# Iterative Goal-Based Voting

**MSc Thesis** *(Afstudeerscriptie)*

written by

**Leyla Ade**
(born May 11th, 1995 in Leimen, Germany)

under the supervision of **Dr. Arianna Novaro** and **Prof. Dr. Ulle Endriss**, and submitted to the Examinations Board in partial fulfillment of the requirements for the degree of

**MSc in Logic**

at the *Universiteit van Amsterdam*.

**Date of the public defense:**
*July 12th, 2021*

**Members of the Thesis Committee:**
Prof. Dr. Yde Venema *(chair)*
Prof. Dr. Ulle Endriss
Dr. Davide Grossi
Dr. Ronald de Haan
Dr. Arianna Novaro
Dr. Zoi Terzopoulou

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

# Abstract

Goal-based voting is a new voting framework in which agents can submit propositional formulae as their goals. We study iterated applications of the majorities and approval rules in this framework. We introduce notions of satisfaction based on the Hamming distance between an agent's goal and the interpretations in the outcome under a given rule. The contribution of this thesis is twofold: First, we analyze the convergence of the iteration. We show that the Majority rules and the Approval rule for some satisfaction functions are not guaranteed to terminate, while other cases of Approval voting do always converge. Second, we study the quality of iteration. The first part of this analysis consists of theoretical results, showing that in cases where termination of Approval voting is guaranteed we also have an improvement of the social welfare. The second part consists of an implementation of the iterative process in Python for the cases not covered by our theoretical results, which gives us preliminary insights on the frequency and quality of iteration.

# Acknowledgements

First and foremost, I want to thank my supervisors Ulle and Arianna for supporting me in realizing this thesis. Arianna, I cannot count the number of Zoom calls that we had over the past months. Thank you, for every hour spent on proof-reading, brain-storming, researching and simply chatting with me. Thank you, Ulle, for taking time to give me honest feedback and for widening my view on the topic. I was fortunate to call two such intelligent, kind and passionate people my supervisors.

Thank you, to all members of my thesis committee, Yde Venema, Davide Grossi, Ronald de Haan and Zoi Terzopoulou, for taking time to read my thesis and participate in my defence.

I also want to thank Tanja and Benedikt, for guiding me through the Masters program, always being available to advise me on practical matters.

Further, a special thanks goes to Giovanni, Arthur, Max and Oliviero for answering all of my programming questions.

Thank you to my Family for never doubting that everything will turn out fine, even when I had doubts of my own. I especially want to thank my sister, Eda, you are always on my side.

Without my friends I could have never finished this project. Special thanks to Matteo, you were my companion in this Masters from day 0, I will not forget about our promises. Thank you to Nicolien and Daniela for always being available for deep and not-so-deep talks and for creating beautiful memories. Likewise, I want to thank Antonio, you were always available for any LaTeX question and I enjoyed our time as flat-mates. Thank you, Alyssa and Fiona, for proofreading parts of this work.

I am thankful for my church family from the ICF Amsterdam, you filled my Sundays with joy and constantly reminded me that there is more to life than work.

Last, but first in my heart, I want to thank Marcel. Without you I could have neither started nor finished this Masters. My thesis marks the finish of one big project in my life, but there are many more to come in ours. I could not be more blessed as to soon call you my husband. I love you.

# Contents

# 1 | Introduction

Making decisions can be as easy as choosing a banana rather than an apple in the morning or as hard as deciding which PhD offer to take. Every one of us has probably dealt with some harder decisions before, collecting all the evidence needed to understand the situation, imagining all the possible effects the result could have, trying to understand one's emotions towards the alternatives. While taking a decision alone can be challenging, letting a group decide on a given issue at stake might lead to even more complicated situations. What to do if there is no objective right or wrong and the opinions of a group clash? How should we aggregate diverse preferences in order to reach a good result? What does it even mean for a result to be good? Such questions, and many more, are studied within the field of social choice theory.

While the question of how to make collective decision has been studied since ancient times, Brandt et al. (2016) name Arrow's Theorem as the starting point of modern social choice theory. Arrow's Theorem (Arrow, 1951) states that any voting rule satisfying a certain set of reasonable properties is a dictatorship. Following this result, many such impossibility theorems as well as other characterizations of rules followed and established the field of social choice theory as we know it today. Computational Social Choice (ComSoc) builds on to this rich history and is additionally interested in the computational aspects of collective decision-making. Allocation algorithms, matching mechanisms, computational complexity results of related decision problems as well as more classical questions like fairness or proportionality in voting are some of the topics which ComSoc researchers are interested in.

One framework that recently won attention is judgment aggregation (Dietrich, 2007; Endriss, 2016; List and Puppe, 2009). In particular, in binary aggregation framework (Grandi and Endriss, 2011), the agents accept or reject binary issues in a given agenda. Based on these votes, a given judgment aggregation rule will output for each issue whether it should be accepted or not.

This thesis will analyze the iterated process of a fairly new framework, goal-based voting, which was first introduced by Novaro et al. (2018). In the current chapter we will motivate iterative goal-based voting by giving a first example, give an overview on the chapters to come and discussing related work.

Goal-based voting was first introduced by Novaro et al. (2018) as a multi-issue decision making process in which agents can hold goals over the issues in a given agenda. Similar to judgment aggregation (List and Puppe, 2009) or belief merging (Konieczny and Pérez, 2011; Everaere et al., 2015), agents formulate their preferences as a formula in propositional logic in which each variable represents one of the issues (for an introduction to propositional logic see Gamut (1991)). In classical judgment aggregation, agents hold complete conjunctions: hence, they either support an issue or reject it. In some extended settings, agents might also be able to abstain from an issue which means that their ballot consists of an incomplete conjunction. Goal-based voting allows more complex connections between issues (as for example implications) which lets the agents express richer preferences than in judgment aggregation.

In belief merging, agents' knowledge or beliefs can be as complex as goals in goal-based voting, but since its aim is to merge propositions that are held by the agents (or sources) into a unique belief or knowledge base, rather than narrowing down the options in a collective decision, the focus is not on resoluteness.

The following example shows a situation where goal-based voting can be used, while judgment aggregation would not be able to capture these complex preferences.

**Example 1.1.** *Three friends, Ann, Betti and Clara, are searching for a shared flat. They already settled for a certain budget, but they still have to decide on whether to have a balcony (or a terrace, in case it is on the ground floor), whether they want to live on the ground floor and whether they want to have a shared living area. These open questions are binary issues and hence this decision can be modelled in goal-based voting. Assume the three friends have the following preferences:*

- *Ann wants to have a balcony in case they are not living at the ground floor. She does not like having a shared living area.*
- *Betti wants a balcony only in case there is no shared living area: she really likes to have a room where they can spend time together. She does not care about living at the ground floor.*
- *Clara does not want to live on the ground floor. She loves spending time in common areas and would therefore love a shared living space as well as a balcony.*

*Assigning propositional variables $b$, $f$ and $\ell$ to the issues balcony, ground floor and living area respectively, we can express the friends' preferences as shown in Table 1.1. In the* *models, $b$, $f$ and $\ell$ are the first, second and third issue respectively. For example in the model $(101)$ $b$ and $\ell$ are true, while $f$ is false, hence it represents a flat with a balcony and a shared living space, which is not on the ground floor. If these friends were to decide on*

| Agent | Goal | Models |
|---|---|---|
| Ann | $(\neg f \to b) \wedge \neg \ell$ | $\{(100), (110), (010)\}$ |
| Betti | $b \leftrightarrow \neg \ell$ | $\{(110), (100), (011), (001)\}$ |
| Clara | $b \wedge \neg f \wedge \ell$ | $\{(101)\}$ |

Table 1.1: The goals of the three agents in Example 1.1. The first column represents the agents' names, the second their goals and the third colums the models of their goals.

*these issues by using judgment aggregation, then only Clara would be able to express her gaol, since Betti's goal, for example, has multiple models and a rational judgment set (in judgment aggregation) could represent only one of them.*

Novaro et al. (2018) introduce different voting rules with which these goals could be aggregated. For example the $Approval$ rule outputs all the interpretations which have the maximal support from the agents in the profile: in Example 1.1 the outcome would thus be $\{(110), (100)\}$. While in this example the $Approval$ rule narrows the decision, not being resolute (meaning that it sometimes can return multiple tied alternatives as the outcome) is indeed a downside of the $Approval$ rule. Among the resolute rules, the $EMaj$ has been introduces as (one possible) generalization to goal-based voting of classical majority. According to this rule every voter and every model of their goal are given equal weight and an issue is accepted exactly if the support received by the agents in the profile is more than half the number of agents. In Example 1.1 the $EMaj$ would, for instance, lead to $\{(011)\}$ as the outcome, i.e., a flat which has no balcony, is on the ground floor and has a common living area.

One might observe that the $Approval$ results in the set $\{(110), (100)\}$: these interpretations are models of Ann's and Betti's goals, but not of Clara's. Thus, she might have an incentive to alter her vote such that she reaches a more favourable outcome. How to assess whether one outcome is more favourable than another is not a simple question. Satisfaction functions are used to model the satisfaction of an agent for the current outcome, given their truthful goal.

In this thesis we will work with two types of satisfaction functions: a dichotomous type, where an agent's satisfaction is solely based on whether a model of their goal is in the outcome or not, and then functions based on the Hamming distance of their truthful goal to the interpretations in the outcome. Considering the Hamming distance, one might see that in Example 1.1, given the two interpretations returned by the $Approval$ rule, Clara would prefer $(100)$ over $(110)$, since the former differs only on one issue to her goal while the latter does on two. Therefore, Clara might have an incentive to alter her vote to $b \land \neg f \land \neg \ell$, so that the new outcome yielded by the $Approval$ rule is $\{(100)\}$ and hence closer to her truthful goal. Assuming that these friends use an online tool to aggregate their preferences, in which anyone can come back, observe the vote of every other agent and change her preference, it is reasonable to believe that this would actually takes place.

Iterative voting studies exactly these scenarios. While in classical voting the agents are assumed to vote only once, in iterative voting agents can revisit and change their vote after observing the current result. This iterative process yields many other questions: Will this voting process terminate? What gives an agent an incentive to change her vote? How big is the computational effort to calculate a better vote? Is the final result better than the initial one? The answers to all these questions depend highly on the set-up of the iteration.

In fact, the underlying rule as well as the rationality of the agents or restrictions on the alterations are important variables. In classic approaches, agents are assumed to be rational, which means that they only alter if it is in some sense *better*, and that they are myopic, i.e., when choosing how to alter their vote, they only take into account the next outcome and not any further iteration. Most often, if there are multiple agents that could possibly improve the result in their favour, the altering agent is chosen randomly.

Reasonable voting rules have been shown to possibly yield circular iterations, this means the agents' end up voting for a profile they have been voting for before, hence possibly voting in a circle. In judgment aggregation, Terzopoulou and Endriss (2018) were able to show that the plurality rule does not always terminate. Novaro (2019) shows that in goal-based voting the Majority rules are in general manipulable, opening the question about iterative voting.

In this work we will be laying the ground work for analyzing iterative goal-based voting. Focusing on the $EMaj$, $TrueMaj$, $2sMaj$ and $Approval$ rules, as defined by Novaro et al. (2018), the aim of this study is to open the question of iteration to this new voting framework. We will be analyzing the termination of these rules as well as the quality of the iteration. We will also implement a program in which randomly chosen profiles are iterated, to get a first impression on how likely and realistic the theoretical results are.

## 1.1 Related Work

For an overview to the field of Computational Social Choice, we recommend the Handbook by Brandt et al. (2016). Relevant to this thesis are especially the introduction to strategy-proofness (Chapter 17) and the discussion on the approval rule (in Chapter 2).

Novaro et al. (2018) introduced goal-based voting: in particular, they defined the $Approval$ and Majority rules which we study in this thesis. Novaro (2019) studied axiomatic analysis of these voting rules and their computational complexity and initial results on the manipulability of the Majority rules, showing these to be manipulable. The manipulability results naturally open the possibility to iterative voting, which was left as a question for future work and which motivates the study in this thesis.

The work closest to ours is the study of iteration in judgment aggregation by Terzopoulou and Endriss (2018), who analyzed the premise-based and plurality rules. They found the former to always terminate in an iterative process, while the latter only does so under certain restrictions. Some of these restrictions limit the knowledge the agents have over the other voters' ballots. Some of the results regarding the plurality rule can be transferred to our setting (see Section 3.3). We direct the reader to Baumeister et al. (2017) for an overview about strategic behaviour in judgment aggregation.

Another related field is that of belief merging, Everaere et al. (2017) compare judgment aggregation with belief merging by defining a projection function between belief bases and agendas, highlighting the similarities and differences between these two research fields. An introduction to belief merging and an overview of the integrity constraints that are commonly used are described by Konieczny and Pérez (2002). Delgrande et al. (2006) define a framework for iterated belief revision which differs from our approach in one key aspect: while we assume agents to keep a truthful goal during the iteration process, trying to optimize according to it during the iteration, in iterated belief revision the agents' change their beliefs, knowledge or opinions through the iteration.

Iterative voting has been a popular research direction in Computational Social Choice. Meir et al. (2010), Reyhani and Wilson (2012) and Brânzei et al. (2013) studied the convergence and quality of iterative voting for different rules and settings. Meir et al. (2017) give an introduction to iterative voting: most classic notions, as being myopic, truth-biased or giving a best response, are explained there. Meir et al. (2010) analyze under which conditions the plurality rule converges to an equilibrium in a game-theoretic framework, showing that it depends on various assumptions (such as, for example, the chosen tie-breaking rule). A similar analysis on the convergence of scoring rules was done by Reyhani and Wilson (2012), who showed that the veto rule is guaranteed to terminate and established a new upper bound on the steps to convergence with the plurality rule. Their results also include examples of circular iterations for other scoring rules, such as the Borda rule. Obraztsova et al. (2015) provide two sufficient conditions for convergence in iterative voting and study whether some classes of voting rules, such as scoring rules, satisfy them.

Beside the convergence of iterative processes, the analysis of their quality also received great attention in the literature. Brânzei et al. (2013) studied the price of anarchy for some scoring rules, i.e., the ratio between the best possible outcome and the worst possible outcome with iteration. They show that the plurality rule can benefit the voters in iteration under this notion. However, they also provide some negative results (for example for the Borda rule).

Finally, a key element of defining iterative goal-based voting is the choice of the satisfaction function for the agents, i.e., defining when an agent has an incentive to possibly alter her vote. While Novaro (2019) defined dichotomous preferences, based on the models and non-models of the agents' goals, we use other classical approaches of preference extensions to define suitable satisfaction functions from the Hamming distance. Our extensions are based on MiniMax and MaxiMax by Packard (1979), who used those to extend preferences on elements to preferences over sets of these elements. More work on this topic has been done by Fishburn (1972) who defined the homonymous extension, in which ties are assumed to be broken by an even-chance lottery. More extensions and their axiomatic characterizations can also be found in Kelly (1977), Pattanaik and Peleg (1984) and Bossert et al. (2000).

## 1.2 Overview

In **Chapter 2** we define goal-based voting following Novaro et al. (2018), also giving some examples and in particular, the $\mathit{EMaj}$, $\mathit{TrueMaj}$, $\mathit{2sMaj}$ and $\mathit{Approval}$ rules. A main part of the chapter will be about defining satisfaction functions, which lay the ground of an agent's alteration. While still including the dichotomous functions defined by Novaro (2019) we will expand the notion of satisfaction to an approach based on the Hamming distance, leading to three types of agents, which we will call Hamming-optimists, Hamming-pessimists and Hamming-realists. The last section of this chapter summarizes and completes the results about manipulation for the rules considered in this work.

**Chapter 3** contains the main theoretical results of this work, in which the termination of the $\mathit{EMaj}$, $\mathit{TrueMaj}$, $\mathit{2sMaj}$ and $\mathit{Approval}$ rules are analyzed. The first section describes the iteration process: in particular, we define best responses, myopic agents, truth-bias and rationality. Then we show that the Majority rules are not guaranteed to terminate if the agents are truth-biased, and that *EMaj* and *TrueMaj* might still yield circular iterations even without truth-bias. We also prove that whether the iterated *Approval* rule terminates or not highly depends on which satisfaction function the agents use. We establish that iteration always terminates with Hamming-pessimists, while some restrictions on the alterations are needed to achieve termination with H-optimists. For Hamming-realists, however, these restrictions are not strong enough to guarantee termination with the *Approval* rule.

In **Chapter** 4, we will introduce the notion of social welfare, based on which we discuss the quality of the iteration. For the *EMaj*, *TrueMaj* and *2sMaj* rules we show that iteration is not guaranteed to yield a higher social welfare. For the *Approval* rule we demonstrate that certain satisfaction functions improve the social welfare in iteration. Then, we present our implementation of iterated *Approval* voting for three agents and three issues, which leads us to formulate the hypothesis that iteration is rare and beneficial.

In the conclusion all results are summed up, discussed and further directions are pointed out.

# 2 | Preliminaries

In this chapter, we will build the basic grounds needed to understand the following chapters and results. We will first define the formal model of goal-based voting, followed by a section discussing strategic voting in this setting. Since strategic behaviour of agents is defined with respect to a given satsifaction function we will introduce in the second part of the chapter the ones we will be studying. This includes satisfaction functions based on the Hamming distance. Lastly, we will summarize and complete the existing results of manipulation in goal-based voting. For a more detailed approach on strategic goal-based voting and further examples, we refer to Novaro (2019).

## 2.1 Goal-Based Voting

In goal-based voting a finite set of agents $\mathcal{N} = \{1, \ldots, n\}$, which will also be called voters, take part in a collective decision on some binary issues from a finite set $\mathcal{I} = \{1, \ldots, m\}$. The vote of agent $i$ is expressed as a propositional formula $\gamma_i$, which is consistent and in which each propositional variable corresponds to one of the issues in $\mathcal{I}$. The set of models of formula $\gamma_i$ will be denoted by $Mod(\gamma_i)$. These models $v \in Mod(\gamma_i)$ will be represented as vectors $(v(1), \ldots, v(m))$ such that $v(j) \in \{0, 1\}$ for any $j \in \mathcal{I}$ describes whether the issue $j$ is true $(1)$ or false $(0)$ in the model. A voting profile $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_n)$ is a vector of every agents' goal. Let $\mathcal{G}$ be the set of possible goals, i.e., all propositional formulae on the variables in $\mathcal{I}$.

Then, a voting rule is a function for any $n$ and $m$

$$F : (\mathcal{G})^n \to \mathcal{P}(\{0, 1\}^m) \setminus \{\emptyset\}$$

mapping voting profiles to a non empty subset of all interpretations. A rule is called *resolute*, if it returns a singleton on every profile. Since we also consider irresolute rules, the co-domain of a rule is defined to be the powerset of all vectors. For a resolute rule $F$, $F(\mathbf{\Gamma})$ is a vector in $\{0, 1\}^m$ and thus, $F(\mathbf{\Gamma})_j$ denotes the $j$th entry of $F(\mathbf{\Gamma})$.

The following rules where introduced by Novaro et al. (2018). The $Approval$ rule coincides with the $\Delta_{\mu}^{\Sigma,d}$-rule in the framework of belief merging by Konieczny and Pérez (2011).

**Definition 2.1** ($Approval$)**.** *For a profile $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_n)$ of $n$ agents and $m$ issues, the $Approval$ rule is defined as:*

$$Approval(\mathbf{\Gamma}) = \mathrm{argmax}_{v \in Mod(\bigvee_{i \in \mathcal{N}} \gamma_i)} |\{i \in \mathcal{N} | v \in Mod(\gamma_i)\}|.$$

One can also define the $Approval$ rule as maximizing the *support* of a model in the given profile. The concept of support will be useful in the following chapters.

**Definition 2.2** (Support)**.** *For a profile $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_n)$ and an interpretation $v \in \{0,1\}^m$ we call the number of agents voting for a goal satisfied by $v$ the support of $v$, denoted by $supp_{\mathbf{\Gamma}}(v) := |\{i \in \mathcal{N} \mid v \in Mod(\gamma_i)\}|$.*

So then $Approval$ can be restated as $Approval(\mathbf{\Gamma}) = \mathrm{argmax}_{v \in Mod(\bigvee_{i \in \mathcal{N}} \gamma_i)} supp_{\mathbf{\Gamma}}(v)$.

The next rules are generalizations of the well-known Majority rule in voting and judgment aggregation. These rules belong to the more general class of threshold rules, where a threshold of acceptance (in this case, $\frac{n}{2}$) is specified for each issue. We follow Novaro et al. (2018) in defining $m_{ij}^x := |\{v \in Mod(\gamma_i) \mid v(j) = x\}|$ as the number of models of agent $i$'s goal that assigns to issue $j \in \mathcal{I}$ the value $x \in \{0,1\}$. The $EMaj$ rule is a resolute voting rule which an issue is accepted, if and only if it surpasses the threshold of $\frac{n}{2}$ votes, where each agent and each model for an agent's goal have equal voting weight. The following definition describes the rule formally.

