5943. doi: 10.1017/S0003055420001045.

[16] John G Kemeny and James L Snell. Preference ranking: An axiomatic approach. *Mathematical models in the social sciences*, pages 9–23, 1962.

[17] Conal Duddy and Ashley Piggins. A measure of distance between judgment sets. *Social Choice and Welfare*, 39(4):855–867, 2012. ISSN 0176-1714.

[18] Wade D. Cook and Lawrence M. Seiford. Priority Ranking and Consensus Formation. *Management Science*, 24(16):1721–1732, December 1978. ISSN 0025-1909. doi: 10.1287/mnsc.24.16.1721.

[19] Benjamin Golub and Matthew O. Jackson. Naïve Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics*, 2(1):112–149, February 2010. ISSN 1945-7669. doi: 10.1257/mic.2.1.112.