Universiteit van Amsterdam

MSc Artificial Intelligence
Master Thesis

Nash Social Welfare in Judgement Aggregation: Introducing the Kemeny-Nash Rule

by
Pauline Baanders
10818561

July 13, 2023

48 ECTS
Sep 2022 – Jul 2023

Supervisor:
Prof. Dr. Ulle Endriss

Second reader:
Dr. Ronald de Haan

Institute for Logic, Language and Computation

Abstract

Social choice theory studies collective decision making problems. Judgement aggregation, a branch of social choice theory, studies the problem of how individual judgements on a set of (possibly) logically interconnected issues are to be aggregated into a collective judgement. One of the main concerns in judgement aggregation is how to guarantee that the computed collective judgement is logically consistent. Importantly, the intuitively appealing majority procedure—which collectively accepts an issue if and only if it is accepted by at least half of the individuals—does not fulfill this requirement. The Kemeny rule is one of the procedures that has been introduced to solve this problem; the rule is defined in such a way that it always returns a logically consistent collective judgement. Using ideas from welfare economics, and on the basis of results in other social choice theory areas, we argue for a potentially interesting modification of the Kemeny rule: we introduce the Kemeny-Nash judgement aggregation procedure. The Kemeny-Nash rule is potentially interesting because we expect the rule to provide collective judgements that are both efficient and fair. The Kemeny-Nash rule is studied throughout this thesis, in particular in relation to the Kemeny rule. We study axiomatic, computational complexity, and experimental properties.

Acknowledgements

First and foremost I gratefully thank my supervisor, Ulle Endriss. This project was an enormous time investment and did not always go the way we planned it. No matter what, I felt you were on my side, and believed in this project; it is unlikely that I would have been able to finish this thesis without that support. Besides your role as my supervisor, I also want to thank you for your role as teacher. Apart from introducing me to the beautiful field of collective decision making, you encouraged me to be a hard worker, and to pay attention to details. I find it very impressive how you build this computational social choice community, and greatly enjoyed being part of it. Finally, I want to thank you for the opportunity of being a teaching assistant, this was a very special experience, in which I learned a lot.

Further, I would like to thank Ronald de Haan for helping me with my implementation, and for taking the time to read my work.

To my friends and family: you turn my sixes to nines! At the end of the day, you are the ones that give meaning to anything I do.

Contents

Chapter 1

Introduction

Judgement aggregation is a branch of social choice theory that studies how individual judgements on a set of possibly interconnected issues, should be aggregated into a single collective judgement. This problem is central to democratic decision making and arises in various contexts, such as examination boards, expert panels, and multimember courts; but also in informal settings, e.g., when a group of friends decides on their holiday plans.

That this problem is not trivial was first recognised by Pettit (2001), who showed that the interconnections between the issues, in many cases, cause serious difficulties. In particular, it was shown that majority voting, traditionally seen as the triumph of democratic decision-making, fails to ensure consistent collective judgements in many cases. The field of judgement aggregation emerged from this observation, and finding suitable ways to circumvent this problem is still one of the fundamental questions of the field. With this thesis we directly contribute to this line of research; we propose an aggregation method that guarantees consistent collective judgements.

The theory of judgement aggregation provides an important framework for studying how groups can reach consistent collective judgements; in a broader sense our work contributes to any discipline in which such problems arise. In particular, the relevance for AI lies (a.o.) in its applications to multiagent systems and abstract argumentation; where virtual entities are to make a collective decision on the basis of conflicting information, see e.g., Awad et al. (2017).

The remainder of this chapter is structured as follows. In Section 1.1, we motivate our work and discuss its relation to other judgement aggregation research. The structure of this thesis is described in Section

1.1 Motivation and Related Work

In this section we treat the problem that was first described by Pettit (2001) in more detail. We discuss several lines of research that sprung from this work, and posit our own work in this context.

The problem that initiated the field of judgement aggregation was directly motivated by work from the legal philosophers Kornhauser and Sager (1986, 1993), who considered a variant of the following problem.

Suppose that a defendant is found guilty, but urges the verdict should be reversed on two grounds: (i) her confession was inadmissible and (ii) the jury was biased. Legal theory requires that the verdict is reversed if and only if at least one of the claims of the defendant is true. Three judges consider the argument:

	Confession inadmissable?	*Biased jury?*	*Conclusion* (Reverse Yes/No)
Judge A	Yes	No	Yes
Judge B	No	Yes	Yes
Judge C	No	No	No

Will the verdict of the defendant be reversed or not? Well, this depends on our exact question to the judges. If we ask them to judge the premises (i) and (ii) separately, the majority says that the verdict should not be reversed. However, if we ask the juries to judge only the conclusion, and ask every one of them if there is a reason to reverse the verdict, the outcome is opposite; then the verdict will be reversed. This is exactly the point of the legal scholars, it is a doctrinal paradox; two equally reasonable approaches lead to opposite juridical verdicts.