**Definition 2.3** ($EMaj$)**.** *For a profile $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_n)$ with $n$ agents and a finite set of $m$ issues, the $EMaj$ rule is defined as:*

$$EMaj(\mathbf{\Gamma})_j = 1 \quad \textit{iff} \quad \sum_{i \in \mathcal{N}} \frac{m_{ij}^1}{|Mod(\gamma_i)|} > \frac{n}{2}.$$

Note that the $EMaj$ rule is resolute by definition, while $Approval$ is not. Another non-resolute rule is $TrueMaj$, which requires us to define a Majority rule $M(\mathbf{\Gamma})_j$ for each issue $j \in \mathcal{I}$:

$$M(\mathbf{\Gamma})_j = \begin{cases} \{x\} & \text{if } \sum_{i \in \mathcal{N}} \frac{m_{ij}^x}{|Mod(\gamma_i)|} > \frac{n}{2} \\ \{0,1\} & \text{if } \sum_{i \in \mathcal{N}} \frac{m_{ij}^1}{|Mod(\gamma_i)|} = \frac{n}{2} \end{cases}$$

We can then define *TrueMaj* as a Cartesian product of the majority outcomes over all issues. That is, *TrueMaj* coincides with the *EMaj* in case the majorities are strict for each issue. For every issue $j$ on which the majority is not strict, the rule returns in the outcome one interpretation where $j$ has value $1$, and one where it has value $0$.

**Definition 2.4** ($\mathit{TrueMaj}$)**.** *For a profile $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_n)$ with $n$ agents and $m$ issues, the $\mathit{TrueMaj}$ is defined as:*

$$\mathit{TrueMaj}(\mathbf{\Gamma}) = \prod_{j \in \mathcal{I}} M(\mathbf{\Gamma})_j.$$

The $\mathit{2sMaj}$ first computes the $\mathit{EMaj}$ over the set of each individual agent's goal's models to then apply a majority function on these outcomes. We will thus first define this majority function. Given a vector of vectors $(v_1, \ldots, v_n)$ we can define a majority function such that

$$\mathit{Maj}(v_1, \ldots, v_n)_j = \begin{cases} 1 & \text{if} \sum\limits_{i \in \{1,\ldots,n\}} v_i(j) > \frac{n}{2} \\ 0 & \text{otherwise} \end{cases}$$

The $\mathit{2sMaj}$ is thus defined as follows:

**Definition 2.5** ($\mathit{2sMaj}$)**.** *For a profile $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_n)$ with $n$ agents and $m$ issues, the $\mathit{2sMaj}$ is defined as:*

$$\mathit{2sMaj}(\mathbf{\Gamma}) = \{\mathit{Maj}(\mathit{EMaj}(\gamma_1), \ldots, \mathit{EMaj}(\gamma_n))\}.$$

We can now properly formulate our first example from the introduction.

**Example 2.1.** *Ann, Betti and Clara are looking for an apartment. Remember that they had three issues to decide on, the balcony $(b)$, the ground floor $(f)$ and whether or not to have a shared living area $(\ell)$. Thus, we have a set of three agents $\mathcal{N} = \{1, 2, 3\}$ and a set of three issues $\mathcal{I} = \{b, f, \ell\}$, we use $b$ to denote the first issue, $f$ for the second and $\ell$ as the third instead of $1$, $2$ and $3$ for ease of notation. Therefore the voting profile $\mathbf{\Gamma} = (\gamma_1, \gamma_2, \gamma_3)$ is such that:*

$$\gamma_1 = (\neg f \to b) \land \neg \ell, \quad \gamma_2 = b \leftrightarrow \neg \ell \quad \textit{and} \quad \gamma_3 = b \land \neg f \land \ell.$$

*The models of agent $1$'s goal are $\mathit{Mod}(\gamma_1) = \{(100), (110), (010)\}$, agent $2$'s goal is modelled by $\mathit{Mod}(\gamma_2) = \{(110), (100), (011), (001)\}$ and $\mathit{Mod}(\gamma_3) = \{(101)\}$.*

*Let us calculate the outcome of our rules defined above. First, the Approval rule returns the interpretation with the maximal support. Since none of the interpretations model all agents' goals, the maximal support here is 2. We thus have:* $Approval(\mathbf{\Gamma}) = \{(110), (100)\}$. *For* $EMaj$ *and* $TrueMaj$ *it is enough to calculate* $\sum_{i \in \mathcal{N}} \frac{m^1_{ij}}{|Mod(\gamma_i)|}$ *for each issue* $j \in \mathcal{I}$. *Note that* $\frac{n}{2} = \frac{3}{2}$, *so we get:*

$$\sum_{i \in \mathcal{N}} \frac{m^1_{ib}}{|Mod(\gamma_i)|} = \frac{2}{3} + \frac{1}{2} + 1 = \frac{13}{6} > \frac{3}{2}$$
$$\sum_{i \in \mathcal{N}} \frac{m^1_{if}}{|Mod(\gamma_i)|} = \frac{2}{3} + \frac{1}{2} + 0 = \frac{7}{6} < \frac{3}{2}$$
$$\sum_{i \in \mathcal{N}} \frac{m^1_{i\ell}}{|Mod(\gamma_i)|} = 0 + \frac{1}{2} + 1 = \frac{3}{2} = \frac{3}{2}$$

*Since the first issue got a majority, the second did not and the third tied, and since* $EMaj$ *breaks ties against an issue, while* $TrueMaj$ *keeps both variations, they result in* $EMaj(\mathbf{\Gamma}) = \{(100)\}$ *and* $TrueMaj(\mathbf{\Gamma}) = \{(100), (101)\}$. *Applying the* $EMaj$ *first to each agent's goals, results in:*

$$EMaj(\gamma_1) = \{(110)\} \quad EMaj(\gamma_2) = \{(000)\} \quad EMaj(\gamma_3) = \{(101)\}.$$

*Then, applying* $Maj$, *the* $2sMaj$ *returns* $\{(100)\}$.

In further sections we will skip these calculations, it is left to the reader to check the results. From now on we will represent the profiles and results as in Table 2.1. We hope to increase readability in this way, since, even though our voting rules take the tuple of goals as their input, they actually compute the result based on the goals' models.

If the friends would choose to use $EMaj$ or $2sMaj$, they would end up searching for a flat which has a balcony, is not on the ground floor and has no shared living area. If they would have chosen the $Approval$ or $TrueMaj$ rule, they would have not gotten a definite decision yet. The former leaves them to choose between a flat with a balcony, not on the ground floor, with no shared living area and an apartment with a terrace on the ground floor and no shared living area. The $TrueMaj$ follows the $EMaj$ in the first two issues but leaves the decision on whether to have a shared living area open. While in this example $EMaj$ and $2sMaj$ coincide, Novaro (2019) has shown that this is not always the case.

| | $\mathbf{\Gamma}$ |
|---|---|
| $Mod(\gamma_1)$ | (100) |
| | (110) |
| | (010) |
| $Mod(\gamma_2)$ | (110) |
| | (100) |
| | (011) |
| | (001) |
| $Mod(\gamma_3)$ | (101) |
| *Approval* | (110) |
| | (100) |
| *EMaj*/ *2sMaj* | (100) |
| *TrueMaj* | (100) |
| | (101) |

Table 2.1: Voting rules applied to the profile in Example 1.1.

## 2.2 Strategic Goal-Based Voting

A *satisfaction function* measures for each agent how content they are with the result of a collective decision. Formally, it maps the tuple consisting of the agent's goal and the outcome, i.e., a propositional formula and a set of models, to a real number.

$$sat : \mathcal{G} \times \mathcal{P}(\{0,1\}^m) \to \mathbb{R}.$$

Given an agent's goal, a set of interpretations is said to be preferred by an agent $i$ to another set of interpretations if the satisfaction for the former is higher than for the latter. In particular, given two profiles $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$, agent $i$ prefers the outcome $F(\mathbf{\Gamma})$ of a rule $F$ over the outcome $F(\mathbf{\Gamma}')$ if the following holds:

$$F(\mathbf{\Gamma}) \succ_i F(\mathbf{\Gamma}') \quad \text{iff} \quad sat(\gamma_i, F(\mathbf{\Gamma})) > sat(\gamma_i, F(\mathbf{\Gamma}')).$$

The following sections will introduce the examples of satisfaction functions which are relevant for our work.

### 2.2.1 Dichotomous Satisfaction Functions

The following satisfaction functions are dichotomous in the sense that the agents are only concerned about the interpretations of the outcome being a model of their goal or not. Novaro (2019) defined three different forms of this satisfaction. The *opt* function returns 1 if the intersection of an agent's goal and the outcome is not empty and 0 otherwise. If an agent wants all interpretations of the outcome to be a model of her goal, her satisfaction coincides with $pess$. Lastly, for $eum$ the proportion of the interpretations satisfying the agent's goal in the outcome corresponds to her satisfaction. These functions correspond to the index functions $i_{dw}$, $i_{ds}$ and $i_p$ used by Everaere et al. (2007) to study strategic behaviour in belief merging.

**Definition 2.6.** *For a profile $\mathbf{\Gamma} = (\gamma_1, \dots, \gamma_n)$ of $n$ agents and $m$ issues, and a voting rule $F$, the opt, $pess$ and $eum$ are defined as:*

- $opt(i, F(\mathbf{\Gamma})) = \begin{cases} 1 & \textit{if } F(\mathbf{\Gamma}) \cap Mod(\gamma_i) \neq \emptyset \\ 0 & \textit{otherwise} \end{cases}$
- $pess(i, F(\mathbf{\Gamma})) = \begin{cases} 1 & \textit{if } F(\mathbf{\Gamma}) \subseteq Mod(\gamma_i) \\ 0 & \textit{otherwise} \end{cases}$
- $eum(i, F(\mathbf{\Gamma})) = \frac{|Mod(\gamma_i) \cap F(\mathbf{\Gamma})|}{|F(\mathbf{\Gamma})|}$

The agents will be called optimists, pessimists and expected utility maximizers accordingly. While these functions induce a natural notion of satisfaction, it is not very fine-grained, as illustrated by the following observation.

Suppose none of the interpretations in the outcome is a model of an agent's goal: according to all of these three notions of satisfaction the agent will be unsatisfied. But we may want to be more specific in how unsatisfied the agent is with the resulting options in the outcome. For example, consider two issues $p, q$ and an agent whose goal is $p \wedge q$: the only model satisfying the agent's goal is $(11)$. Now suppose that, on a given profile, one voting rule outputs $(00)$ and another rule results in $(10)$. Do we really want to say that the agent is equally unhappy about those?

Intuitively, the Hamming distance counts the number of issues on which two given interpretations disagree. In the former example of issues $p$ and $q$, the Hamming distance from the agent's model to the first outcome would be $2$ while the distance to the second outcome is only $1$. Hence it gives us a way to distinguish satisfaction of non-models in more detail. In the next section, we will see satisfaction functions based on the Hamming distance, which capture this intuition.

### 2.2.2 Hamming Distance Extensions

The Hamming distance is often used to compare models and induce satisfaction (Dietrich and List, 2007b; Endriss et al., 2012) it is defined between two given vectors of a fixed length $m$.

**Definition 2.7** (Hamming Distance). *Given two interpretations $v, w \in \{1, 0\}^m$ and a set of issues $\mathcal{I} = \{1, \ldots, m\}$, the Hamming distance is defined as*

$$H(v, w) = |\{j \in \mathcal{I} \mid v(j) \neq w(j)\}|.$$

The Hamming distance is often used in judgement aggregation to define a preference, i.e., a linear order, over the possible outcomes (Baumeister et al., 2017). However, since an agent's goal might have multiple models and because we also consider irresolute voting rules, we need to extend the notion of preference to sets of models.

Kelly (1977) was one of the first to extend preferences over objects to preferences over sets of objects, introducing the so called *Kelly extension*. Intuitively speaking, one set is Kelly-preferred over another if every element in the former is at least weakly preferred over every element in the latter. However, there are many sets which are left incomparable by this extension. For example, if an agent prefers elements $a$ over $b$ over $c$ then the Kelly extension cannot tell the agent whether outcome $\{a, c\}$ is preferred to $\{b\}$ or vice-versa, given her initial preference over the elements.

Our approach is inspired by the *Maxi-Max* and *Maxi-Min* as well as the *Averaging* extension as introduced by Packard (1979). These extensions are total; hence any two given sets will be comparable. Given a certain preference over elements of a set, these approaches try to maximize the minimum, the maximum or the average satisfaction, with respect to a given order; see Definitions 2.8, 2.9 and 2.10.

**Definition 2.8** (Maxi-Max). *Given a preference order $\succ$ on a set $S$ and non-empty subsets $X, Y \subseteq S$, let $max(X)$ and $max(Y)$ be the maximal elements in $X$ and $Y$ according to $\succ$. The $MaxiMax$ extension is defined as:*

$$X \succ_{maximax} Y \textit{ iff } max(X) \succ max(Y).$$

**Definition 2.9** (Maxi-Min). *Given a preference order $\succ$ on a set $S$ and non-empty subsets $X, Y \subseteq S$, let $min(X)$ and $min(Y)$ be the minimal elements in $X$ and $Y$ according to $\succ$. The $MaxiMin$ extension is defined as:*

$$X \succ_{maximin} Y \textit{ iff } min(X) \succ min(Y).$$

**Definition 2.10** (Averaging)**.** *Given a set $S$, a function $u : S \to \mathbb{R}$ and non-empty subsets $X, Y \subseteq S$. The $Averaging$ extension is defined as:*

$$X \succ_{averaging} Y \textit{ iff } \sum_{x \in X} \frac{u(x)}{|X|} > \sum_{y \in Y} \frac{u(y)}{|Y|}.$$

The function $u$ can be seen as a utility function. We will be using the Hamming distance to implement these extensions as Barrot et al. (2017) did for Maxi-Max and Maxi-Min in a setting of approval voting for committee elections (corresponding to a complete conjunction in our setting). Following their example of calling the former *Optimistic* and the latter *Pessimistic*, we will introduce three types of agents: *Hamming-optimist*, *Hamming-pessimist* and *Hamming-realist*. Hamming-optimists and Hamming-pessimists will only consider the minimal and maximal Hamming distance of a set of interpretations to their goal, as per Definitions 2.8 and 2.9. The Hamming-realist will consider the average Hamming distance, such that the utility $u$ in Definition 2.10 corresponds to the negative of this distance (e.g. -2 if the distance is 2). Implementing these extensions in our setting is not straightforward.

We not only have multiple interpretations in the outcome, but an agent's goal might also have multiple models. Here, there is no unique preference order induced by the Hamming distance. Thus, there is not just one way to define the minimal, maximal or average distance of an outcome compared to the models of an agent's goal, as shown by the following example.

**Example 2.2.** *Assume agent $i$ has the goal $\gamma_i = (p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge r)$ for three given issues $\mathcal{I} = \{p, q, r\}$. Thus the models of her goal are $Mod(\gamma_i) = \{(100), (001)\}$. Now take $O = \{(111), (110)\}$ to be an outcome of some voting rule. When comparing the models of agent $i$'s goal to the elements of $O$, there are four Hamming distances we could consider: namely, the ones between each $v \in O$ and each $w \in Mod(\gamma_i)$. These distances are $H((111), (100)) = 2$, $H((111), (001)) = 2$, $H((110), (100)) = 1$ and $H((110), (001)) = 3$.*

Which one is the minimal or maximal element of $O$ according to $i$'s goal in Example 2.2? Given any interpretation in an outcome and its Hamming distance to each of the models of the agent's goal, we will always consider the lowest of these distances. That is, in Example 2.2 the interpretation $(111)$ has distance $2$ to both models of $i$'s goal, and thus the lowest distance is $2$; while the lowest distance of $(110)$ to the models of $i$'s goal is $1$. Thus we will consider $(110)$ to be the minimal element in $O$ according to $i$'s goal.

One could also choose the average over these distances to be the relevant factor of the comparison. These would be 2 for both interpretations in $O$ and hence $(111)$ and $(110)$ would be equally preferred. However, the following example motivates our choice.

**Example 2.3.** *Suppose agent $i$ has the goal $\gamma_i = p \wedge (q \leftrightarrow r)$ for three given issues $\mathcal{I} = \{p, q, r\}$, hence $Mod(\gamma_i) = \{(111), (100)\}$. Consider two outcomes of a voting rule, namely $O = \{(111)\}$ and $O' = \{(110)\}$. The interpretation in $O$ is a model of agent $i$'s goal, while the one in $O'$ is not. Looking at the Hamming distances, we get $H((111), (111)) = 0$ and $H((100), (111)) = 2$ as well as $H((111), (110)) = 1$ and $H((100), (110)) = 1$.*

Intuitively, we want an agent to be satisfied if $O \subseteq Mod(\gamma_i)$, since the agent's goal is guaranteed to be fulfilled in this case. The lowest distance as discussed above ensures this for $O$ in Example 2.3, since $H((111), (111)) = 0$. However, considering the average distance the agent in Example 2.3 would be indifferent between $O$ and $O'$, since both have an average distance of 1, even though the interpretation in $O'$ does not satisfy her goal. Following this argument we will define the *lowest Hamming distance*, which is analogous to the *minimal distance* as Everaere et al. (2007) use it in belief merging. Based on the lowest Hamming distance we will then define the satisfaction functions of Hamming-optimist, Hamming-pessimist and Hamming-realist, which will be denoted by H-optimist, H-pessimist and H-realist from now on.

**Definition 2.11** (Lowest Hamming Distance)**.** *For a given set of vectors $M \in \mathcal{P}(\{0,1\}^m)$ and a vector $w \in \{0,1\}^m$, the lowest Hamming distance is defined as the minimal Hamming distance the vector $w$ has to a model in $M$. Namely, we have*

$$lowH(M, w) = \min_{v \in \mathcal{M}} H(v, w).$$

Given an agent's goal $\gamma_i$, the lowest Hamming distance induces a preference order $\succ_i$ over all possible interpretations in an outcome. Namely,

$$v \succ_i w \quad \text{iff} \quad lowH(Mod(\gamma_i), v) < lowH(Mod(\gamma_i), w).$$

This weak linear order is total and can be extended to a preference over sets.

**Definition 2.12** (H-Optimist, H-Pessimist, H-Realist)**.** *Given a profile $\boldsymbol{\Gamma} = (\gamma_1, \ldots, \gamma_n)$ and a voting rule $F$, the H-optimist, H-pessimist and H-realist satisfaction of the agents are induced by the following distances:*

$$optH(\gamma_i, F(\boldsymbol{\Gamma})) = \min_{w \in F(\boldsymbol{\Gamma})} lowH(Mod(\gamma_i), w)$$

$$pessH(\gamma_i, F(\mathbf{\Gamma})) = \max_{w \in F(\mathbf{\Gamma})} lowH(Mod(\gamma_i), w)$$

$$realH(\gamma_i, F(\mathbf{\Gamma})) = \frac{\sum_{w \in F(\mathbf{\Gamma})} lowH(Mod(\gamma_i), w)}{|F(\mathbf{\Gamma})|}$$

*then for a given distance* $d_i \in \{optH, pessH, realH\}$ *the satisfaction is defined as* $sat_i\gamma_i, (F(\mathbf{\Gamma}')) = m - d_i$*.*

In the following we will write $lowH(\gamma_i, v)$ instead of $lowH(Mod(\gamma_i), v)$ for ease of notation and $optH(\gamma_i, F(\mathbf{\Gamma}))$, $pessH(\gamma_i, F(\mathbf{\Gamma})))$ as well as $realH(\gamma_i, F(\mathbf{\Gamma})))$. Since the distances $optH$, $pessH$ and $realH$ range between 0 and $m$, the former marking a more satisfied agent, the satisfaction $sat_i$ ranges between $m$ and 0, respectively.

These preference is an instance of Definition 2.9, where the interpretation in the outcome holding the distance $optH$, i.e., $\text{argmin}_{w \in F(\mathbf{\Gamma})} lowH(\gamma_i, w)$, is the minimal element. Similarly $\text{argmax}_{w \in F(\mathbf{\Gamma})} lowH(\gamma_i, w)$ is the maximal element in line with Definition 2.8. The preference of an H-realist coincides with Definition 2.10 when $sat_i$ is taken to be the utility function.

Barrot et al. (2017) used a similar approach when discussing *Approval* voting for committee elections. They also assumed satisfaction to be based on Hamming distances and defined the utility of one agent as the difference between the number of elected committee members and the distance to those. We follow this approach by defining the satisfaction of an agent as the difference between the number of issues and her distance to a given outcome under some rule.

Note that, if her satisfaction is induced by the distance in Definition 2.12, an agent $i$ will *(strictly) prefer* the outcome of a voting rule $F$ on a profile $\mathbf{\Gamma}'$ over the outcome of $F$ on the profile $\mathbf{\Gamma}$ if and only if according to her type the distance to the former is (strictly) smaller than the distance to the latter. For an H-optimist $i$ for example this means: $F(\mathbf{\Gamma}') \succ_i F(\mathbf{\Gamma})$ iff $optH(\gamma_i, F(\mathbf{\Gamma}')) < optH(\gamma_i, F(\mathbf{\Gamma}))$. The following example will clarify these definitions.

**Example 2.4.** *Consider three friends wanting to go out for dinner. They need to decide if they want to have an appetizer (a), main dish (m) and dessert (d) at their night out. The first friend wants to either have all or not go out at all, that is* $\gamma_1 = a \leftrightarrow m \leftrightarrow d$*. The second friend is on a diet and would prefer not to have a dessert,* $\gamma_2 = a \wedge m \wedge \neg d$*. The third wants the whole experience and we therefore get* $\gamma_3 = a \wedge m \wedge d$*. Assume these friends use the* $TrueMaj$ *rule to narrow down their decision. We get the profile as presented in Table 2.2.*

| | |
|---|---|
| $Mod(\gamma_1)$ | (000) |
| | (111) |
| $Mod(\gamma_2)$ | (110) |
| $Mod(\gamma_3)$ | (111) |
| $TrueMaj(\mathbf{\Gamma})$ | (111) |
| | (110) |

Table 2.2: Outcome of $TrueMaj$ on the profile from Example 2.4

*The lowest Hamming distance between the models of agent* $1$*'s goal and the interpretations in the outcome are:*

$$lowH(\gamma_1,(111)) = H((111),(111)) = 0 \quad lowH(\gamma_1,(110)) = H((110),(111)) = 1.$$

*Thus this profile induces the* $optH$*,* $pessH$ *and* $realH$ *distances for agent* $1$ *as below*

$$\begin{aligned} optH(\gamma_1, TrueMaj(\mathbf{\Gamma})) &= \min\{lowH(\gamma_1,(111)), lowH(\gamma_1,(110))\} = 0 \\ pessH(\gamma_1, TrueMaj(\mathbf{\Gamma})) &= \max\{lowH(\gamma_1,(111)), lowH(\gamma_1,(110))\} = 1 \\ realH(\gamma_1, TrueMaj(\mathbf{\Gamma})) &= \frac{lowH(\gamma_1,(111)) + lowH(\gamma_1,(110))}{2} = \frac{1}{2}. \end{aligned}$$

As an H-optimist, the first friend in Example 2.4 believes that they will have all three courses and therefore is totally happy. As an H-pessimist or H-realist, she is more skeptical. As H-pessimist she is least satisfied because she would be convinced that only the first two courses will be served. The H-realist on the other hand still believes both results to be possible and holds an average satisfaction.

The different types of agents could therefore also be seen as risk-taking or risk-averse. This interpretation is formalized by the uncertainty appeal and uncertainty aversion axiom by Bossert et al. (2000) for preference extensions from elements to sets of elements.

**Definition 2.13** (Simple Uncertainty Appeal)**.** *For any three elements* $x, y, z$*, if we have* $x \succ y \succ z$*, then the extension is such that* $\{x, z\} \succ \{y\}$*.*

**Definition 2.14** (Simple Uncertainty Aversion)**.** *For any three elements* $x, y, z$*, if we have* $x \succ y \succ z$*, then the extension is such that* $\{y\} \succ \{x, z\}$*.*

Bossert et al. (2000) use these notions to characterize their Mini-Max and Maxi-Min extensions, which are not to be confused with *Maxi-Max* and *Maxi-Min* from Packard

(1979). While our notions only compare the minimal or maximal element respectively, the Mini-Max and Maxi-Min first compare the minimal (maximal) element and then the maximal (minimal), in case the former are equally preferred. Thus, these Maxi-Min and Mini-Max extensions are even more fine grained. However, if the first elements which are compared differ, the two approaches coincide. If the premise of simple uncertainty aversion and appeal is satisfied, we know that the minimal and maximal element of the given sets will differ. Thus, we can transfer the results from Bossert et al. (2000) that Maxi-Min satisfies simple uncertainty appeal and Mini-Max satisfies uncertainty aversion.

**Proposition 2.1.** *The $optH$ function satisfies simple uncertainty appeal.*

*Proof.* Take $\mathcal{N} = \{1, \dots, n\}$ to be the set of agents, deciding over issues $\mathcal{I} = \{1, \dots, m\}$. Assume $u, v, w \in \{0,1\}^m$ and suppose the goal $\gamma_i$ of agent $i$ is such that $lowH(\gamma_i, u) < lowH(\gamma_i, v) < lowH(\gamma_i, w)$, i.e., the induced preference gives $u \succ_i v \succ_i w$.

For $O = \{u, w\}$ and $O' = \{v\}$ we get

$$optH(\gamma_i, O) = \min\{lowH(\gamma_i, u), lowH(\gamma_i, w))\} = lowH(\gamma_i, u) \text{ and}$$
$$optH(\gamma_i, O') = lowH(\gamma_i, v).$$

By assumption we have $lowH(\gamma_i, u) < lowH(\gamma_i, v)$, which gives us that the optimists distance from $O'$ is bigger, i.e., $optH(\gamma_i, O) < optH(\gamma_i, O')$, and therefore agent $i$ prefers $\{u, w\}$ over $\{v\}$. □

**Proposition 2.2.** *The $pessH$ function satisfies simple uncertainty aversion.*

The proof of this Proposition 2.2 is analogue to the one of Proposition 2.1. Our three extensions collapse in case the outcome is a singleton, for example, if the voting rule is resolute. This is the case because, given only one interpretation in the outcome, there is also only one lowest Hamming distance to consider. Thus, the satisfaction for the interpretation in the outcome is exactly $lowH$.

**Proposition 2.3.** *If the voting profile $\mathbf{\Gamma}$ and the voting rule $F$ are such that $|F(\mathbf{\Gamma})| = 1$, then the $realH$, $pessH$ and $optH$ functions coincide.*

*Proof.* Take $F$ to be a voting rule and $\mathbf{\Gamma} = (\gamma_1, \dots, \gamma_n)$ to be a voting profile on issues $\mathcal{I} = \{1, \dots, m\}$, such that $F(\mathbf{\Gamma}) = \{v\}$ for some interpretation on $\mathcal{I}$. Take any $i \in \mathcal{N}$, then $\{lowH(\gamma_i, w) \mid w \in F(\mathbf{\Gamma})\}$ has only one element, namely $lowH(\gamma_i, v)$. Therefore, since it is the case that $|F(\mathbf{\Gamma})| = 1$ we get

$$lowH(\gamma_i, v) = \min\{lowH(\gamma_i, w) \mid w \in F(\mathbf{\Gamma})\}$$

$$= \max\{lowH(\gamma_i, w) \mid w \in F(\mathbf{\Gamma})\}$$
$$= \frac{\sum_{w \in F(\mathbf{\Gamma})} lowH(\gamma_i, w)}{|F(\mathbf{\Gamma})|}.$$

Hence the distances $realH$, $pessH$ and $optH$ coincide. □

**Corollary 2.1.** *If the voting rule is resolute, then the $realH$, $pessH$ and $optH$ functions coincide.*

However, none of these three notions implies another, i.e., one agent type preferring a certain outcome over another does not imply that another agent type does as well. This tells us that all three notions are actually different and have to be considered separately in the remainder of this thesis.

**Proposition 2.4.** *The satisfaction functions yielded by $optH$, $pessH$ or $realH$ differ.*

*Proof.* In order to prove this proposition it suffices to show examples in which each satisfaction function, H-optimist, H-pessimist and H-realist, disagrees with the others.

Take an agent with the goal $\gamma = (p \wedge \neg q \wedge \neg r) \vee (p \leftrightarrow q \wedge \neg r))$ for three issues. Then the models of her goal are $Mod(\gamma) = \{(100), (000), (110)\}$. Let now $O = \{(111)\}$ and $O' = \{(111), (011)\}$ be some possible outcomes. The agent will have the satisfaction as presented below:

$$optH(\gamma, O) = 1 = optH(\gamma, O')$$
$$pessH(\gamma, O) = 1 < 2 = pessH(\gamma, O')$$
$$realH(\gamma, O) = 1 < \frac{3}{2} = realH(\gamma, O').$$

Hence an H-pessimist and H-realist prefer $O'$ over $O$, while an H-optimist does not.

Take an agent to have the goal $\gamma = p \wedge (q \leftrightarrow r)$ for three issues. So then the models of her goal are $Mod(\gamma) = \{(100), (111)\}$. Let now $O = \{(101)\}$ and $O' = \{(101), (100)\}$ be some possible outcomes. The agent will have the satisfaction given by:

$$optH(\gamma, O) = 1 > 0 = optH(\gamma, O')$$
$$pessH(\gamma, O) = 1 = pessH(\gamma, O')$$
$$realH(\gamma, O) = 1 > \frac{1}{2} = realH(\gamma, O').$$

Hence an H-optimist and H-realist prefers $O'$ over $O$, while a H-pessimist does not. □

The H-optimist and H-pessimist can be seen as a generalization of the notion of a optimist and pessimist as defined in Section 2.2. The following propositions show that the preference of an optimist (pessimist) implies a preference of an H-optimist (H-pessimist).

**Proposition 2.5.** *If for some voting rule $F$ and two profiles $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$ it is the case that a optimist prefers $F(\mathbf{\Gamma})$ over $F(\mathbf{\Gamma}')$, then the same holds for an H-optimist.*

*Proof.* Take $\mathcal{N} = \{1, \ldots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \ldots, m\}$ a set of issues and take $F$ to be a voting rule. Let the profiles $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$ be such that agent $i \in \mathcal{N}$ prefers $F(\mathbf{\Gamma})$ over $F(\mathbf{\Gamma}')$ as an optimist, i.e., we have $F(\mathbf{\Gamma}) \cap Mod(\gamma_i) \neq \emptyset$ and $F(\mathbf{\Gamma}') \cap Mod(\gamma_i) = \emptyset$. Hence there is $v \in F(\mathbf{\Gamma})$ with $v \in Mod(\gamma_i)$, thus $lowH(\gamma_i, v) = 0$ and therefore $optH(\gamma_i, F(\mathbf{\Gamma})) = 0$. But we also get that for all $v \in F(\mathbf{\Gamma}')$ we have that $v \notin Mod(\gamma_i)$, therefore for all $v \in F(\mathbf{\Gamma}')$ it holds that $lowH(\gamma_i, v) > 0$ and so $optH(\gamma_i, F(\mathbf{\Gamma})) > 0$. This leads to $optH(\gamma_i, F(\mathbf{\Gamma})) < optH(\gamma_i, F(\mathbf{\Gamma}'))$ and hence H-optimists prefer $F(\mathbf{\Gamma})$ over $F(\mathbf{\Gamma}')$. □

**Proposition 2.6.** *If for some voting rule $F$ and two profiles $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$ it is the case that a pessimist prefers $F(\mathbf{\Gamma})$ over $F(\mathbf{\Gamma}')$, then the same holds for an H-pessimist and H-realist.*

*Proof.* Take $\mathcal{N} = \{1, \ldots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \ldots, m\}$ a set of issues and assume $F$ to be a voting rule. Let the profiles $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$ be such that agent $i \in \mathcal{N}$ prefers $F(\mathbf{\Gamma})$ over $F(\mathbf{\Gamma}')$ as a pessimist, i.e., we have $F(\mathbf{\Gamma}) \subseteq Mod(\gamma_i)$ and $F(\mathbf{\Gamma}') \nsubseteq Mod(\gamma_i)$. Hence we get $lowH(\gamma_i, w) = 0$ for all $w \in F(\mathbf{\Gamma})$ and so $pessH(\gamma_i, F(\mathbf{\Gamma})) = realH(\gamma_i, F(\mathbf{\Gamma})) = 0$. Additionally, there is at least one $w \in F(\mathbf{\Gamma}')$, $w \notin Mod(\gamma_i)$ and so $lowH(\gamma_i, w) > 0$. Thus we have $pessH(\gamma_i, F(\mathbf{\Gamma'})) > 0 = pessH(\gamma_i, F(\mathbf{\Gamma}))$ and $realH(\gamma_i, F(\mathbf{\Gamma'})) > 0 = realH(\gamma_i, F(\mathbf{\Gamma}))$ and hence H-pessimists and H-realists prefer $F(\mathbf{\Gamma})$ over $F(\mathbf{\Gamma}')$. □

### 2.2.3 Manipulable Voting Rules

In this section we will be discussing strategic voting. We will be looking at results from Novaro (2019) and extend them to the satisfaction functions based on Hamming distance and the *Approval* rule. It is essential for iteration to understand under which conditions manipulation is possible, since there cannot be an iteration if the rule is strategy-proof.

For any given profile $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_n)$, $(\mathbf{\Gamma}_{-i}, \gamma_i')$ will denote the profile in which $\gamma_i$ is substituted with $\gamma_i'$ and all the other goals are the same. A voting rule $F$ is said to be *manipulable* if there is a profile $\mathbf{\Gamma}$ and a formula $\gamma_i'$ such that $F((\mathbf{\Gamma}_{-i}, \gamma_i')) \succ_i F(\mathbf{\Gamma})$; in this case we say that agent $i$ has an *incentive to manipulate*. If a rule is not manipulable, it is called *strategy-proof*. Novaro (2019) has proven *EMaj*, *TrueMaj* and *2sMaj* to be manipulable in general. Strategy-proofness, however, can be achieved by restricting the manipulation as well as the language of goals. Given that $F((\mathbf{\Gamma}_{-i}, \gamma_i')) \succ_i F(\mathbf{\Gamma})$ holds for optimists or pessimists, we can apply Propositions 2.5 and 2.6 to derive the following corollaries.

**Corollary 2.2.** *If an optimist can manipulate a rule, so can an H-optimist.*

**Corollary 2.3.** *If a pessimist can manipulate a rule, so can an H-pessimist and H-realist.*

**Corollary 2.4.** *If a rule is strategy-proof for an H-optimist or H-pessimist, then it is for an optimist or pessimist, respectively.*

Novaro (2019) defined three different types of manipulation for an agent $i$ in goal-based voting, where $\gamma_i$ represents the original goal and $\gamma_i'$ the manipulation:

- Erosion: $Mod(\gamma_i') \subseteq Mod(\gamma_i)$.
- Dilatation: $Mod(\gamma_i) \subseteq Mod(\gamma_i')$.
- Unrestricted: $Mod(\gamma_i') \neq \emptyset$.

The following results hold for optimists, pessimists or expected utility maximizers.

**Proposition 2.7** (Novaro 2019)**.** *$EMaj$, $TrueMaj$ and $2sMaj$ can be manipulated by both erosion and dilatation.*

This result also holds for H-optimists, H-pessimists and H-realists by Corollary 2.2 and 2.3. For dichotomous agents strategy-proofness can be secured by restricting the language to conjunctions, as shown by Novaro (2019). The language only consisting of conjunctions of positive and negative literals will be denoted by $\mathcal{L}^\wedge$.

**Proposition 2.8** (Novaro 2019)**.** *For any profile $\mathbf{\Gamma}$ where $\gamma_i \in \mathcal{L}^\wedge$ for some $i \in \mathcal{N}$, agent $i$ has no incentive to manipulate unrestrictedly the rules $2sMaj$, $EMaj$ and $TrueMaj$.*

However, even with a restricted language (such as conjunctions) we still find manipulable profiles for the $Approval$ rule, also for agents basing their satisfaction on Hamming distance. This can be derived from Terzopoulou and Endriss (2018) where it was shown that agents can manipulate the voting process. They were able to show this for the plurality rule, which is a resolute version of our $Approval$ rule. Note that goal-based voting coincides with classical judgment aggregation in case of goals being restricted to conjunctions. In general, if the goals are (incomplete) conjunctions, goal-based voting can be seen as a form of judgment aggregation with abstentions.

**Proposition 2.9.** *The $Approval$ rule can be manipulated by H-optimists, H-pessimist and H-realists by unrestricted manipulation.*

*Proof.* Table 2.3 shows a profile in which agent 1 can manipulate the outcome to be more preferred by an H-optimist on the left and H-pessimist on the right, respectively. Both alterations would also take place if agent 1 was an H-realist, since the average $lowH$ distance decreases in both examples. □

| | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ | | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ |
|---|---|---|---|---|---|
| $Mod(\gamma_1)$ | (1111)<br>(1110) | **(1111)**<br>**(1110)**<br>**(1100)** | $Mod(\gamma_1)$ | (0001) | **(0000)** |
| $Mod(\gamma_2)$ | (0000) | (0000) | $Mod(\gamma_2)$ | (0000)<br>(1000) | (0000)<br>(1000) |
| $Mod(\gamma_3)$ | (0000) | (0000) | $Mod(\gamma_3)$ | (0000)<br>(1000) | (0000)<br>(1000) |
| $Mod(\gamma_4)$ | (1100) | (1100) | | | |
| $Approval$ | (0000) | (0000)<br>(1100) | $Approval$ | (0000)<br>(1000) | (0000) |

Table 2.3: Manipulation example for $Approval$ voting with $optH$, $pessH$ and $realH$.

The $Approval$ rule for optimists, pessimists and expected utility maximizers will be shown to be strategy-proof in Proposition 2.10. This result is a special case of a result by Everaere et al. (2007), since the $Approval$ rule corresponds to the $\Delta_{\mu}^{\Sigma,d}$ rule in belief merging where $d$ is the drastic distance $d_D$ and the satisfaction functions $opt$, $pess$ and $eum$ represents the indices $i_{dw}$, $i_{ds}$ and $i_p$ as defined in their paper, respectively. They first proved that strategy-proofness with $i_p$ implies the strategy-proofness of the other satisfaction functions. Then they prove that $\Delta_{\mu}^{d_D,f}$ is strategy-proof for the drastic distance $d_D$ and any aggregation function $f$ and constraint $\mu$. Considering $f = \Sigma$ and $\mu = \top$ this theorem coincides with Proposition 2.10.

**Proposition 2.10** (Everaere et al., 2007)**.** *$Approval$ rule is strategy-proof for optimists, pessimists and expected utility maximizers.*

Note that an agent who bases her satisfaction on a dichotomous function only has an incentive to alter her goal if she can either further support some models of her truthful goal to be in the outcome or eliminate some non-models of it from the outcome. However, their truthful vote, in which they approve exactly the models of their goal, does exactly this: they support all models of their goals and none of the non-models. This is, intuitively speaking, the reason why the result of Everaere et al. (2007) holds.

Given these manipulation results the question of iterated voting arises, since one might ask what happens if changes are allowed and hence agents can alter their vote to reach a more preferred outcome. In the following section we will define the iteration process and analyze the termination for the majorities as well as $Approval$ voting. Given the result of Proposition 2.10 in this section our analysis of the iterated $Approval$ rule can be limited to H-optimists, H-pessimists and H-realists only.

# 3 | Iterated Majorities and Approval

Iteration of a voting process can serve multiple purposes. Imagine an engaged couple and their families, who uses the platform Doodle to decide on what to serve for dinner on the wedding, with only the options to approve or disapprove a given list of courses. Now assume everyone is entering their votes in a strategic way: when in doubt, they rather disapprove. If by acting in such a way no dinner is found, the families might prefer to reevaluate their vote before the dinner is chosen in another way. Usually in these cases an approval rule is used, which outputs all meals with the highest support, from which then the final course can be chosen. In case the votes for none of the options are overlapping, the vote would just return the list of all meals and hence not helping the bride and groom to plan their wedding dinner. This is just one example on how iterative voting can be worthwhile to be analyzed.

The termination, complexity and quality of iterative processes, for voting and judgment aggregation, has experienced a rise of interest in recent years (Meir et al., 2017; Terzopoulou and Endriss, 2018). Some of the classic assumptions, e.g., the best response dynamic and agents being myopic, will also be addressed here.

In this chapter we will first define the iteration process formally for the framework of goal-based voting: fixing notation, assumptions and agents' characteristics. Then, we will analyze the Majority rules and the $Approval$ rule, focusing on the question of whether or not the iterated process of these rules is guaranteed to terminate.

## 3.1 The Iteration Process

An iteration is a repeated voting process at which in each step the result is computed by a fixed rule, while the profile changes during the process. We define *stages* denoted by a natural number $t \in \mathbb{N}$, to count the number of iteration steps taken in one process. The stage $t = 0$ describes the initial situation. Any further stage $t > 0$ describes the situation after $t$ alteration steps. At stage $t$, the goal of an agent $i \in \mathcal{N}$ is denoted by $\gamma_i^t$, and all the agents' goals build the current profile $\mathbf{\Gamma}^t$. The initial profile $\mathbf{\Gamma}^0$ is assumed to contain the truthful goals of all agent.

In the following we will speak of an agent altering their vote rather than manipulating. We choose this terminology to avoid the bad connotation associated to an agent changing her vote: while agents act strategically when altering their vote, it should be understood as them optimizing their vote according to their goal rather than manipulating the result. An agent has an *incentive to alter* her vote if she has an incentive to manipulate as defined in Section 2.2.3. According to a satisfaction function $sat$, this means that an agent has such an incentive if $sat(\gamma_i^0, F(\mathbf{\Gamma}^t)) < sat(\gamma_i^0, F(\mathbf{\Gamma}^{t+1}))$ for the new profile $\mathbf{\Gamma}^{t+1} = (\mathbf{\Gamma}^t_{-i}, \gamma_i')$, in which $i's$ goal is substituted by $\gamma_i'$. Note that, for the satisfaction the outcome at each stage will always be compared to the original goal $\gamma_i^0$ of an agent $i$. We will call the goals which yield the best possible improvement at a certain stage *best responses*, as per the following definition:

**Definition 3.1** (Best Response)**.** *If agent $i$ has an incentive to alter her vote, then a goal $\gamma'$ is a best response at step $t$, if for any other goal $\gamma$ it is the case that:*

$$sat(\gamma_i^0, F(\mathbf{\Gamma}^t_{-i}, \gamma)) \leq sat(\gamma_i^0, F(\mathbf{\Gamma}^t_{-i}, \gamma')).$$

Observe that at any stage $t$, there could be multiple possible improvements and multiple best responses. An iteration from one stage $(t)$ to another $(t+1)$ will be taking place if there is an agent who has an incentive to alter her vote. If multiple agents have such an incentive, one will be chosen randomly. Each step only includes an alteration of one agent. Thus, in principle multiple different iterations are possible from a given profile. An iteration *terminates* if after a finite number of steps, no agent has an incentive to alter her vote. Note that, if at some stage $t$ it is $\mathbf{\Gamma}^t = \mathbf{\Gamma}^s$ for a stage $s < t$, the iteration does not terminate. An iteration in which this happens will be called *circular*.