Pettit (2001), and later List and Pettit (2002) revisited the problem that was posed by the legal scholars, and recognised that the scope of this problem reaches far beyond legal theory. They showed that it is impossible to satisfy a small number of salient principles, and to guarantee consistent collective judgements at the same time. In JA examples of this kind are now known as *discursive dilemmas* or *discursive paradoxes* (Endriss, 2016; List and Pettit, 2002).

After the work of List and Pettit various approaches have been taken to circumvent the impossibility. One possibility is to focus on the role of the judgement aggregation procedure; several rules have been put forward that guarantee consistent collective judgements. The Kemeny rule is one of the examples, this procedure collectively accepts the consistent judgements that are, in some well-defined way, closest to the list of individual judgements. This thesis falls within this line of research: we study a modification to the Kemeny rule, we introduce the Kemeny-Nash rule, which guarantees a consistent collective judgement.

1.2 Thesis Overview

This thesis is structured as follows:

Chapter 2. In this chapter we present the judgement aggregation framework proposed by Endriss et al. (2020), which is used throughout the rest of this thesis. The distinctive feature of this framework is that the demands on individual judgements are distinguished from the demands on a collective judgement. We discuss different kinds of aggregation procedures that have been studied in the literature. One of these procedures, that is of particular relevance for this work, is the Kemeny rule. We use concepts from the field of welfare economics to propose a modification to the Kemeny rule that should make the rule sensitive to fairness considerations. We introduce two variants of the Kemeny-Nash rule.

Chapter 3. In all parts of social choice theory, the axiomatic approach is one of the fundamental tools that is used to scrutinise collective decision problem procedures. In Chapter 3 we study the axiomatic properties of the Kemeny-Nash rule, in particular, in comparison to the axiomatic properties of the Kemeny rule.

Chapter 4. Besides the normative principles that a procedure complies with, it is also important what computational resources are necessary to use the procedure in practice. Employing tools from computational complexity theory, in Chapter 4, we study the *outcome determination problem in judgement aggregation*—the problem of computing a collective judgement. For the Kemeny rule, this problem is known to be Θ_2^p-complete; we reiterate the result from Endriss et al. (2020). We show that for both variants of the Kemeny-Nash rule, this problem is contained in Δ_2^p and hard for Θ_2^p. Whether the problem of one (or both) of the variants of the Kemeny-Nash rule is complete for either one of these classes is unanswered.

Chapter 5. At the end of the day, collective decision procedures are meant to be used in the tangible world, in which theoretic results are hardly ever the hole story. In Chapter 5, we study the differences between collective judgements that have been computed with the Kemeny-Nash rule and the Kemeny rule in an experimental way. Our implementation is based on the `Jaggpy` library.[1]. Although we consider other aspects as well, the major part of this chapter is devoted to a qualitative analysis. We introduce criteria to evaluate both the efficiency and the fairness of the collective judgements, as computed by the different rules.

We conclude in Chapter 6, where we summarise our results and indicate possible directions for future research.

[1] `https://pypi.org/project/jaggpy/`

Chapter 2

Background

This chapter is meant to serve as a foundational resource for the mathematical language used throughout this thesis. In Section 2.1 we discuss the judgement aggregation framework introduced by Endriss et al. (2020). In particular, we explain the formal description of scenarios in which judgement aggregation can be applied. Moving on, Section 2.2 outlines the formalism for judgement aggregation procedures which are used to compute collective judgements. In Section 2.3 we make a detour into welfare economics, examining judgement aggregation procedures in this context. Our study suggests a novel judgement aggregation procedure, which we introduce as the Kemeny-Nash rule. Our final section, provides an initial exploration of the Kemeny-Nash rule (Section 2.4).

2.1 Framework

The first formal framework for judgement aggregation was established by List and Pettit (2002) and is now known as *formula-based judgement aggregation*. In their framework issues are represented as formulas in classical propositional logic. Although several related frameworks exist in the judgement aggregation literature today, the original framework is still widely used (Endriss et al., 2020).

In this thesis, we use the framework of Endriss et al. (2020), which is more general than any other known framework in the literature. This framework distinguishes an input constraint Γ_{in} for individual judgements from an output constraint Γ_{out} for collective judgements.

While the full generality of the framework is only used in Chapter 3, we believe that the explicitness of the framework clarifies matters. We deviate from the original exposition by Endriss et al. (2020) only in the representation of a collection of individual judgements (we use an unordered representation).

In simple terms, a *judgement aggregation scenario* is the mathematical description of a situation that can be evaluated using judgement aggregation. This section provides an in-depth explanation of the mathematical descriptions for such scenarios, using the framework by Endriss et al. (2020). We further discuss the influence of the representation of a collection (or profile) of individual judgements, ordered or unordered, on the total number of different collections. Finally, we revisit the discursive dilemma from Chapter 1, and render it using our formal framework.

Symbols. To get an idea of where we are heading, let us just look at the symbols that we need to fully specify a judgement aggregation scenario. Formally, a scenario is a tuple $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ containing a set of issues Φ, a constraint Γ_{in} for individual judgements, a constraint Γ_{out} for collective judgements, and a collection $\boldsymbol{J}$ of individual judgements. The issues and constraints contain formulas that are expressed in classical propositional logic.