We will assume agents to have certain characteristics in order to predict their behaviour in the iteration process. First, agents are *myopic* as defined by Meir et al. (2017), i.e., they only think about the next step and are not able to predict what happens after that. A best response is only defined for one further iteration steps. Second, agents are *improvement-driven*, i.e., with a given satisfaction function an agent will always want to alter her vote if she can. Note that any iteration step only allows one agent to change their vote, so this assumptions says that, from all agents who have an incentive to vote, if one agent is chosen to do so, she will. Last, agents are *fully informed* (Terzopoulou and Endriss, 2018): they know about any other agent's vote and how the voting rule works. However, agents being myopic limits this information to only one further step of iteration. These characteristics will be assumed for any agent from now on. Additionally, we will sometimes consider agents to be *truth-biased*, as per the following definition.

**Definition 3.2** (Truth-Bias)**.** *An agent $i$ is truth-biased if, in case she has no best response as per Definition 3.1 under which the outcome is strictly better than the current one, and the current outcome $F(\mathbf{\Gamma}^t)$ is equally satisfying as the outcome under the profile $(\mathbf{\Gamma}^t_{-i}, \gamma_i^0)$ in which she votes for her initial goal, i.e., $sat(\gamma_i^0, F(\mathbf{\Gamma}^t)) = sat(\gamma_i^0, F(\mathbf{\Gamma}^t_{-i}, \gamma_i^0))$, she then has an incentive to alter her vote to the truthful goal $\gamma_i^0$.*

Thus, these agents want to return to their truthful goal rather than keeping the altered vote, if the former has no negative consequence according to the given satisfaction.

## 3.2 Iterated Majority Rule

The majority rule in voting has been shown to be the only anonymous, neutral, and monotonic rule (May, 1952). The majority rule is a special instance of the quota rule with threshold $\frac{n}{2}$. An introduction to general quota rules or majority voting in judgment aggregation can be found in Dietrich and List (2007a) and Endriss (2016). In the context of goal-based voting, Novaro (2019) proved that $TrueMaj$ is the only rule which satisfies a list of axioms, including versions for goal-based voting of anonymity, neutrality, independence and monotonicity. However, while the $TrueMaj$ rule satisfies some desirable axioms that might be argued to add to the fairness of voting, $EMaj$ and $2sMaj$ could still be preferred because of their resoluteness.

In goal-based voting the Majority rules have been shown to be manipulable (Novaro, 2019). In voting already, manipulability results gave rise to the analysis of iterated Majority voting (Airiau and Endriss, 2009). One result by Novaro (2019) states that assuming agents to be optimists, pessimists or expected utility maximizers, if the their goals are conjunctions in $\mathcal{L}^\wedge$, then $EMaj$, $TrueMaj$ and $2sMaj$ are strategy-proof, extending known results in judgment aggregation. In some way, strategy-proofness can be seen as a strong termination result, where the iteration terminates at $t = 0$. In this section we will show that, assuming truth-biased agents and any type of satisfaction function, the iteration of the Majority rules we consider, is not guaranteed to terminate.

Our first result, Proposition 3.1, proves precisely this: that there is a profile and feasible alterations, such that any of the considered Majority rules yields a circular iteration. This holds even for agents having dichotomous preferences, since through the alteration a non-model changes to a model for each agent altering.

**Proposition 3.1.** *Iterated $EMaj$, $TrueMaj$ or $2sMaj$ voting does not always terminate for truth-biased agents with satisfaction $opt$, $pess$, $eum$, $optH$, $pessH$ and $realH$.*

*Proof.* Table 3.1 shows an initial profile $\mathbf{\Gamma}^0$ for three agents and three issues, where none of the three Majority rules terminate for any of the satisfaction functions listed above.

Note that $EMaj$ and $TrueMaj$ coincide, since they yield a singleton under the considered profiles. By Proposition 2.3 the satisfaction of H-pessimist, H-optimist and H-realist coincide in case of resoluteness.

To see that higher satisfactions are reached note that by altering from $\mathbf{\Gamma}^0$ to $\mathbf{\Gamma}^1$ agent 3 changes the outcome from an interpretation that is not a model of her goal $(111)$ to one that is $(101)$. This means that the lowest Hamming distance strictly decreases from $lowH(\gamma_3, (111)) = 1$ to $lowH(\gamma_3, (101)) = 0$. The same is the case for agent 1's alteration from $\mathbf{\Gamma}^1$ to $\mathbf{\Gamma}^2$. The last two alterations are due to the truth bias: since submitting their truthful goal or the alteration leads to the same outcome for agent 3 and 1 in profiles $\mathbf{\Gamma}^2$ and $\mathbf{\Gamma}^3$, respectively, they choose to alter back to their truthful goals. Since $\mathbf{\Gamma}^4$ is identical to the initial profile $\mathbf{\Gamma}^0$, the iteration does not terminate. □

| | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ | $\mathbf{\Gamma}^2$ | $\mathbf{\Gamma}^3$ | $\mathbf{\Gamma}^4$ |
|---|---|---|---|---|---|
| $Mod(\gamma_1)$ | $(100)$ | $(100)$ | $\mathbf{(111)}$ | $(111)$ | $\mathbf{(100)}$ |
| | $(111)$ | $(111)$ | | | $\mathbf{(111)}$ |
| $Mod(\gamma_2)$ | $(111)$ | $(111)$ | $(111)$ | $(111)$ | $(111)$ |
| | $(101)$ | $(101)$ | $(101)$ | $(101)$ | $(101)$ |
| | $(110)$ | $(110)$ | $(110)$ | $(110)$ | $(110)$ |
| $Mod(\gamma_3)$ | $(101)$ | $\mathbf{(101)}$ | $(101)$ | $\mathbf{(101)}$ | $(101)$ |
| | $(010)$ | | | $\mathbf{(010)}$ | $(010)$ |
| | $(011)$ | | | $\mathbf{(011)}$ | $(011)$ |
| $E/True/2s-Maj$ | $(111)$ | $(101)$ | $(111)$ | $(111)$ | $(111)$ |

Table 3.1: Example of non-terminating Majority rules with truth-biased agents.

This result shows that iterated Majority voting with truth-biased agents is not guaranteed to terminate. However, truth-bias is quite a strong assumption: it adds to the set of possible alterations in the iterative process thus increasing the chance of returning to the initial profile and creating a circular iteration. By lifting this assumption one might think that termination is guaranteed. However, we can show that even when restricting the best responses to alterations which strictly increase the satisfaction, i.e., lifting the truth-bias, the $EMaj$ and $TrueMaj$ rules are still not guaranteed to terminate for any satisfaction function, as shown in Proposition 3.2.

**Proposition 3.2.** *Iterated $EMaj$ and $TrueMaj$ voting does not always terminate for the satisfaction functions $opt$, $pess$, $eum$, $optH$, $pessH$ and $realH$.*

*Proof.* Table 3.2 shows an initial profile $\mathbf{\Gamma}^0$ of truthful goals for four agents to which they return after four iteration steps. The given alterations are best responses for agents with any of the above mentioned satisfaction functions, since each alteration turns the outcome from a non-model to a model for the altering agent's goal. □

| | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ | $\mathbf{\Gamma}^2$ | $\mathbf{\Gamma}^3$ | $\mathbf{\Gamma}^4$ |
|---|---|---|---|---|---|
| $Mod(\gamma_1)$ | (100) | **(100)** | (100) | **(100)** | (100) |
| | (001) | **(101)** | (101) | **(001)** | (001) |
| | | **(111)** | (111) | | |
| | | **(010)** | (010) | | |
| $Mod(\gamma_2)$ | (101) | (101) | **(101)** | (101) | **(101)** |
| | (110) | (110) | **(111)** | (111) | **(110)** |
| | (000) | (000) | **(001)** | (001) | **(000)** |
| $Mod(\gamma_3)$ | (101) | (101) | (101) | (101) | (101) |
| | (010) | (010) | (010) | (010) | (010) |
| | (000) | (000) | (000) | (000) | (000) |
| $Mod(\gamma_4)$ | (101) | (101) | (101) | (101) | (101) |
| | (010) | (010) | (010) | (010) | (010) |
| | (000) | (000) | (000) | (000) | (000) |
| $E/TrueMaj$ | (000) | (100) | (101) | (001) | (000) |

Table 3.2: Example of non-terminating $EMaj$ and $TrueMaj$ rules without truth-bias.

This cycle is only established since the alterations of both agent 1 and agent 2 are such that once the latter agent altered, the alteration of the former is still effective but not beneficial anymore and vice-versa. Then each agent profits from again altering back to their initial goal. This is a quite specific situation and might have not happened if the agents chose different best responses. If agent 2, for example, chose a goal supporting $(000)$ rather than $(101)$ when altering from $\mathbf{\Gamma}^1$ to $\mathbf{\Gamma}^2$ by submitting for example, a goal $\gamma_2^2$ whose model is $Mod(\gamma_2^2) = \{(000)\}$, she would have determined the rejection of all issues together with agents 3 and 4. Then the same profile would have terminated with a different outcome under the use of the $EMaj$ and $TrueMaj$ rules.

The example seen in Table 3.2 is not enough to prove an equivalent statement for $2sMaj$. In fact, remember that $2sMaj$ first calculates $EMaj$ on each agent's goal and then applies the strict Majority rule to those intermediate results. Applying $EMaj$ to agent 3 and 4's goals returns $(000)$. Hence, these two agents already determine the outcome. The $2sMaj$ rule will return $(000)$ and neither agent 1 nor 2 will be able to alter. Hence, an interesting and still open question is whether or not lifting restrictions are enough in order to guarantee termination for $2sMaj$. The strategy-proofness results of Novaro (2019) show that restricting the language to $\mathcal{L}^{\wedge}$ will lead to no iteration at all. A natural conjecture would be that there are some restrictions building a middle ground, allowing iteration while still guaranteeing termination.

The Majority rule, despite its popularity and wide-spread use, earns reasonable criticism when defined in an issue-independent way, as done in judgment aggregation and goal-based voting. Since the issues might not be independent they should not be decided as such. For example, take an agent voting for a goal $p \leftrightarrow \neg q$ where $p$ and $q$ are the first and second issue, respectively. The Majority rules $EMaj$ and $TrueMaj$ assume the agent to split in half her voting power between $p$ and $q$. This suggests that the agent is indifferent whether or not $p$ or $q$ are chosen. However, this is not the case: the agent only wants one of them to be made true by the chosen interpretation and simply does not care which one. For example, we could imagine some agents to decide on to whether to go to Spain or Italy for their holiday and one agent submitting this goal due to her being indifferent where to go, but she still wants to go to exactly one location. The issues are not independent in this agent's goal.

In case the issues are also subject to an external integrity constraint this might even lead to inconsistent results, a problem known in the judgment aggregation literature as the *doctrinal paradox* (Kornhauser and Sager, 1993) or as the *discursive dilemma* (Pettit, 2001). Dietrich and List (2007b) analyze under which conditions quota rules do not suffer from the *discursive dilemma*. Botan and Endriss (2020) also address this dilemma and forward a solution by defining a novel notion of strategy-proofness. For this reason the Majority rule might not be the best choice for certain decision problems.

This dilemma yields one strong argument for the $Approval$ rule over the issue-wise Majority rules. As we shall see in the next section, the $Approval$ rule is a reasonable alternative for dependent issues, as it does not compute an outcome issue by issue, but considers the goal's models as one object, returning the interpretations with the maximal amount of support.

## 3.3 Iterated Approval Rule

Laslier and Sanver (2010) give a nice overview on the history, axiomatization and experimental findings of the approval rule in voting. The sincerity and strategy-proofness of the approval rule has been studied by Endriss (2013) and Brams and Fishburn (1978). For a definition, axiomatic characterization and analysis of strategy-proofness in judgment aggregation the reader is referred to Terzopoulou (2021).

In this section, we want to explore iterative $Approval$ voting in goal-based voting. Recall that in Definition 2.1, the $Approval$ rule was defined in Section 2.1 as $Approval(\mathbf{\Gamma}) = \text{argmax}_{v \in Mod(\bigvee_{i\in\mathcal{N}} \gamma_i)} \; supp_{\mathbf{\Gamma}}(v)$ for a profile $\mathbf{\Gamma}$ of $n$ agents and $m$ issues, where $supp_{\mathbf{\Gamma}}(v)$ is the number of supporters for interpretation $v \in \bigcup_{i\in\mathcal{N}} Mod(\gamma_i)$.

For a given profile $\mathbf{\Gamma}$, we denote as $k_{\mathbf{\Gamma}} = supp_{\mathbf{\Gamma}}(v)$ for $v \in Approval(\mathbf{\Gamma}^t)$ the support of the winning interpretations $v \in Approvla(\mathbf{\Gamma})$. In the iterative setting, if the profile is clear from context, we will write $k_t$ instead of $k_{\mathbf{\Gamma}^t}$. This notion will help to keep track of the support throughout the process of iteration.

The $Approval$ rule as defined in 2.1 is closely related to the plurality rule as defined by Terzopoulou and Endriss (2018) for judgment aggregation. The plurality rule explored in their setting chooses from all judgments the most reported one. Some of their results will be shown to transfer to goal-based voting (see Proposition 3.3). First, we consider no assumptions on the iteration process. Then we will introduce some restrictions restrictions that will yield in some positive termination results.

### 3.3.1 Unrestricted Iteration

The $Approval$ rule is strategy-proof (see Theorem 2.10), for optimists, pessimists and expected utility maximizers and will therefore not admit iteration in those cases.

**Corollary 3.1.** *Iterated $Approval$ voting with optimists, pessimists or expected utility maximizers terminates at $t = 0$.*

Strategy-proofness cannot be established for agents whose satisfaction is based on Hamming distance as shown by Proposition 2.9. However, for H-pessimists the iterated $Approval$ rule is guaranteed to terminate in finitely many steps without any further restrictions. This will be shown in Theorem 3.1, of which the proof essentially depends on the following two lemmas.

**Lemma 3.1.** *By considering H-pessimist agents only, no agent under the $Approval$ rule would alter her vote from stage $t$ to $t+1$ if it would result in $k_{t+1} < k_t$.*

*Proof.* Take $\mathcal{N} = \{1, \dots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \dots, m\}$ a set of issues. Let $\mathbf{\Gamma}^0 = (\gamma_1, \dots, \gamma_n)$ be the initial voting profile.

Assume now that iteration is at a stage $t$ and $Approval(\mathbf{\Gamma}^t)$ is the current outcome. Suppose agent $i$ can alter her goal such that $k_{t+1} < k_t$. The only way $k_{t+1} < k_t$ can hold is if all the interpretations in $Approval(\mathbf{\Gamma}^t)$ lose at least one agent's support in the iteration to stage $t + 1$. Since agent $i$ is the only one altering her goal she has to withdraw support from every $v \in Approval(\mathbf{\Gamma}^t)$. Thus, for any $v \in Approval(\mathbf{\Gamma}^t)$ we have $supp_{\mathbf{\Gamma}^{t+1}}(v) = k_t - 1$. Since $k_{t+1} < k_t$ we get that this support is now the highest, so $k_{t+1} = k_t - 1$ and therefore $v \in Approval(\mathbf{\Gamma}^{t+1})$ for all $v \in Approval(\mathbf{\Gamma}^t)$. But then all interpretations which were furthest in Hamming distance from $i$'s goal, i.e., $w \in Approval(\mathbf{\Gamma}^t)$ s.t. $lowH(\gamma_i, w) = pessH(i, Approval(\mathbf{\Gamma}^t))$ are still in the outcome of $Approval(\mathbf{\Gamma}^{t+1})$. Thus as an H-pessimist the satisfaction of $i$ would not increase, since $pessH(i, Approval(\mathbf{\Gamma}^{t+1})) \leq lowH(\gamma_i, w) = pessH(i, Approval(\mathbf{\Gamma}^t))$.

Hence agent $i$ would not alter her vote if it would result in $k_{t+1} < k_t$. □

**Lemma 3.2.** *By considering H-pessimist agents only, no agent under the Approval rule would alter her vote from stage $t$ to $t+1$ if it would result in $k_{t+1} = k_t$.*

*Proof.* Take $\mathcal{N} = \{1, \ldots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \ldots, m\}$ a set of issues. Let $\mathbf{\Gamma}^0 = (\gamma_1, \ldots, \gamma_n)$ be the voting profile.

Take stage $t$ to be the first stage at which by an alteration of some agent $i$ we have $k_{t+1} = k_t$. Note that since every agent is an H-pessimist, by Lemma 3.1 we get that $k_{r+1} > k_r$ for any $r < t$. This means that and any interpretation $v \in Approval(\mathbf{\Gamma}^r)$ for any $r \leq t$ was already in the initial outcome $Approval(\mathbf{\Gamma}^0)$. This is so since the support of the winning interpretations $k_r$ strictly increased in these alterations and the support of any interpretation can only increase by one with one alteration. If for example one interpretation $v$ was such that it was missing only the support of one more agent to be approved in the initial profile $supp_{\mathbf{\Gamma}^0}(v) = k_0 - 1$, then after an alteration step $v$ gained at most one support, but the support of the winning interpretations also increased by one since $k_{r+1} > k_r$, hence we would have $supp_{\mathbf{\Gamma}^1}(v) = k_1 - 1$. Thus, we get $Approval(\mathbf{\Gamma}^{r+1}) \subseteq Approval(\mathbf{\Gamma}^r)$ for any $r < t$.

Remember that since the agents are H-pessimists we have that for any $i \in \mathcal{N}$:

$$Approval(\mathbf{\Gamma}^{t+1}) \succ_i Approval(\mathbf{\Gamma}^t) \Leftrightarrow$$
$$pessH(\gamma_i, Approval(\mathbf{\Gamma}^{t+1})) < pessH(\gamma_i, Approval(\mathbf{\Gamma}^t)).$$

Note that an H-pessimist does not prefer an outcome $Approval(\mathbf{\Gamma}^{t+1})$ if it is a superset of $Approval(\mathbf{\Gamma}^t)$, since the interpretation with the lowest distance would still be in the altered outcome. Thus as H-pessimist $i$ would only alter her vote resulting in $k_{t+1} = k_t$ if she can withdraw her support from the least preferred interpretations of the current outcome $\{w_0, \ldots, w_l\} \subseteq Approval(\mathbf{\Gamma}^t)$, i.e., those interpretations with $lowH(\gamma_i, w_j) = pessH(i, Approval(\mathbf{\Gamma}^t))$ for any $j \in \{0, \ldots, l\}$. This alteration will only increase her satisfaction in two cases:

(i) if for any other interpretation $v \in \mathit{Approval}(\mathbf{\Gamma}^t) \setminus \{w_0, \dots, w_l\}$ and any $j \in \{0, \dots, l\}$ it is the case that $lowH(\gamma_i, v) < lowH(\gamma_i, w_j)$.

(ii) if $\{w_0, \dots, w_l\} = \mathit{Approval}(\mathbf{\Gamma}^t)$ and there is a $v \in \{0,1\}^m$ with $lowH(\gamma_i, v) < lowH(\gamma_i, w_j)$ for any $j \in \{0, \dots, l\}$, $supp_{\mathbf{\Gamma}^t}(v) = k_t - 1$ and $v \notin Mod(\gamma_i^t)$, such that agent $i$ can then support $v$ to be in $\mathit{Approval}(\mathbf{\Gamma}^{t+1})$.

Either way this means that $lowH(\gamma_i, w_j) \neq 0$ and hence $w_j \notin Mod(\gamma_i)$ for any $j \in \{0, \dots, l\}$. This means that since these interpretations $w_j$ are not models of her initial goal $\gamma_i^0$ and yet they are in $Mod(\gamma_i^t)$, agent $i$ must have altered her vote before at some stage $r_j < t$ for every such $j$. These stages actually must have been the same for all $w_j$, since $\mathit{Approval}(\mathbf{\Gamma}^{r+1}) \subseteq \mathit{Approval}(\mathbf{\Gamma}^r)$, i.e., anything that is in an outcome $\mathit{Approval}(\mathbf{\Gamma}^r)$ must have been in the outcome of any earlier stage and hence gained support at each iteration step, since the $k_r$ are strictly increasing. If the $w_j$'s were added in different stages, they could not all be in $\mathbf{\Gamma}^t$, since agent $i$ can only increase their support by one and the support of the winning interpretations is strictly increasing with each stage.

Thus, take $s$ to be the stage at which agent $i$ altered in favour of all the $w_j$. We will now distinguish the two cases (i) and (ii) from above.

First, consider $v \in \mathit{Approval}(\mathbf{\Gamma}^{t+1}) \setminus \{w_0, \dots w_l\}$: we know that $v$ has also been part of the outcome at stage $s$ and agent $i$ increased its support with the alteration at this stage. But this means that the alteration from agent $i$ at stage $s$ was not a best response. It would have been better if she chose to only increase the support of $v \in \mathit{Approval}(\mathbf{\Gamma}^{t+1})$ and none of the $\{w_0, \dots, w_l\}$. This is a contradiction to agent $i$ altering to a best response.

Second, assume instead that the agent $i$ withdraw the support from all interpretations in $\mathit{Approval}(\mathbf{\Gamma}^t)$. This means that all $v \in \mathit{Approval}(\mathbf{\Gamma}^{t+1})$ are such that $v \notin Mod(\gamma_i^t)$ and $supp_{\mathbf{\Gamma}^t}(v) = k_t - 1$, but then for similar reasons as seen before it must have been that $supp_{\mathbf{\Gamma}^r}(v) \geq k_r - 1$ at any stage $r \leq t$. Let $s_i < t$ be the latest stage at which agent $i$ was altering her goal. Since $supp_{\mathbf{\Gamma}^r}(v) \geq k_r - 1$ at any stage $r \leq t$ we also have $supp_{\mathbf{\Gamma}^{s_i}}(v) \geq k_{s_i} - 1$. If it was the case that $supp_{\mathbf{\Gamma}^{s_i}}(v) = k_{s_i}$, then:

- either agent $i$ did not hold them as models of $\gamma_i^{s_i}$: but then supporting $v$ instead of the $w_j$ for $\{w_0, \dots, w_l\}$ would have been a better response. This contradicts the assumption that agents only alter to best responses.

- or agent $i$ had them as models of $\gamma_i^{s_i}$: but then since they are not in $Mod(\gamma_i^t)$ we know that she did not support them anymore after this alteration. Which means that the support of $v$ decreased to $supp_{\mathbf{\Gamma}^{s_i+1}}(v) = supp_{\mathbf{\Gamma}^{s_i}}(v) - 1 = k_{s_i} - 1$. Thus, since the support of the winning interpretation strictly increases, we get $k_{s_i+1} = k_{s_i} + 1$

and $supp_{\mathbf{\Gamma}^{s_i+1}}(v) = k_{s_i+1} - 2$. This contradicts the fact that $supp_{\mathbf{\Gamma}^r}(v) \geq k_r - 1$ at any stage $r \leq t$.

Therefore we know that $supp_{\mathbf{\Gamma}^{s_i}}(v) = k_{s_i} - 1$. And since agent $i$ increased the support of the winning interpretations but did not support any of these $v$'s, we have $supp_{\mathbf{\Gamma}^{s_i+1}}(v) \leq k_{s_i+1} - 2$. However, this again contradicts the fact that $supp_{\mathbf{\Gamma}^r}(v) \leq k_r - 1$ at any stage $r \leq t$. Hence, no agent $i$ can alter as it is presented by (ii).

In conclusion, there will never be such an alteration at stage $t$ by an H-pessimist. □

**Theorem 3.1.** *Iterated Approval voting with H-pessimists terminates after at most* $|\mathcal{N}| - k_0$ *rounds.*

*Proof.* Take $\mathcal{N} = \{1, \ldots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \ldots, m\}$ a set of issues. Let $\mathbf{\Gamma}^0 = (\gamma_1, \ldots, \gamma_n)$ be the voting profile.

By Lemma 3.1 and Lemma 3.2 we know that for any alteration of an H-pessimist from any stage $t$ to $t+1$ it is the case that $k_{t+1} > k_t$. Hence an agent can only add support to interpretations already in the outcome. Since an interpretation can only have at most $|\mathcal{N}|$ support and any $v \in Approval(\mathbf{\Gamma}^0)$ has support $k_0$, we get that there can be at most $|\mathcal{N}| - k_0$ many alterations. Thus, the iteration is guaranteed to terminate. □

So the iteration with H-pessimists is always guaranteed to terminate in at most as many steps as there are agents disagreeing with the initial result. The picture looks different when considering H-optimists.

Terzopoulou and Endriss (2018) could already prove that in judgment aggregation the plurality rule with truth-biased agents, whose satisfaction is based on the Hamming distance, is not guaranteed to terminate. Judgment aggregation coincides with goal-based voting where goals are restricted to complete conjunctions. By adapting their Approval rule to our irresolute version and by observing that a restriction of the language implies the statement for the general language, their result implies the following proposition for H-optimists and H-realists. Their counterexample included five agents, four issues and four iteration steps until it circled back. Calculations show that even though they consider a different type of satisfaction, once we extend the rule to the irresolute $Approval$ rule and check satisfactions, the alterations are still favourable for H-optimists and H-realists.

**Proposition 3.3** (Terzopoulou and Endriss 2018)**.** *Iterated Approval voting with truth-biased agents might not terminate with H-optimists or H-realists.*

Truth-bias allows agents to have more possible responses in the iteration process. So one might think that lifting this assumption would yield a positive result about termination of the iteration, as we have seen for H-pessimists. Surprisingly, this result that the

*Approval* rule always terminates for H-pessimists cannot be extended to H-optimists nor H-realists, as shown by Theorem 3.2.

**Theorem 3.2.** *Iterated Approval voting might not terminate, if agents are H-optimists or H-realists.*

| | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ | $\mathbf{\Gamma}^2$ | $\mathbf{\Gamma}^3$ | $\mathbf{\Gamma}^4$ | $\mathbf{\Gamma}^5$ | $\mathbf{\Gamma}^6$ |
|---|---|---|---|---|---|---|---|
| $Mod(\gamma_1)$ | (000000) | **(000000)**<br>**(110000)** | (000000)<br>(110000) | **(000000)**<br>**(100000)**<br>**(111110)** | (000000)<br>(100000)<br>(111110) | **(000000)**<br>**(110000)** | (000000)<br>(110000) |
| $Mod(\gamma_2)$ | (111111) | (111111) | **(111111)**<br>**(111100)**<br>**(100000)** | (111111)<br>(111100)<br>(100000) | **(111111)**<br>**(111110)** | (111111)<br>(111110) | **(111111)**<br>**(111100)**<br>**(100000)** |
| $Mod(\gamma_3)$ | (111000)<br>(100000)<br>(110000) | (111000)<br>(100000)<br>(110000) | (111000)<br>(100000)<br>(110000) | (111000)<br>(100000)<br>(110000) | (111000)<br>(100000)<br>(110000) | (111000)<br>(100000)<br>(110000) | (111000)<br>(100000)<br>(110000) |
| $Mod(\gamma_4)$ | (111000)<br>(110000)<br>(111100) | (111000)<br>(110000)<br>(111100) | (111000)<br>(110000)<br>(111100) | (111000)<br>(110000)<br>(111100) | (111000)<br>(110000)<br>(111100) | (111000)<br>(110000)<br>(111100) | (111000)<br>(110000)<br>(111100) |
| $Mod(\gamma_5)$ | (111000)<br>(111100)<br>(111110) | (111000)<br>(111100)<br>(111110) | (111000)<br>(111100)<br>(111110) | (111000)<br>(111100)<br>(111110) | (111000)<br>(111100)<br>(111110) | (111000)<br>(111100)<br>(111110) | (111000)<br>(111100)<br>(111110) |
| *Approval* | (111000) | (111000)<br>(110000) | (111000)<br>(110000)<br>(111100) | (111000)<br>(100000)<br>(111100) | (111000)<br>(111110) | (111000)<br>(110000) | (111000)<br>(110000)<br>(111100) |

Table 3.3: Example of non-terminating *Approval* rule for H-optimists and H-realists.

| | $\mathbf{\Gamma}^0$ | | $\mathbf{\Gamma}^1$ | | $\mathbf{\Gamma}^2$ | | $\mathbf{\Gamma}^3$ | | $\mathbf{\Gamma}^4$ | | $\mathbf{\Gamma}^5$ | | $\mathbf{\Gamma}^6$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $optH(\gamma_1, Approval(\mathbf{\Gamma}^t))$ | 3 | $\rightarrow$ | 2 | | 2 | $\rightarrow$ | 1 | | 3 | $\rightarrow$ | 2 | | 2 |
| $optH(\gamma_2, Approval(\mathbf{\Gamma}^t))$ | 3 | | 3 | $\rightarrow$ | 2 | | 2 | $\rightarrow$ | 1 | | 3 | $\rightarrow$ | 2 |
| $realH(\gamma_1, Approval(\mathbf{\Gamma}^t))$ | 3 | $\rightarrow$ | $\frac{5}{2}$ | | 3 | $\rightarrow$ | $\frac{8}{3}$ | | 4 | $\rightarrow$ | $\frac{5}{2}$ | | 3 |
| $realH(\gamma_2, Approval(\mathbf{\Gamma}^t))$ | 3 | | $\frac{7}{2}$ | $\rightarrow$ | 3 | | $\frac{10}{3}$ | $\rightarrow$ | 2 | | $\frac{7}{2}$ | $\rightarrow$ | 3 |

Table 3.4: Hamming distances of H-optimists and H-realists for the profile in Table 3.3.

*Proof.* Table 3.3 shows a profile in which at each stage one agent has an incentive to alter her vote as an H-realist or H-optimist. Table 3.4 lists the $optH$ and $realH$ distances of agent 1 and 2 proving that these agents have incentives to alter to the bold models. Since we have $\mathbf{\Gamma}^6 = \mathbf{\Gamma}^2$, the iteration of the *Approval* rule will not terminate. □

Note that, since an iteration process where agents are truth-biased admits more alterations than one without, Theorem 3.2 extends the result of Proposition 3.3 in which agents are truth-biased.

In the light of these negative result it is natural to ask which restrictions are needed in order to guarantee a termination. One might have noticed that the alterations in Table 3.3 are only possible because both agents $1$ and $2$ alter in a somewhat unexpected way. They submit goals whose models are further in Hamming distance from their truthful goal. This additional support for a model is the only reason why then the other agent can support this model even further, such that it gets approved. Consider agent 2's alteration from profile $\mathbf{\Gamma}^1$ to $\mathbf{\Gamma}^2$. She not only included $(111100)$ in her new goal, which is the model she wanted to support such that it will be approved, but she also supports $(100000)$. By doing so, she is adding a model that is closer to the first agent's goal than her own. The former agent then withdraws her vote from the just approved interpretations and supports the model that the second agent just added to her vote. This is how the iteration in Table 3.3 takes place. In the following section we restrict alterations, excluding ones of this type.

### 3.3.2 Weakly Truth-Biased Agents and Minimal Alterations

We want agents to only alter their vote such that it is effective for the change they want to create in the outcome, without further unnecessary alterations. We therefore define three *minimal alterations* to capture the basic changes an agent can do in order to have an effect on the outcome of the $Approval$ rule. The restricted alterations are: adding support to already winning alternatives, adding support to not yet winning alternatives or withdrawing support from winning alternatives. These changes are special forms of dilatation and erosion manipulation.

**Definition 3.3** (Minimal Alteration)**.** *Given a voting profile $\mathbf{\Gamma} = (\gamma_1, \dots, \gamma_n)$, an agent $i$ with an incentive to alter her vote performs a minimal alteration from stage $t$ to stage $t+1$ if one of the following three cases holds:*

*(1)* $Mod(\gamma_i^{t+1}) = Mod(\gamma_i^t) \cup \{v_1, \dots, v_l\}$ *with* $v_j \notin Mod(\gamma_i^t)$ *and it is the case that* $v_j \in Approval(\mathbf{\Gamma}^t)$ *for all* $j \in \{1, \dots, l\}$.

*(2)* $Mod(\gamma_i^{t+1}) = Mod(\gamma_i^t) \cup \{v_1, \dots, v_l\}$ *with* $v_j \notin Mod(\gamma_i^t)$ *and it is the case that* $supp_{\mathbf{\Gamma}^t}(v_j) = k_t - 1$ *for all* $j \in \{1, \dots, l\}$.

*(3)* $Mod(\gamma_i^{t+1}) = Mod(\gamma_i^t) \setminus \{v_1, \dots, v_l\}$ *with* $\{v_1, \dots, v_l\} \subseteq Approval(\mathbf{\Gamma}^t)$.

Assuming agents to only alter their vote in these ways, we restrict the iteration. This restriction however can be argued to be rational and in line with the agents being myopic.

In fact, since agents are only concerned about changing the current outcome to a better one, minimal alterations only allow actions that have an effect on these outcomes. Any effective change of an unrestricted best response can also be achieved through a best response which is of the kind (1)-(3) from Definition 3.3 or a combination of those. We do not allow additional changes that are ineffective at the time, but might have an effect on the iteration later on. This conforms with myopic agents as they are assumed to not care about changes further than the next iteration step. Considering minimal alterations only, we can exactly predict the outcome of an iteration under the $Approval$ rule, as shown by the following proposition.

**Proposition 3.4.** *Given a voting profile $\mathbf{\Gamma}^t$, an agent $i$'s minimal alteration (as per Definition 3.3) of a given kind will result in the following outcome:*

*(i) (1) leads to $Approval(\mathbf{\Gamma}^{t+1}) = \{v_1, \dots, v_l\}$.*

*(ii) (2) leads to $Approval(\mathbf{\Gamma}^{t+1}) = Approval(\mathbf{\Gamma}^t) \cup \{v_1, \dots, v_l\}$.*

*(iii) (3) leads to*

$$Approval(\mathbf{\Gamma}^{t+1}) = \begin{cases} Approval(\mathbf{\Gamma}^t) \setminus \{v_1, \dots, v_l\} \\ \quad \text{if } \{v_1, \dots, v_l\} \subset Approval(\mathbf{\Gamma}^t) \\ Approval(\mathbf{\Gamma}^t) \cup \{v \in \{0,1\}^m \mid supp_{\mathbf{\Gamma}^t}(v) = k_t - 1\} \\ \quad \text{if } \{v_1, \dots, v_l\} = Approval(\mathbf{\Gamma}^t). \end{cases}$$

*Proof.* In the following we will prove that the minimal alterations of kinds (1)-(3) as per Definition 3.3 will lead to the outcomes described by (i)-(iii). Recall that $k_t = supp_{\mathbf{\Gamma}^t}(v)$ for $v \in Approval(\mathbf{\Gamma}^t)$ is the support of the winning interpretations.

(i) Assume a profile $\mathbf{\Gamma}^t = (\gamma_1, \dots, \gamma_n)$ is such that agent $i$ has an incentive to alter her vote by (1). So then $\mathbf{\Gamma}^{t+1} = (\mathbf{\Gamma}^t_{,-i}, \gamma'_i)$ s.t. $Mod(\gamma'_i) = Mod(\gamma^t_i) \cup \{v_1, \dots, v_l\}$ for $v_j \notin Mod(\gamma^t_i)$ and $v_j \in Approval(\mathbf{\Gamma}^t)$ for all $j \in \{1, \dots, l\}$.

   For any interpretation $w \notin \{v_1, \dots, v_l\}$ it is the case that $supp_{\mathbf{\Gamma}^{t+1}}(w) = supp_{\mathbf{\Gamma}^t}(w)$ and $supp_{\mathbf{\Gamma}^{t+1}}(v_j) = supp_{\mathbf{\Gamma}^t}(v_j) + 1 = k_t + 1$ for all $j \in \{1, \dots, l\}$. Hence there are no other models with as much support as the $v_j$ and we have $k_{t+1} = supp_{\mathbf{\Gamma}^{t+1}}(v_j)$ for all $j \in \{1, \dots, l\}$. So it is the case that $Approval(\mathbf{\Gamma}^{t+1}) = \{v_1, \dots, v_l\}$.

(ii) Assume a profile $\mathbf{\Gamma}^t = (\gamma_1, \dots, \gamma_n)$ is such that agent $i$ has an incentive to alter her vote by (2). So then $\mathbf{\Gamma}^{t+1} = (\mathbf{\Gamma}^t_{,-i}, \gamma'_i)$ s.t. $Mod(\gamma^{t+1}_i) = Mod(\gamma^t_i) \cup \{v_1, \dots, v_l\}$ with $v_j \notin Mod(\gamma^t_i)$ and $supp_{\mathbf{\Gamma}^t}(v_j) = k_t - 1$ for all $j \in \{1, \dots, l\}$.

For any interpretation $w \notin \{v_1, \dots, v_l\}$ it is the case that $supp_{\mathbf{\Gamma}^{t+1}}(w) = supp_{\mathbf{\Gamma}^t}(w)$ and $supp_{\mathbf{\Gamma}^{t+1}}(v_j) = supp_{\mathbf{\Gamma}^t}(v_j) + 1 = k_t$ for all $j \in \{1, \dots, l\}$. Thus $k_{t+1} = k_t$, but additionally to the models in $Approval(\mathbf{\Gamma})$ also all $v_j$ for $j \in \{1, \dots, l\}$ now have this much support. Hence $Approval(\mathbf{\Gamma}^{t+1}) = Approval(\mathbf{\Gamma}^t) \cup \{v_1, \dots, v_l\}$.

(iii) Assume a profile $\mathbf{\Gamma}^t = (\gamma_1, \dots, \gamma_n)$ to be such that agent $i$ has an incentive to alter her vote by (3). This will lead to a profile $\mathbf{\Gamma}^{t+1} = (\mathbf{\Gamma}^t_{,-i}, \gamma_i^{t+1})$ such that $Mod(\gamma_i^{t+1}) = (Mod(\gamma_i^t) \setminus \{v_1, \dots, v_l\})$ with $\{v_1, \dots, v_l\} \subseteq Approval(\mathbf{\Gamma}^t)$.

For any interpretation $w \notin \{v_1, \dots, v_l\}$ we have $supp_{\mathbf{\Gamma}^{t+1}}(w) = supp_{\mathbf{\Gamma}^t}(w)$ and $supp_{\mathbf{\Gamma}^{t+1}}(v_j) = supp_{\mathbf{\Gamma}^t}(v_j) - 1 = k_t - 1$ for all $j \in \{1, \dots, l\}$. We now distinguish two cases:

- Assume $\{v_1, \dots, v_l\} = Approval(\mathbf{\Gamma}^t)$. Then, for all $w \in \{0,1\}^m$ we have $supp_{\mathbf{\Gamma}^t}(w) < k_t$. Thus $k_t - 1$ is the maximal support and $supp_{\mathbf{\Gamma}^{t+1}}(v_j) = k_t - 1$ for all $j \in \{1, \dots, l\}$. Therefore we have $Approval(\mathbf{\Gamma}^{t+1}) = Approval(\mathbf{\Gamma}^t) \cup \{v \in \{0,1\}^m \mid supp_{\mathbf{\Gamma}^t}(v) = k_t - 1\}$.

- Now assume $\{v_1, \dots, v_l\}$ is a proper subset of $Approval(\mathbf{\Gamma}^t)$. Then, we have $supp_{\mathbf{\Gamma}^{t+1}}(v_j) = k_t - 1$ for all $j \in \{1, \dots, l\}$ and there are still $w \in \{0,1\}^m$ with $supp_{\mathbf{\Gamma}^{t+1}}(w) = supp_{\mathbf{\Gamma}^t}(w) = k_t$. We have $k_{t+1} = k_t$ and thus $Approval(\mathbf{\Gamma}^{t+1}) = Approval(\mathbf{\Gamma}^t) \setminus \{v_1, \dots, v_l\}$.

This finalizes the proof of the proposition. □

Another restriction to ensure more realistic alterations will be called *weak truth-bias*. With this restriction we will assume all agents to always support their truthful goal's models. We call it a weak truth-bias since it could also be understood as agents always keeping their truthful goal as a disjunct in any alternative goal. Nevertheless, while the classic truth-bias (as per Definition 3.2) allows more possible alterations, the weak truth-bias is a real restriction on the set of alterations. In that sense the bias is not so much towards the truthful goal, as it is a bias on the choice within the set of best responses. In *Approval* voting with H-optimists, H-pessimists or H-realists any set of best responses includes at least one best response with a weak truth-bias, since still supporting her truthful goal will have no effect on an agent's alteration.

**Definition 3.4** (Weak Truth-Bias). *An agent $i$ is weakly truth-biased, if for any stage $t$ and altered goal $\gamma_i^t$ it is the case that $Mod(\gamma_i^0) \subseteq Mod(\gamma_i^t)$.*

A minimal alteration with a weak truth-bias will be of the form as described in Definition 3.3 where $Mod(\gamma_i^0) \subseteq Mod(\gamma_i^{t+1})$. This is not a restriction for alterations of kind $(1)$ and $(2)$, since here the model of the previous stages are always included in the new goal's models. For type $(3)$, on the other hand, it restricts the withdrawing of votes to interpretations which are not models of the agent's truthful goal, i.e., $v \notin Mod(\gamma_i^0)$. Consequently, type $(3)$ alterations cannot take place at $t = 0$.

These restrictions are enough to ensure termination of iterated $Approval$ voting with H-optimists, as will be shown in Theorem 3.4. However, H-realists are still able to alter their vote such that iteration can be circular.

**Theorem 3.3.** *Iterated* $Approval$ *voting might not terminate, if agents are H-realists, even with minimal alterations and weak truth-bias.*

*Proof.* Table 3.5 shows that there is a profile in which at each stage one of the agents has an incentive to alter their vote as an H-realist. Since $\mathbf{\Gamma}^{13} = \mathbf{\Gamma}^1$ we know that this profile would not terminate for the $Approval$ rule. The alterations are either adding interpretations to the outcome by an additional vote (alteration of kind (1) and (2)) or withdrawing this vote again, such that the model disappears from the outcome, which corresponds to an alteration of kind (3).

Note that agent 1 has an incentive to alter her vote from profile $\mathbf{\Gamma}^0$ to $\mathbf{\Gamma}^1$ since she can turn $realH(\gamma_1^0, Approval(\mathbf{\Gamma}^0)) = \frac{7}{6}$ into $realH(\gamma_1^0, Approval(\mathbf{\Gamma}^1)) = 1$, by an alteration of kind $(1)$, pushing the model of agent 2's goal to be the only outcome. From then on the same process repeats: another voter has now an incentive to push the model of the voter's goal (the one who just altered) into the outcome, so that there are two interpretations in the result. The voter who just altered before has now an incentive to withdraw her vote from the model she just added one stage earlier, so the outcome is a singleton again. This process repeats until every agent altered to vote a model in and out of the $Approval$ outcome and we arrive back at $\mathbf{\Gamma}^1$. □

This negative result for this type of agents can be explained by the fact that an H-realist, since she is always taking all the interpretations in the outcome into account, has more possibilities to alter in her interest. She has an incentive to push any interpretation into the outcome which is closer to her goal or withdraw the support of any that is further away than the current average result. The H-optimists (and H-pessimists) on the other hand are always only concerned with the best (or worse) interpretation in the outcome. They will get to a point at which there is no better interpretation to push into the outcome or worse one to eliminate.

| | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ | $\mathbf{\Gamma}^2$ | $\mathbf{\Gamma}^3$ | $\mathbf{\Gamma}^4$ | $\mathbf{\Gamma}^5$ | $\mathbf{\Gamma}^6$ | ... |
|---|---|---|---|---|---|---|---|---|
| $Mod(\gamma_1)$ | (0000) | **(0000)**<br>**(1000)** | (0000)<br>(1000) | **(0000)** | (0000) | (0000) | (0000) | |
| $Mod(\gamma_2)$ | (1000) | (1000) | (1000) | (1000) | (1000) | (1000) | (1000) | |
| $Mod(\gamma_3)$ | (1100) | (1100) | (1100) | (1100) | (1100) | (1100) | (1100) | |
| $Mod(\gamma_4)$ | (0100) | (0100) | (0100) | (0100) | (0100) | (0100) | **(0100)**<br>**(0110)** | |
| $Mod(\gamma_5)$ | (0110) | (0110) | (0110) | (0110) | **(0110)**<br>**(0010)** | (0110)<br>(0010) | (0110)<br>(0010) | |
| $Mod(\gamma_6)$ | (0010) | (0010) | **(0010)**<br>**(0000)** | (0010)<br>(0000) | (0010)<br>(0000) | **(0010)** | (0010) | |
| $Approval$ | (0000)<br>(1000)<br>(1100)<br>(0100)<br>(0110)<br>(0010) | (1000) | (1000)<br>(0000) | (0000) | (0000)<br>(0010) | (0010) | (0010)<br>(0110) | |

| | $\mathbf{\Gamma}^6$ | $\mathbf{\Gamma}^7$ | $\mathbf{\Gamma}^8$ | $\mathbf{\Gamma}^9$ | $\mathbf{\Gamma}^{10}$ | $\mathbf{\Gamma}^{11}$ | $\mathbf{\Gamma}^{12}$ | $\mathbf{\Gamma}^{13}$ |
|---|---|---|---|---|---|---|---|---|
| $Mod(\gamma_1)$ | (0000) | (0000) | (0000) | (0000) | (0000) | (0000) | **(0000)**<br>**(1000)** | (0000)<br>(1000) |
| $Mod(\gamma_2)$ | (1000) | (1000) | (1000) | (1000) | **(1000)**<br>**(1100)** | (1000)<br>(1100) | (1000)<br>(1100) | **(1000)** |
| $Mod(\gamma_3)$ | (1100) | (1100) | **(1100)**<br>**(0100)** | (1100)<br>(0100) | (1100)<br>(0100) | **(1100)** | (1100) | (1100) |
| $Mod(\gamma_4)$ | **(0100)**<br>**(0110)** | (0100)<br>(0110) | (0100)<br>(0110) | **(0100)** | (0100) | (0100) | (0100) | (0100) |
| $Mod(\gamma_5)$ | (0110)<br>(0010) | **(0110)** | (0110) | (0110) | (0110) | (0110) | (0110) | (0110) |
| $Mod(\gamma_6)$ | (0010) | (0010) | (0010) | (0010) | (0010) | (0010) | (0010) | (0010) |
| $Approval$ | (0010)<br>(0110) | (0110) | (0110)<br>(0100) | (0100) | (0100)<br>(1100) | (1100) | (1100)<br>(1000) | (1000) |

Table 3.5: Example of non-terminating iterative $Approval$ voting with weakly truth-biased H-realists using minimal alterations.

Lemma 3.3 shows that H-optimists only consider certain minimal alterations given the weak truth-bias: an H-optimist only has an incentive to alter if she can ensure that there will be a better interpretation in the new outcome, where *better* means a smaller $lowH$ distance to her goal. This causes an H-optimist to only consider an alteration of kind $(2)$, since, for example, an alteration of type $(1)$ restricts the new outcome to a subset of the old, which is not in favour of an H-optimist. This result will be used to prove Theorem 3.4, which is the main result of this section, stating that iterated $Approval$ voting always terminates if we assume weak truth-bias and minimal alterations for H-optimists.

**Lemma 3.3.** *A weak truth-biased H-optimist using minimal alterations in iterated Approval voting, only considers alterations of kind* $(2)$ *according to Definition 3.3.*

*Proof.* Take $\mathcal{N} = \{1, \ldots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \ldots, m\}$ a set of issues. Let $\mathbf{\Gamma}^0 = (\gamma_1, \ldots, \gamma_n)$ be the initial voting profile.

For an arbitrary $i \in \mathcal{N}$ take $v \in Approval(\mathbf{\Gamma}^0)$ with $optH(\gamma_i, Approval(\mathbf{\Gamma}^0)) = lowH(\gamma_i, v)$. By definition, as an H-optimist $i$ would only alter her vote to

$$
\begin{aligned}
&Approval(\mathbf{\Gamma}^{t+1}) \succ_i Approval(\mathbf{\Gamma}^t) \Leftrightarrow \\
&optH(\gamma_i, Approval(\mathbf{\Gamma}^{t+1})) < optH(\gamma_i, Approval(\mathbf{\Gamma}^t)) \\
&\Leftrightarrow \exists w \in Approval(\mathbf{\Gamma}^{t+1}) \text{ s.t. } lowH(\gamma_i, w) < lowH(\gamma_i, v) \\
&\Rightarrow \exists w \in Approval(\mathbf{\Gamma}^{t+1}) \text{ s.t. } w \notin Approval(\mathbf{\Gamma}^t)
\end{aligned}
$$

since $v$ is such that $lowH(\gamma_i, v)$ is minimal.

Note that, an alteration to $\mathbf{\Gamma}^{t+1}$ by $(1)$ as defined in Definition 3.3 will always cause $Approval(\mathbf{\Gamma}^{t+1}) \subseteq Approval(\mathbf{\Gamma}^t)$, as proven by Proposition 3.4. So, there is no interpretation $w \in Approval(\mathbf{\Gamma}^{t+1})$ such that $w \notin Approval(\mathbf{\Gamma}^t)$, therefore it is the case that $optH(\gamma_i, Approval(\mathbf{\Gamma}^{t+1})) \geq optH(\gamma_i, Approval(\mathbf{\Gamma}^t))$. Hence an H-optimist would never alter her vote by using alterations of type (1).

In case of an alteration type $(3)$ we have to distinguish two scenarios:

(i) If agent $i$ withdraws her support only from a proper subset of the outcome, i.e., $\{v_1, \ldots, v_l\} \subset Approval(\mathbf{\Gamma}^t)$, by Proposition 3.4 we get that the new outcome is $Approval(\mathbf{\Gamma}^{t+1}) = Approval(\mathbf{\Gamma}^t) \setminus \{v_1, \ldots, v_l\}$. Then $Approval(\mathbf{\Gamma}^{t+1})$ is a subset of $Approval(\mathbf{\Gamma}^t)$, and thus the argument from case $(1)$ applies.

(ii) The agent $i$ withdraws her vote from all interpretations $v \in Approval(\mathbf{\Gamma}^t)$. Then we get, $Approval(\mathbf{\Gamma}^{t+1}) = Approval(\mathbf{\Gamma}^t) \cup \{w \in \{0,1\}^m \mid supp_{\mathbf{\Gamma}^{t+1}}(w) = k_t - 1\}$, i.e., the new result is a superset of the old, additionally containing the models which were short of one support in stage $t$. By the weak truth-bias we can assume no agent withdraws her vote from her original goal's models. Hence $k_t$ will always be at least $k_0$. Note also that via an alteration at stage $t$ as described here, we have $k_{t+1} = k_t - 1$. For this alteration to be rational for an agent we therefore need $k_t > k_0$. But only an alteration of type $(1)$ causes $k_t$ to increase. However, no H-optimist would alter by $(1)$ as established above. So, for H-optimists we get $k_{t+1} = k_t$ and hence $k_t = k_0$ for any stage $t$ and so no H-optimist would alter her vote as $(3)$ suggests.

So since we have shown that a weak truth-biased H-optimist would not alter by (1) or (3), we can conclude that she only considers alterations of type (2). □

Lemma 3.3 shows that weak truth-biased H-optimists that choose minimal alterations are very restricted. These restrictions enable us to directly show that the $Approval$ rule terminates for this type of agents.

**Theorem 3.4.** *Iterated $Approval$ voting with weakly truth-biased H-optimists using minimal alterations, terminates after at most* $|\{w \in \{0,1\}^m \mid supp_{\mathbf{\Gamma}^0}(w) = k_0 - 1\}|$ *rounds.*

*Proof.* Take $\mathcal{N} = \{1, \ldots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \ldots, m\}$ a set of issues. Let $\mathbf{\Gamma}^0 = (\gamma_1, \ldots, \gamma_n)$ be the initial voting profile.

Let $W_t = \{w \in \{0,1\}^m \mid supp_{\mathbf{\Gamma}^t}(w) = k_t - 1\}$, be the models that could possibly be voted into the outcome of $Approval$ by an alteration of type $(2)$. By Lemma 3.3, this is the only kind of alteration we have to consider.

Observe that for any stage $t \in \mathbb{N}$ at which an alteration of type $(2)$ takes place we get $W_{t+1} \subset W_t$. This is so, since for any alteration of this kind at stage $t$, some models $\{v_1, \ldots, v_l\} \subseteq W_t$ will get one additional support $supp_{\mathbf{\Gamma}^{t+1}}(v) = supp_{\mathbf{\Gamma}^t} + 1(v)$. Thus $W_{t+1} = W_t \setminus \{v_1, \ldots, v_l\}$ and $Approval(\mathbf{\Gamma}^{t+1}) = Approval(\mathbf{\Gamma}^t) \cup \{v_1, \ldots, v_l\}$. Hence, $W_t$ properly decreases with each iteration step of $Approval$ with only H-optimists. Note that if $W_t = \emptyset$ no H-optimist has any incentive to alter anymore, therefore the iteration process will terminate after at most $|W_0|$ rounds. □

By Theorem 3.1 we already know that $Approval$ voting with H-pessimist is guaranteed to terminate even without restrictions. Hence it is immediate that by restricting the alterations, termination is still guaranteed.

**Corollary 3.2.** *Iterated $Approval$ voting with weakly truth-biased H-pessimists using minimal alterations always terminates.*

However, one might still ask if the minimal alteration restriction has an effect on the iteration of the *Approval* voting for H-pessimists. Lemmas 3.1 and 3.2 showed that H-pessimists only consider alterations which will increase the support of the winning interpretations. By Proposition 3.4 we can then see that H-pessimists would actually only consider minimal alterations of kind (1), since alterations of type (2) and (3) either decrease or don't change the support of the interpretations of the outcome. Restricting to weakly truth-biased agents only considering minimal alterations would have no effect on the iteration with H-pessimists. In fact, any unrestricted alteration is already a minimal alteration of kind (1) plus possibly some changes that have no effect on the process. Hence, for any unrestricted alteration we can find a similar restricted one which has the same effect on the iteration process. It also does not terminate in fewer steps with a restricted iteration, since any alteration step can take place exactly as it does in the unrestricted case. However, we have seen that this is not the case for the H-optimist (Theorem 3.4), since the restrictions are enough to ensure termination, while iteration might be circular with general alterations with the *Approval* rule for these agents (Theorem 3.3).

## 3.4 Discussion

In this chapter we explored the iterative voting process for different satisfaction functions using Majority rules and the *Approval* rule. In Section 3.2 we have seen that the *EMaj* and *TrueMaj* are generally not guaranteed to terminate, no matter which of the considered functions underlie the agents' satisfactions. On the other hand, the *Approval* rule gives a more diverse picture.

| | *EMaj* | *TrueMaj* | *2sMaj* |
|---|---|---|---|
| no truth-bias | ✗ | ✗ | ? |
| truth-bias | ✗ | ✗ | ✗ |
| | | *Approval* | |
| | Hopt | Hpess | Hreal |
| no truth-bias | ✗ | ✓ | ✗ |
| minimal alterations and weak truth-bias | ✓ | ✓ | ✗ |

Table 3.6: Overview of the results in Chapter 3.

Table 3.6 gives an overview of all results of this chapter. A guaranteed termination is marked with a check (✓), while a cross (✗) indicates a possible circular iteration with the respective satisfaction function, and a question mark (?) identifies an open problem. Note that there is no extra row dividing the Majority rules into the different satisfaction functions, since all the results in Section 3.2 apply to any satisfaction function, also including the dichotomous ones.

In *Approval* voting, termination of the iteration process highly depends on the satisfaction function considered. We have seen that *Approval* voting with H-pessimists (Theorem 3.1) is guaranteed to terminate, while we needed some minor restrictions on the alterations to guarantee termination for the H-optimists (Theorem 3.4). The *Approval* rule with H-realists, however, is still not guaranteed to terminate even with these restrictions (Theorem 3.3).

Endriss (2013) already showed the importance of the choice of satisfaction by analyzing best responses in classic approval voting under different preference extensions. He showed that the preference extension, which corresponds to our satisfaction functions, affect the set of best responses in manipulation. Thus, it is not surprising that we also found differences in the iteration procedure when considering different satisfactions. In our work we have effectively defined two possible sets of best responses: alterations without restrictions and minimal alterations with a weak truth-bias. Following Endriss' approach, it would be an interesting open problem to characterize what types of satisfaction functions cause the iteration to terminate under these assumptions.

For the Majority rule we have observed no differences in the satisfaction functions, since the voting rules themselves seem to admit large possibilities to alterations and thus circular iterations. Note that *TrueMaj* and *EMaj* are quota rules, i.e., an issue gets rejected or accepted based on the ratio of approvals and disapprovals in the given profiles. Considering a higher threshold might raise the bar for alteration and yield a more positive result for iteration. Also, considering Novaro (2019) result about the strategy-proofness of the Majority rules under different restrictions on the goals' language, it is natural to assume that some weaker restriction on the language will lead to a positive termination result without hindering iteration completely.

In conclusion, we have seen in this chapter non-termination results that we labelled as negative. However, one might ask how realistic these results are. In fact, most of these are established by specific counterexamples, which were found by carefully selecting and adapting profiles and even pushing them into absurd narratives. One question worth asking is: How likely circular iterations are?

In the next Chapter we address this question, by implementing the iterative *Approval* voting into a Python program. Based on the data collected from this implementation, we will build first hypothesis on how likely circularity and iteration are in general. Additionally, we will analyze the quality of the iteration, based on a notion of social welfare.

# 4 | Social Welfare

In the first part of this chapter we will define a notion of social welfare and analyze the quality of iteration for $Approval$ voting and Majority voting in the light of it. The second part contains an analysis of the data obtained by implementing iterative $Approval$ voting, with randomly chosen profiles. The goal of this chapter is to formulate some hypothesis about how often iteration takes place and how beneficial it is for the agents involved.

## 4.1 Social Welfare in Iteration

Satisfaction measures based on different generalizations of the Hamming distance have been defined and discussed in Section 2.2.2. In this section we want to establish a notion of group satisfaction, based on the individual agents' satisfaction. In the literature this is referred to as *social welfare*.

Note that there are multiple notions of social welfare. While we are following an utilitarian (additive) approach one could also consider the Nash (multiplicative) social welfare as studied by DeMeyer and Plott (1971) and Kaneko and Nakamura (1979) or the egalitarian social welfare as introduced by Rawls (1971). In the literature the *utilitarian social welfare* is defined as the sum over all the agents' utilities (Caragiannis and Procaccia, 2011; Barrot et al., 2017). Recall that, in Section 2.2 we have defined a notion of satisfaction for an agent based on the distance functions $optH$, $pessH$ and $realH$. The social welfare will be defined as the sum over all satisfactions.

**Definition 4.1** (Social Welfare)**.** *For a set of agents $\mathcal{N} = \{1, \dots, n\}$ and a set of issues $\mathcal{I} = \{1, \dots, m\}$, the social welfare for a truthful profile $\mathbf{\Gamma}$ and the outcome of a voting rule $F : (\mathcal{G})^n \to \mathcal{P}(\{0,1\}^m) \setminus \{\emptyset\}$ under a profile $\mathbf{\Gamma}'$ is defined as:*

$$sw(F(\mathbf{\Gamma}'), \mathbf{\Gamma}) = \sum_{i \in \mathcal{N}} sat_i(\gamma_i, F(\mathbf{\Gamma}'))$$

We will call the social welfare of H-optimists, H-pessimists and H-realists $sw_o$, $sw_p$

and $sw_r$ respectively. Then for a truthful profile $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_n)$ by Definition 4.1 we get the following social welfare:

$$sw_o(F(\mathbf{\Gamma}'), \mathbf{\Gamma}) = \sum_{i \in \mathcal{N}} (m - optH(\gamma_i, F(\mathbf{\Gamma}')))$$
$$sw_p(F(\mathbf{\Gamma}'), \mathbf{\Gamma}) = \sum_{i \in \mathcal{N}} (m - pessH(\gamma_i, F(\mathbf{\Gamma}')))$$
$$sw_r(F(\mathbf{\Gamma}'), \mathbf{\Gamma}) = \sum_{i \in \mathcal{N}} (m - realH(\gamma_i, F(\mathbf{\Gamma}'))) = \sum_{i \in \mathcal{N}} (m - \frac{\sum_{w \in F(\mathbf{\Gamma}')} lowH(\gamma_i, w)}{|F(\mathbf{\Gamma}')|})$$

Note the social welfare ranges between $0$ and $n \cdot m$ for any distance we consider. A higher value of social welfare means a higher satisfaction for the group. Note that the social welfare only equals $n \cdot m$ if every agent $i$ has distance $d_i = 0$. In case of H-optimists, this means that for each agent there is at least one interpretation in the outcome that models her goal, while for H-pessimists and H-realists it means that all interpretations in the outcome model all agents' goals.

### 4.1.1 Optimality Results

In goal-based voting one might reasonably think that, if all agents' goals have at least one model in common, these interpretations build the best outcome. The following notion of an outcome being *optimal* captures exactly this idea.

**Definition 4.2** (Optimal)**.** *Given a truthful profile $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_n)$ and a voting rule $F$ we call the outcome $F(\mathbf{\Gamma}')$ under a profile $\mathbf{\Gamma}'$ optimal if $\bigcap_{i \in \mathcal{N}} Mod(\gamma_i) \neq \emptyset$ implies $F(\mathbf{\Gamma}') \subseteq \bigcap_{i \in \mathcal{N}} Mod(\gamma_i)$.*
*A rule is optimal if it always returns an optimal outcome if such exist.*

Note the difference between this notion of optimality and being Pareto optimal. An outcome is called *Pareto optimal*, if it is not dominated by any other outcome, where an outcome $o$ is dominated by $o'$, if there is an agent who strictly prefers $o'$ over $o$ and no agent prefers $o$ over $o'$. An outcome being optimal as defined by Definition 4.2 is stronger: it requires that no interpretation in the outcome is dominated, and thus is also Pareto optimal, but additionally demands all interpretations to be a model of each agents' goal. Therefore, if the optimal outcome exists it also is Pareto optimal. But in case the intersection of all agents' models is empty, there is no optimal, but there might still be a Pareto optimal outcome. Additionally we will call an alteration or step in the iteration a *Pareto improvement*, if none of the agents' satisfaction decreased and at least one agent's satisfaction increased.

A rule being optimal is closely related to the *model unanimity* axiom as defined by Novaro (2019), stating that voting rules satisfying this axiom will return the optimal outcome, which consists of exactly all interpretations in the intersection of all goal's models if they exist. Our notion of a rule being optimal coincides with one direction of model unanimity, only requiring the outcome to be a subset of the set of these interpretations. Since Novaro (2019) showed that the $\mathit{Approval}$ rule satisfies model unanimity axiom, we know that it guarantees to return an optimal outcome. For the Majority rules, however, this may not be the case, as Example 4.1 shows.

**Example 4.1.** *Consider three agents and three issues and let their voting profile be as in Table 4.1. All the agents hold the same goal, which has two models. Note that any outcome including only* $(010)$ *or* $(101)$ *is optimal.*

*Every agent gives a support of* $\frac{1}{2}$ *towards each issue. Hence each issue* $j \in \mathcal{I}$ *ties within one agent's goal. Due to its tie-breaking rule,* $\mathit{EMaj}$ *favours* $0$ *over* $1$ *and hence will return* $(000)$ *as the only interpretation in the outcome. The same is true for* $\mathit{2sMaj}$ *since it computes* $\mathit{EMaj}$ *for each agent's goal.* $\mathit{TrueMaj}$ *on the other hand, does not have a tie-breaking rule and thus returns the whole set of interpretations. Only the* $\mathit{Approval}$ *rule chooses an optimal outcome.*

| | $\mathbf{\Gamma}$ |
|---|---|
| $\mathit{Mod}(\gamma_1)$ | $(010)$ |
| | $(101)$ |
| $\mathit{Mod}(\gamma_2)$ | $(010)$ |
| | $(101)$ |
| $\mathit{Mod}(\gamma_3)$ | $(010)$ |
| | $(101)$ |
| $\mathit{E/2sMaj}$ | $(000)$ |
| $\mathit{TrueMaj}$ | $\{0,1\}^3$ |
| $\mathit{Approval}$ | $(010)$ |
| | $(101)$ |

Table 4.1: Example of the Majority rules not choosing the optimal result.

Observe that optimal results always yield the highest possible social welfare of $n \cdot m$, since any interpretation in the outcome is a model of all agents' goals and hence it has a $\mathit{lowH}$ distance of $0$ to any agent. Therefore, the social welfare will be maximal for any satisfaction function.

One might note that, in case the agents are H-optimists the result of the $\mathit{TrueMaj}$ still yields the best possible social welfare. This holds, since all interpretations are in the outcome, for every agent there is at least one interpretation which is a model of her goal. Hence, this model will have lowest Hamming distance $0$ and the social welfare for H-optimists will be maximal. However, the result also includes many more interpretations which are not models of the agent's goal. While these additional interpretations often build some kind of compromise, in this case they are redundant, as they are all dominated by the models of the agents' goals.

For all of these rules, agents basing their satisfaction on any of the dichotomous functions, on the $optH$ or $realH$ (and for $EMaj$ and $2sMaj$ also the $pessH$) have an incentive to alter their vote. In fact, in Example 4.1, iteration would terminate with the optimal outcome: if one of the agents alters to one of her goal's models, any of the rules (except the $2sMaj$) would return this instead and they would reach the maximal social welfare. The good news is that iteration can help to reach the optimal result. Unfortunately, it is not guaranteed to do so, as we will see in Example 4.2.

**Example 4.2.** *Consider three agents and three issues with the profile as presented in Table 4.2. None of the Majority rules would result in the optimal outcome* $(100)$. *And since agent* $1$ *can alter to* $(110)$, *the iteration terminates at another non-optimal outcome.*

| | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ |
|---|---|---|
| $Mod(\gamma_1)$ | $(101)$ | $\mathbf{(110)}$ |
| | $(010)$ | |
| | $(110)$ | |
| | $(001)$ | |
| $Mod(\gamma_2)$ | $(010)$ | $(010)$ |
| | $(101)$ | $(101)$ |
| $Mod(\gamma_3)$ | $(010)$ | $(010)$ |
| | $(101)$ | $(101)$ |
| $EMaj$ | $(000)$ | $(110)$ |
| $TrueMaj$ | $\{0,1\}^3$ | $(110)$ |

Table 4.2: Example of the $EMaj$ and $TrueMaj$ rules not terminating to the optimal result after iteration.

This profile could have terminated in an optimal outcome, if either another agent altered her vote or agent $1$ decided to be more mindful of the other agents' goals. This example shows that this is nevertheless not guaranteed. Note that the $2sMaj$ would also yield $(000)$ as the outcome of profile $\mathbf{\Gamma}^0$; however, Example 4.2 is not enough to also show that the $2sMaj$ would not terminate at the optimal outcome in case iteration starts, since here no agent has the power to change the outcome with an alteration. Whether or not $2sMaj$ always terminates at the optimal outcome in iteration, is an open question.

### 4.1.2 Social Welfare with the Approval Rule

As we have seen in Section 4.1.1, the $Approval$ rule is guaranteed to yield the optimal outcome, if it exists. However, there are multiple profiles which do not have an optimal outcome. In these cases, iteration might occur. In this section we will show that the social welfare will never decrease considering the iterative $Approval$ rule with H-optimists and H-pessimists. For H-realists, on the other hand, the picture is not as clear, as Proposition 4.2 shows that the social welfare considering H-realists with the $Approval$ rule, might increase, decrease or stay constant in iteration.

**Theorem 4.1.** *In the iteration of the $Approval$ rule with H-pessimists, the social welfare always increases.*

*Proof.* We will prove this theorem by showing that any iteration step is a Pareto improvement. Take $\mathcal{N} = \{1, \dots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \dots, m\}$ a set of issues. Let $\mathbf{\Gamma}^0 = (\gamma_1, \dots, \gamma_n)$ be the initial profile.

By Lemmas 3.1 and 3.2 we know that for any iteration step with H-pessimists, we have $k_{t+1} > k_t$ and hence $Approval(\mathbf{\Gamma}^{t+1}) \subseteq Approval(\mathbf{\Gamma}^t)$ at any stage $t$. This directly gives us that the interpretations with the maximal lowest Hamming distance ($lowH$) to the goal of any given agent $i$ are either still in the new outcome, or the distance decreased. This means we have $pessH(i, Approval(\mathbf{\Gamma}^{t+1})) \leq pessH(i, Approval(\mathbf{\Gamma}^t))$. To see that this is the case, assume there is an agent $i$ whose distance increased, so we have $pessH(i, Approval(\mathbf{\Gamma}^{t+1})) > pessH(i, Approval(\mathbf{\Gamma}^t))$. Then since it is the case that $Appoval(\mathbf{\Gamma}^{t+1}) \subseteq Approval(\mathbf{\Gamma}^t)$ the interpretation that holds the lowest distance $pessH(i, Approval(\mathbf{\Gamma}^{t+1}))$ must have been in $Approval(\mathbf{\Gamma}^t)$ before and therefore $pessH(i, Approval(\mathbf{\Gamma}^t))$ was not the maximal distance, which contradicts the definition of $pessH$ (see Definition 2.12).

Additionally we know that the altering agent $j$ changes her goal such that her satisfaction strictly increases, i.e., $pessH(j, Approval(\mathbf{\Gamma}^{t+1})) < pessH(j, Approval(\mathbf{\Gamma}^t))$. This gives us that at any step $t$

$$\sum_{i \in \mathcal{N}} (m - pessH(i, Approval(\mathbf{\Gamma}^{t+1}))) > \sum_{i \in \mathcal{N}} (m - pessH(i, Approval(\mathbf{\Gamma}^t))).$$

Therefore, it is the case that $sw_p(Approval(\mathbf{\Gamma}^{t+1}), \mathbf{\Gamma}^0) > sw_p(Approval(\mathbf{\Gamma}^t), \mathbf{\Gamma}^0)$ for any stage $t$. This shows that any step is a Pareto improvement. Since by Theorem 3.1 the iteration with H-pessimists always terminates we know that the final stage will yield an outcome that has a higher social welfare than the outcome under the initial profile. □

**Theorem 4.2.** *In case the iteration of the $Approval$ rule with H-optimists terminates, the social welfare does not decrease.*

*Proof.* Take $\mathcal{N} = \{1, \dots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \dots, m\}$ a set of issues. Let $\mathbf{\Gamma}^0 = (\gamma_1, \dots, \gamma_n)$ be the initial profile whose outcome is $Approval(\mathbf{\Gamma}^0)$. We show the theorem by proving the following lemma:

**Lemma 4.1.** *For H-optimists at any stage $t$ we have $Approval(\mathbf{\Gamma}^0) \subseteq Approval(\mathbf{\Gamma}^t)$.*

Assume this lemma to be true, then at any possible stage $t$ for any agent $i$ the lowest possible satisfaction coincides with the highest possible $optH(i, Approval(\mathbf{\Gamma}^t))$ which will never drop below $optH(i, Approval(\mathbf{\Gamma}^0))$. This is so since the best interpretation in the initial outcome is guaranteed to be included in any later outcome as well. Therefore we have $optH(i, Approval(\mathbf{\Gamma}^t)) \leq optH(i, Approval(\mathbf{\Gamma}^0))$ for any agent $i$ and any stage $t$. This gives us that

$$\sum_{i \in \mathcal{N}} optH(i, Approval(\mathbf{\Gamma}^t)) \leq \sum_{i \in \mathcal{N}} optH(i, Approval(\mathbf{\Gamma}^0)).$$

Therefore, by the definition of social welfare we get that any later stage yields a higher satisfaction for the group, i.e., $sw_o(Approval(\mathbf{\Gamma}^t), \mathbf{\Gamma}^0) \leq sw_o(Approval(\mathbf{\Gamma}^0), \mathbf{\Gamma}^0)$ for any stage $t$. Thus, in case the iteration terminates it will lead to an outcome with the same or a higher social welfare. Hence, it suffices to prove Lemma 4.1.

*Proof of Lemma 4.1.* Note that any agent $i$ who includes a winning interpretation in her current goal, which also models her truthful goal, has no incentive to alter at this stage. This is the case since H-optimists only consider the minimal $lowH$, which in this case is 0 and hence cannot be decreased. This also gives us that the support of these models will not decrease, i.e., $supp_{\mathbf{\Gamma}^t}(v) \geq supp_{\mathbf{\Gamma}^0}(v)$ for any stage $t$ and all interpretations $v \in Approval(\mathbf{\Gamma}^0)$.

Further, note that the support of the winning interpretations from one stage $t$ to the next will not increase, i.e., $k_{t+1} \leq k_t$. Suppose this would not be the case, so there is an iteration step such that $k_{t+1} > k_t$. Since support of interpretations can only be raised by 1 per step we have $k_{t+1} = k_t + 1$. Hence, the new outcome must have been a subset of the previous $Approval(\mathbf{\Gamma}^{t+1}) \subseteq Approval(\mathbf{\Gamma}^t)$. We would have $optH(i, Approval(\mathbf{\Gamma}^{t+1})) \geq optH(i, Approval(\mathbf{\Gamma}^t))$, hence no H-optimist has an incentive to induce such a change, because one of their closest interpretations is either still in the outcome, which makes no difference to them, or now a worse interpretation is their closest, which would make them worse off. Therefore, the support of the winning interpretations from one stage $t$ to the next will not increase, i.e., $k_{t+1} \leq k_t$.

In conclusion, the initial winning interpretations will have a stable support of $k_0$ and the support of the winning interpretations $k_t$ does not increase during the iteration, i.e., $k_t = k_0$ for all stages $t$. Therefore, the interpretations $v \in Approval(\mathbf{\Gamma}^0)$ will always hold the maximal support. Hence $Approval(\mathbf{\Gamma}^0) \subseteq Approval(\mathbf{\Gamma}^t)$. □

Thus, social welfare for H-optimists does not decrease in non-circular iteration. □

Note that we only considered profiles which terminate, since a circular iteration, while possibly yielding changes in social welfare, has no end profile whose social welfare can be compared to the initial one. The following proposition shows that in case of weakly truth-biased H-optimists who only consider minimal alterations, we get that each iteration step is a Pareto improvement.

**Proposition 4.1.** *The iterated $Approval$ rule with weakly truth-biased H-optimists considering only minimal alterations, always yields an increase in social welfare.*

*Proof.* We will prove this theorem by showing that any iteration step is a Pareto improvement. Take $\mathcal{N} = \{1, \ldots, n\}$ to be a set of voters and $\mathcal{I} = \{1, \ldots, m\}$ a set of issues. Let $\mathbf{\Gamma}^0 = (\gamma_1, \ldots, \gamma_n)$ be the voting profile.

An iteration with weakly truth-biased H-optimists who only consider minimal alterations is known by Lemma 3.3 to only include alterations of kind (2) from Definition 3.3. That is, any agent $i$ will only alter such that $Mod(\gamma_i^{t+1}) = Mod(\gamma_i^t) \cup \{v_1, \ldots, v_l\}$ with $v_j \notin Mod(\gamma_i^t)$ and $supp_{\mathbf{\Gamma}^t}(v_j) = k_t - 1$ for all $j \in \{1, \ldots, l\}$.

By Theorem 3.4 we know that this iteration will terminate and hence by Theorem 4.2 it will not yield a worse outcome. Additionally, we know that the altering agent's satisfaction strictly increases. This means that for any stage $t$, all agents are at least as satisfied as before, which gives us $optH(i, Approval(\mathbf{\Gamma}^{t+1})) \leq optH(i, Approval(\mathbf{\Gamma}^t))$, and at least for one the altering agent ($j \in \mathcal{N}$) the satisfaction increases: $optH(j, Approval(\mathbf{\Gamma}^{t+1})) < optH(j, Approval(\mathbf{\Gamma}^t))$. This is enough to see that any step is a Pareto improvement and leads to:

$$\sum_{i \in \mathcal{N}} (m - optH(i, Approval(\mathbf{\Gamma}^{t+1}))) > \sum_{i \in \mathcal{N}} (m - optH(i, Approval(\mathbf{\Gamma}^t))).$$

Therefore, $sw_o(Approval(\mathbf{\Gamma}^{t+1}), \mathbf{\Gamma}^0) > sw_o(Approval(\mathbf{\Gamma}^t), \mathbf{\Gamma}^0)$ at any stage $t$. □

These results cannot be extended to *Approval* voting with H-realists. The next proposition shows that we can find examples for an increase, decrease and constant social welfare through iteration.

**Proposition 4.2.** *In case the iteration of the Approval rule with H-realists terminates, the social welfare might increase, decrease or stay the same.*

*Proof.* The following examples in Table 4.3 show that social welfare in iterated *Approval* rule with H-realists might increase, decrease or stay constant.

| | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ | | $\mathbf{\Gamma}'^0$ | $\mathbf{\Gamma}'^1$ | | $\mathbf{\Gamma}''^0$ | $\mathbf{\Gamma}''^1$ |
|---|---|---|---|---|---|---|---|---|
| $Mod(\gamma_1)$ | (000) | **(010)** | $Mod(\gamma_1)$ | (010) | **(000)** | $Mod(\gamma_1)$ | (000) | **(010)** |
| $Mod(\gamma_2)$ | (011) | (011) | $Mod(\gamma_2)$ | (000) | (000) | $Mod(\gamma_2)$ | (010) | (010) |
| | (010) | (010) | | (111) | (111) | | (111) | (111) |
| $Mod(\gamma_3)$ | (011) | (011) | $Mod(\gamma_3)$ | (111) | (111) | $Mod(\gamma_3)$ | (111) | (111) |
| | (010) | (010) | | | | | | |
| *Approval* | (011) | (010) | *Approval* | (111) | (111) | *Approval* | (111) | (111) |
| | (010) | | | | (000) | | | (010) |

Table 4.3: Example of social welfare increasing, decreasing and being constant for H-realists under the *Approval* rule.

The social welfare of these H-realists from Table 4.3 are as follows: for the first table we have $sw_r(Approval(\mathbf{\Gamma}^0), \mathbf{\Gamma}^0) = 1.5 + 3 + 3 = 7,5$ for the first profile and a higher social welfare of $sw_r(Approval(\mathbf{\Gamma}^1), \mathbf{\Gamma}^0) = 2 + 3 + 3 = 8$ for the second profile. The profiles in the second table yield decreasing social welfare, $sw_r(Approval(\mathbf{\Gamma}'^0), \mathbf{\Gamma}'^0) = 1 + 3 + 3 = 7$ and $sw_r(Approval(\mathbf{\Gamma}'^1), \mathbf{\Gamma}'^0) = 1.5 + 3 + 1.5 = 6$. The last table gives us a social welfare of $sw_r(Approval(\mathbf{\Gamma}''^0), \mathbf{\Gamma}''^0) = 0 + 3 + 3 = 6 = 1 + 3 + 2 = sw_r(Approval(\mathbf{\Gamma}''^1), \mathbf{\Gamma}''^0)$. This concludes the proof, that iterative *Approval* voting can yield a higher, lower or constant social welfare. □

One could consider a result *better for the group of agents* engaged in the voting process, if it yields a higher social welfare. In this sense we have seen that iteration with the *Approval* rule with H-optimists and H-realists will lead to better outcomes while the Majority rules and the *Approval* rule with H-realists might not. In the next section, we want to get a glimpse on how likely it is for an iteration to result in a worse outcome. Using an implementation of the iterative *Approval* voting with H-realists, we will see that iteration seem to be rare, but worth considering.

## 4.2 Implementation of Iterative Approval Voting

A lot of results in this thesis showed that there are multiple instances in which the termination of an iterative goal-based voting process is not guaranteed. For example $Approval$ voting possibly yields circular iterations as shown in Theorem 3.2. However, the counterexamples of these seemingly negative results often use very special profiles to construct a circular iteration. Iterative voting might still be considered profitable to the agents if circularity occurs rarely and the social welfare increases regularly with iteration. In order to shine some light on this question we implemented a program which iterates goal-based voting and computes the social welfare for each visited profile.

The program is written in Python, and it is for the $Approval$ rule with H-realists. This satisfaction type seems to be the most interesting and least explored, since this is the only satisfaction function for which no definite theoretical result could be reached, neither for termination of the considered rules, nor for their social welfare in iteration. The iterations are run for profiles chosen uniformly at random. One iteration step consists of computing the set of agents which have an incentive to alter, by computing the agent's satisfaction under the current profile and comparing it to the satisfaction of outcomes under possible variants. If the set of agents who benefit from alteration is non-empty, the program first chooses one of these agents and then one of their best responses randomly. Like this we ensure that there is no bias towards a specific agent or response. Since we ran the program multiple times, one profile could have been visited multiple times. Note, however, that since we chose the altering agents and their best response randomly, visiting the same profile more often can still increase our insight, if then other agents or responses are chosen in the iteration.

We ran the program with $n = 3$ agents and $m = 3$ issues. Although this is a restriction to a special case, we were forced to choose a minimal implementation due to computational restrictions. If one were to add more issues, the exponent in the number of models ($2^m$) and so also in the goals ($2^{2^m}$) would increase. In fact, observe that already with these numbers of agents and issues we are faced with $2^8 - 2 = 254$ goals and thus $254^3 = 16.387.064$ possible profiles (though if one were to implement more experiments, a way to reduce the search space would be to remove symmetrical profiles). Note that there are $8$ possible models for $3$ issues and that goals are either including or excluding each of these models, so since we do not allow neither contradictions nor tautologies as goals, the number of goals is calculated as above. While technically allowed by the framework of goal-based voting, we decided to exclude tautologies from the set of goals, as in $Approval$ voting an agent who approves every model is basically absent from the vote.

With this implementation we will observe the frequency as well as the quality of iterations. We have analyzed $500$ randomly chosen profiles, which, while not significant (considering the total number of possible profiles), gives us a first impression on the question raised, in order to formulate some weak hypothesis and identify interesting further research directions.

### 4.2.1 Frequency of Iteration

In this section we discuss how often iteration actually took place in the $500$ sample profiles of our implementation. We will also look at some examples of those to understand how profiles that yield iterations commonly look like.

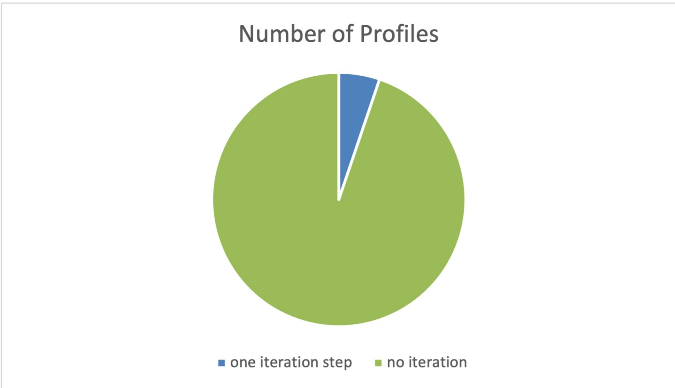


Figure 4.1: Number of profiles iterated with $Approval$ rule and H-realists.

Figure 4.1 shows that only very few profiles actually started iteration. In our case only $26$ out of $500$ and thus $5.2\%$ of the profiles did and all of these took exactly one step to terminate, except for one profile which took two iteration steps.

Looking at these profiles carefully, one can recognize most of them to follow a similar pattern: there are at least two interpretations in the outcome with support $2$, hence two agents agreeing on at least two interpretations, while the third agent's goal has no common model with these two. Some of the interpretations are more preferred by this third agent than the others. This causes the third agent to have an incentive to manipulate. An example can be seen in Table 4.4.

Not all of the iterated profiles do look like this. One can also find profiles as the one which took two iteration steps, presented in Table 4.5. Here, the support $k_0$ for the approved interpretations is only one. This means that the models of any two agents' goals are disjoint and therefore all of them will be in the outcome. In this case, an agent might have an incentive to alter her vote, in case there are more undesirable interpretations in

| | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ |
|---|---|---|
| $Mod(\gamma_1)$ | $(101)$ | $\mathbf{(011)}$<br>$\mathbf{(100)}$ |
| $Mod(\gamma_2)$ | $(000)$<br>$(100)$<br>$(111)$ | $(000)$<br>$(100)$<br>$(111)$ |
| $Mod(\gamma_3)$ | $(000)$<br>$(100)$ | $(000)$<br>$(100)$ |
| $Approval(\mathbf{\Gamma})$ | $(000)$<br>$(100)$ | $(100)$ |

Table 4.4: Example of iterated $Approval$ voting with H-realists.

the outcome than there are ones she likes. Then, this agent can choose to support one of the other agents' models. This causes her own models to not be in the outcome anymore, but since we are assuming H-realists, this might still yield a lower average distance for the altering agent. In the second alteration agent $1$ was able to support some interpretations such that her average satisfaction will increase. This alteration coincides with a minimal alteration of type (2) of Definition 3.3.

### 4.2.2 Quality of Iteration

In this section we inspect the social welfare in iteration. In order to analyze it, we have computed the social welfare for each profile visited during the iteration process. The following figures present the results of our implementation.

In Figure 4.2 we only consider the profiles which yield iteration. It shows that almost $70\%$ of the iterated profiles were such that the social welfare increased with the iteration. For about $15\%$ of the profiles the social welfare decreased. Due to the low number of inspected profiles, we cannot draw a definite conclusion. However, this is an indication that iterations is beneficial, hence running more test iterations is worth exploring.

In Figure 4.3 one can see that none of the iterated profiles reached the maximal social welfare of $9$, while almost $70\%$ of the initial profiles did. About $65\%$ of the iterated profiles reached social welfare of $8$. The iterated profiles have another smaller peak at social welfare $7$ with roughly $25\%$ of the profiles terminating at it. The initial profiles also have one more value over $20\%$ which is at social welfare $8$. Any other value has only been reached by less than $10\%$ of the profiles considered. The average final social

| | $\mathbf{\Gamma}^0$ | $\mathbf{\Gamma}^1$ | $\mathbf{\Gamma}^2$ |
|---|---|---|---|
| $Mod(\gamma_1)$ | (111) | (111) | **(011)**<br>**(100)**<br>**(101)**<br>**(110)** |
| $Mod(\gamma_2)$ | (000) | **(000)**<br>**(001)** | (000)<br>(001) |
| $Mod(\gamma_3)$ | (001)<br>(010)<br>(011)<br>(101)<br>(110) | (001)<br>(010)<br>(011)<br>(101)<br>(110) | (001)<br>(010)<br>(011)<br>(101)<br>(110) |
| $Approval(\mathbf{\Gamma})$ | $Mod(\bigvee_{i \in \mathcal{N}} \gamma_i)$ | (001) | (001)<br>(011)<br>(101)<br>(110) |

Table 4.5: Examples of iterated *Approval* voting with H-realists with two steps.

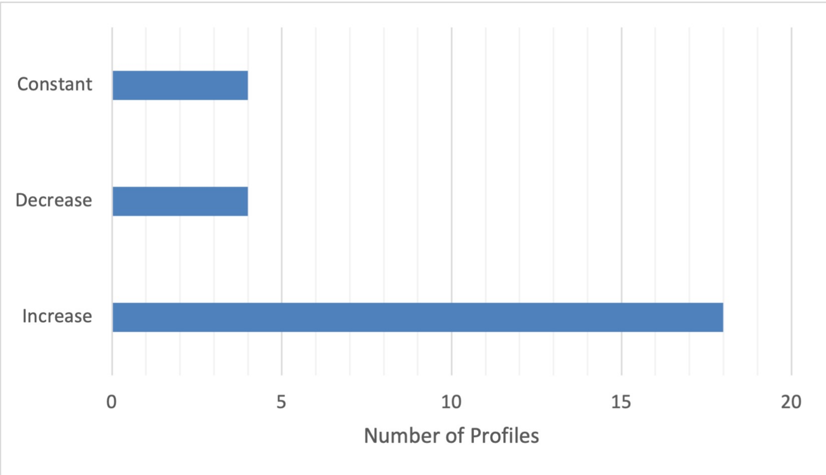


Figure 4.2: Change in social welfare for profiles where iteration took place with *Approval* voting and H-realists.

welfare after iteration (including the social welfare of the profiles who did not iterate) was 8.619 while in the initial profile the average was 8.598. As one can tell, this difference is very small. When only considering the profiles which did take an iteration step, the initial average of the social welfare was 7.259 and the final one was 7.664.

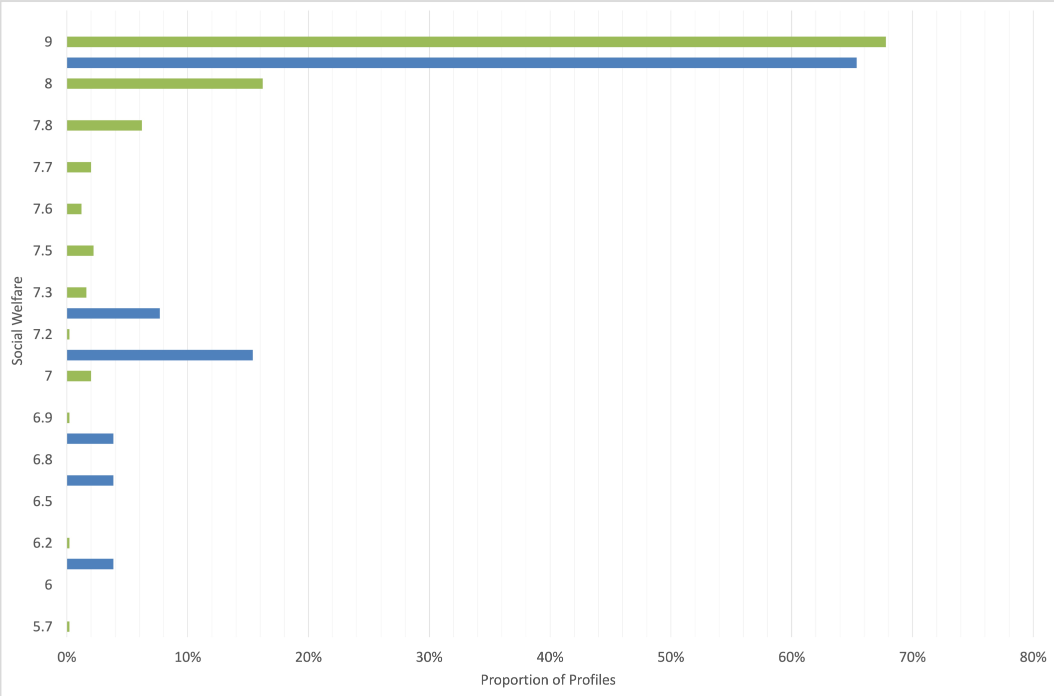


Figure 4.3: Social welfare for *Approval* voting with H-realists. The blue bars represent social welfare for profiles with iteration, the green bars for profiles without iteration.

## 4.3 Discussion

In our implementation only very few profiles ($5.2\%$) iterated and none of them were circular. This gives a first impression that termination might be very likely after all. However, note that our dataset itself does not even cover close to $1\%$ of the possible profiles due to the double exponential number of profiles. One has to keep this in mind when analyzing this data.

Figure 4.2 indicates that iteration might be beneficial for social welfare. Almost $70\%$ of the iterated profiles lead to an increase of the social welfare, although Figure 4.3 shows that the average social welfare after iteration is close to the average initial one. Most of the initial profiles hold the highest possible social welfare, while none of the iterated ones do. One may conclude that, since iteration also yields some costs (computational, effort, etc.) it might not be worth it.

However, Figure 4.2 might be misleading: since many of the initial profiles have such a high social welfare to begin with, iteration does not appear often. If for example a profile yields a maximal social welfare of $9$, none of the agents has an incentive to alter her vote, since all of them hold the highest utility. The reason why the differences of averages ($8.619$ compared to $8.598$) is small can be explained by the fact that the profiles which do iterate are such a small fraction of the number of all profiles. In order to study this in more depth one needs to run more implementations of this form for more agents and issues.

Note also that we built a deadline into our implementation which would have stopped the iteration if it would have exceeded a certain number of stages. In our case the bound was set to 10 steps and hence never triggered. Assuming such a bound on the iteration steps seems natural since most decisions have a date by which they must be taken. This has been considered before, for example, Airiau and Endriss (2009) study iterated Majority voting with a bound.

# 5 | Conclusion

This thesis has laid the groundwork for iterative goal-based voting. Through defining novel satisfaction functions, $optH$, $pessH$ and $realH$, we extended the notion of strategic behaviour beyond dichotomous functions to more complex reasoning. These diversified results for the $Approval$ rule: while the strategy-proofness of this rule is limited to agents who base their satisfaction on dichotomous functions (Proposition 2.10), our results demonstrated that iteration is possible for H-optimists, H-pessimists and H-realists, showing the convergence of the iteration to be highly dependent on the satisfaction function being considered.

In particular, we proved that among H-pessimists, iterated $Approval$ voting is guaranteed to terminate (Theorem 3.1). For H-optimists, we needed to restrict agents' actions to what we called *minimal alterations* with a *weak truth-bias* in order to guarantee convergence (Theorem 3.4). Minimal alterations force an agent to alter only in such a way that it is effective, that is, she shall not add any unnecessary changes to her alteration.

Minimal alterations and weak truth-bias restrict alterations in such a way that iterative $Approval$ voting terminates for H-optimists. This restriction is concerned with the effectiveness of an alteration. Sincerity, as used by Endriss (2013) and first introduced by Brams and Fishburn (1978) in the context of voting, might be an alternative approach to this problem. Sincerity means that all models in an agent's goal would have to be weakly preferred to all models this agent does not approve. Note that even though all alterations in the circular profile of $Approval$ voting with H-realists (Table 3.5) are sincere, and thus, this restriction would not be enough to ensure the termination for H-realists, it could be sufficient for the H-optimists. How minimal alterations with weak truth-bias and sincerity correlate is an interesting direction for future work.

In contrast, we did not find constraints for H-realists under which the $Approval$ rule is guaranteed to terminate (Theorem 3.3). One reason why we found profiles for which the iterated $Approval$ rule did not terminate, even when assuming minimal alterations and weak truth-bias, is that an H-realist considers the average distance of all interpretations to her goal. In other words, she does not restrict her interest to the best or worst interpretation. Therefore, she has more opportunities to alter her vote in favour of her truthful goal. The question on how these different satisfaction functions yield different results in iterative voting is also related to the more general question on strategic behaviour under irresolute voting rules considering different preference extensions. A similar question was studied for judgment aggregation by Botan et al. (2016) and Brandt and Brill (2011). One could further analyze what influence an extensions property has on the iteration process.

In Chapter 4 we also showed that iteration yields increased (utilitarian) social welfare for both H-pessimists and H-optimists and no agent's individual satisfaction will suffer from iteration (Theorems 4.1, 4.2, 4.1). The fact that termination with H-realists in iterative $Approval$ voting is not guaranteed should not be considered a negative result, yet. The goal of Chapter 4 was to clear up how frequent iteration with H-realists actually is and its impact on social welfare. One way of analyzing this is through the implementation. While the collected data is not enough to draw definite conclusions, our findings suggest iteration to be rare. Out of $500$ of the randomly chosen profiles with three agents and three issues, only $5.2\%$ iterated, almost all of them terminating after one step. Further analysis is needed to conclude whether the restrictions to such a small number of agents and issues are the reason for the rare iterations, rather than the nature of the $Approval$ rule and the $realH$ function. Although most of the iterated profiles did increase social welfare, we also observed that most initial profiles already had a high social welfare. More simulations are needed in order to draw an accurate conclusion.

In Section 3.2 we showed that none of the Majority rules are guaranteed to terminate under standard conditions (truth-bias or no assumptions, as seen in Theorem 3.1). This result holds for all types of satisfaction functions, including the dichotomous ones. Novaro (2019) has shown that, assuming dichotomous satisfactions functions, the Majority rules are strategy-proof for goals which are incomplete conjunctions. This suggests that there is some middle ground, maybe some milder restriction on the language, for which iterated Majority voting takes place and does terminate. Restricting the language is one of multiple possibilities, like decreasing the set of best responses, as we have done with minimal alterations and the weak truth-bias, or like not assuming agents to be fully informed.

Terzopoulou and Endriss (2018) followed this latter approach while analyzing iterated judgment aggregation. They showed that restricting an agent's information has a positive effect on the termination of a given rule. Analyzing the iteration of goal-based voting further, this would be a feasible approach to consider for the Majority rules as well as the $Approval$ rule with H-realists.

In conclusion iterative $Approval$ voting yields desirable properties: first, iteration is possible, second, it guarantees termination under mild constraints after a reasonable amount of steps and third, it is beneficial for each agent, always returning an optimal outcome if one exists. Considering decision problems such as the one Ann, Betti and Clara faced in Example 1.1, about finding a suitable flat, iterative *Approval* voting would be a good procedure to use. It is fairly simple to let everyone vote for their truthful goal and then let the agents sequentially alter, if they want to. We could imagine that the agent with the largest distance (or a random agent if distances are tied) gets to alter her vote first, even without knowing which function underlies the agents' satisfaction.

**Future Research.** As this thesis lays the ground for iterative goal-based voting it also opens a new area for diverse research questions. First, one could explore more termination results, searching for restrictions under which given or novel rules terminate and in which amount of time they do. There are many variables that one could adapt in order to reach additional results. Some interesting approaches consist of adapting more rules to goal-based voting and trying to reproduce our results or expanding the notion of satisfaction. By testing more restrictions on how agents are allowed to formulate their goals or alter those, we would expect more termination results to arise.

Second, one could follow an axiomatic approach, by analyzing and axiomatizing the rules for which termination can be achieved, as well as the satisfaction functions which yield iteration, as Obraztsova et al. (2015) have done for voting. This may result in impossibility results in the style of the Gibbard-Satterthwaite Theorem (Gibbard, 1973; Satterthwaite, 1975). Additionally, considering group manipulation as it has been done by Botan et al. (2016) for judgment aggregation, would also yield an interesting direction.

Thirdly, in order to further analyze the quality of iteration, we would like to see an application of the *Dynamic Price of Anarchy* as it has been analyzed for iterative voting and judgment aggregation before (Terzopoulou and Endriss, 2018; Koutsoupias and Papadimitriou, 2009; Andelman et al., 2009). Chapter 4 already gives rise to the believe that iteration can be beneficial, however, we have not studied what the possible social costs of iteration under a given rule might be.

Lastly, we would be interested in seeing further implementations. Since our dataset is quite small, it would be sensible to expand the analysis to more agents and issues, as well as to other rules. Taking the research one step further, one could design empirical studies as for example those by Laslier and Sanver (2010) to determine whether and how people actually vote strategically, which satisfaction function they are most likely to follow and if iteration benefits the social welfare.

# Bibliography


Airiau, S. and Endriss, U. (2009). Iterated majority voting. In *Proceedings of the 1st International Conference on Algorithmic Decision Theory (ADT 2009)*.

Andelman, N., Feldman, M., and Mansour, Y. (2009). Strong price of anarchy. *Games and Economic Behavior*, 65(2):289–317.

Arrow, K. J. (1951). *Social choice and individual values*. Monograph / Cowles Commission for Research in Economics ; no. 12. Wiley, New York.

Barrot, N., Lang, J., and Yokoo, M. (2017). Manipulation of Hamming-based approval voting for multiple referenda and committee elections. In *Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017)*.

Baumeister, D., Rothe, J., and Selker, A.-K. (2017). Strategic behavior in judgment aggregation. In Endriss, U., editor, *Trends in Computational Social Choice*, chapter 8. AI Access.

Bossert, W., Pattanaik, P. K., and Xu, Y. (2000). Choice under complete uncertainty: axiomatic characterizations of some decision rules. *Economic Theory*, 16(2):295–312.

Botan, S. and Endriss, U. (2020). Majority-strategyproofness in judgment aggregation. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020)*.

Botan, S., Novaro, A., and Endriss, U. (2016). Group manipulation in judgment aggregation. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)*.

Brams, S. J. and Fishburn, P. C. (1978). Approval voting. *The American Political Science Review*, pages 831–847.

Brandt, F. and Brill, M. (2011). Necessary and sufficient conditions for the strategyproofness of irresolute social choice functions. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2011)*.

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

Brânzei, S., Caragiannis, I., Morgenstern, J., and Procaccia, A. (2013). How bad is selfish voting? In *Proceedings of the 27th Conference on Artificial Intelligence (AAAI 2013)*.

Caragiannis, I. and Procaccia, A. D. (2011). Voting almost maximizes social welfare despite limited communication. *Artificial Intelligence*, 175(9-10):1655–1671.

Delgrande, J. P., Dubois, D., and Lang, J. (2006). Iterated revision as prioritized merging. In *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR 2016)*.

DeMeyer, F. and Plott, C. R. (1971). A welfare function using "relative intensity" of preference. *The Quarterly Journal of Economics*, 85(1):179–186.

Dietrich, F. (2007). A generalised model of judgment aggregation. *Social Choice and Welfare*, 28(4):529–565.

Dietrich, F. and List, C. (2007a). Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics*, 19(4):391–424.

Dietrich, F. and List, C. (2007b). Strategy-proof judgment aggregation. *Economics & Philosophy*, 23(3):269–300.

Endriss, U. (2013). Sincerity and manipulation under approval voting. *Theory and Decision*, 74(3):335–355.

Endriss, U. (2016). Judgment aggregation. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D., editors, *Handbook of Computational Social Choice*, chapter 17. Cambridge University Press.

Endriss, U., Grandi, U., and Porello, D. (2012). Complexity of judgment aggregation. *Journal of Artificial Intelligence Research*, 45:481–514.

Everaere, P., Konieczny, S., and Marquis, P. (2007). The strategy-proofness landscape of merging. *Journal of Artificial Intelligence Research*, 28:49–105.

Everaere, P., Konieczny, S., and Marquis, P. (2015). Belief merging versus judgment aggregation. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015)*.

Everaere, P., Konieczny, S., and Marquis, P. (2017). Belief merging and its links with judgment aggregation. In *Trends in Computational Social Choice*, chapter 7, pages 123–143. AI Access.

Fishburn, P. C. (1972). Even-chance lotteries in social choice theory. *Theory and Decision*, 3(1):18–40.

Gamut, L. T. F. (1991). *Logic, language, and meaning, volume 1: Introduction to logic*, volume 1. University of Chicago Press.

Gibbard, A. (1973). Manipulation of voting schemes: a general result. *Econometrica: Journal of the Econometric Society*, 41(4):587–601.

Grandi, U. and Endriss, U. (2011). Binary aggregation with integrity constraints. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*.

Kaneko, M. and Nakamura, K. (1979). The Nash social welfare function. *Econometrica: Journal of the Econometric Society*, 47(2):423–435.

Kelly, J. S. (1977). Strategy-proofness and social choice functions without singlevaluedness. *Econometrica: Journal of the Econometric Society*, 45(2):439–446.

Konieczny, S. and Pérez, R. P. (2002). Merging information under constraints: a logical framework. *Journal of Logic and Computation*, 12(5):773–808.

Konieczny, S. and Pérez, R. P. (2011). Logic based merging. *Journal of Philosophical Logic*, 40(2):239–270.

Kornhauser, L. A. and Sager, L. G. (1993). The one and the many: Adjudication in collegial courts. *California Law Review*, 81(1):1–59.

Koutsoupias, E. and Papadimitriou, C. (2009). Worst-case equilibria. *Computer science review*, 3(2):65–69.

Laslier, J.-F. and Sanver, M. R. (2010). *Handbook on Approval Voting*. Studies in choice and welfare. Springer, Berlin, Heidelberg.

List, C. and Puppe, C. (2009). Judgement aggregation: A survey. *Oxford Handbook of Rational and Social Choice*.

May, K. O. (1952). A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica: Journal of the Econometric Society*, 20(4):680–684.

Meir, R., Polukarov, M., Rosenschein, J., and Jennings, N. (2010). Convergence to equilibria in plurality voting. In *Proceedings of the 24th Conference on Artificial Intelligence (AAAI 2010)*.

Meir, R., Polukarov, M., Rosenschein, J. S., and Jennings, N. R. (2017). Iterative voting and acyclic games. *Artificial Intelligence*, 252:100–122.

Novaro, A. (2019). *Collective decision-making with goals*. PhD thesis, Université Paul Sabatier-Toulouse III.

Novaro, A., Grandi, U., Longin, D., and Lorini, E. (2018). Goal-based collective decisions: Axiomatics and computational complexity. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018)*.

Obraztsova, S., Markakis, E., Polukarov, M., Rabinovich, Z., and Jennings, N. (2015). On the convergence of iterative voting: how restrictive should restricted dynamics be? In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI 2015)*.

Packard, D. J. (1979). Preference relations. *Journal of Mathematical Psychology*, 19(3):295–306.

Pattanaik, P. K. and Peleg, B. (1984). An axiomatic characterization of the lexicographic maximin extension of an ordering over a set to the power set. *Social Choice and Welfare*, 1(2):113–122.

Pettit, P. (2001). Deliberative democracy and the discursive dilemma. *Philosophical Issues*, 11:268–299.

Rawls, J. (1971). *A Theory of Justice : Original Edition.*, volume Original edition. Harvard University Press.

Reyhani, R. and Wilson, M. (2012). Best-reply dynamics for scoring rules. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*.

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.

Terzopoulou, Z. (2021). *Collective decisions with incomplete individual opinions*. PhD thesis, Universiteit van Amsterdam.

Terzopoulou, Z. and Endriss, U. (2018). Modelling iterative judgment aggregation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)*.