Agendas. Let language $\mathcal{L}$ consist of all formulas constructed from a finite set of propositional variables and the usual logical connectives ($\neg, \vee, \wedge, \rightarrow, \leftrightarrow, \top$ and $\bot$). Double negations are eliminated by defining the complement of a proposition $\sim\varphi$ as follows: $\sim\varphi := \vartheta$ if $\varphi = \neg\vartheta$ (for some ϑ) and $\sim\varphi := \neg\varphi$ otherwise. The set of issues $\Phi \subseteq \mathcal{L}$, known as the *agenda*, is a subset of language $\mathcal{L}$ and is closed under complementation (i.e., $\varphi \in \Phi$ if and only if $\sim\varphi \in \Phi$). The *pre-agenda* $\Phi^+ \subset \Phi$ contains all non-negated formulas that appear in the corresponding agenda Φ. Thus, given an agenda $\Phi = \{\varphi_1, \ldots, \varphi_m, \neg\varphi_1, \ldots, \neg\varphi_m\}$ built from non-negated formulas φ_i, the corresponding pre-agenda is defined as $\Phi^+ = \{\varphi_1, \ldots, \varphi_m\}$. The issues in the agenda are evaluated by a group, formally defined as a set $N = \{1, \ldots, n\}$, of n judges.

Judgement sets. Individual and collective judgements are represented as *judgement sets* $J \subseteq \Phi$ consisting of the accepted agenda-issues. The antipodal of judgement $J \subseteq \Phi$ is denoted as $\overline{J}$, and contains the agenda-issues that are not contained in J, i.e., $\overline{J} = \Phi \setminus J$. A judgement J is *consistent* if J is a satisfiable set of formulas; it is *complete* if for all agenda-issues $\varphi \in \Phi$, it holds that $\varphi \in J$ or $\sim\varphi \in J$; finally, J is *complement-free* if there is no formula $\varphi \in \Phi$ such that both $\varphi \in J$ and $\sim\varphi \in J$.

Further, judgement set J is *rational*, or Γ_{in}-consistent, if it is compatible with the *input constraint* Γ_{in}; that is, $J \cup \{\Gamma_{\text{in}}\}$ is satisfiable. Similarly, a judgement set J is *feasible*, or Γ_{out}−consistent, if J is compatible with the *output constraint* Γ_{out}. Note that both rational and feasible judgements are consistent by definition. The set of all complete and complement-free rational judgements is denoted as $\mathcal{J}(\Phi, \Gamma_{\text{in}})$, similarly, set $\mathcal{J}(\Phi, \Gamma_{\text{out}})$ contains all complete and complement-free feasible judgements.

We introduce some auxiliary notation to describe (properties of) judgement sets. When agenda Φ and constraint Γ (either an input or output constraint, $\Gamma \in \{\Gamma_{\text{in}}, \Gamma_{\text{out}}\}$) are clear from the context, for an arbitrary subset $S \subseteq \Phi$ of agenda-issues we define the set of all complete, complement-free and Γ-consistent *extensions* of S as follows:

$$ext(S) = \{J \in \mathcal{J}(\Phi, \Gamma) \mid J \supseteq S\}$$

We present two measures for the (dis)similarity between complete and complement-free judgements. The *Hamming distance* is widely used in the judgement aggregation literature and is a measure for the dissimilarity between judgement sets; the *agreement* between two judgements is a notion that is closely related, but that measures the similarity instead. Formally, the *Hamming distance* between two complete and complement-free judgements J, J' is defined as the number of issues in Φ^+ on which they do not agree:

$$H(J, J') = |J \setminus J'| = |J' \setminus J|$$

Correspondingly, we define the *agreement* between two complete and complement-free judgements J, J' as the number of issues on which they agree:

$$\mathrm{Agr}(J, J') = |J \cap J'|$$

Thus, for two complete and consistent-free judgements J, J' the Hamming distance and agreement are related as follows: $\mathrm{Agr}(J, J') = |\Phi^+| - H(J, J')$.

Example 2.1. Consider the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2, \varphi_3\}$ containing three propositional variables. Let $J = \{\varphi_1, \varphi_2, \varphi_3\}$ and $J' = \{\varphi_1, \neg\varphi_2, \neg\varphi_3\}$ be two judgements that are complete and complement-free. In this case we have $|\Phi^+| = 3$, $H(J, J') = 2$ and $\mathrm{Agr}(J, J') = 1$. $\triangle$

Profiles. Finally, a *profile* (for n judges) is a collection of n individual judgements. In the original exposition by Endriss et al. (2020) such a profile $\boldsymbol{J} = (J_1, \dots, J_n)$ is represented as a vector, specifying a rational judgement $J_i \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$ for all judges $i \in N$. For such a profile we may write $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$.

Because we never explicitly consider the identity of the judges, in our setting, the vector representation is superfluous, we use *multisets* instead. Informally, a multiset is a set (i.e., unordered representation) in which a single element may be contained more than once.

We represent a profile for n judges as a multiset $\boldsymbol{J}$ of cardinality n, with underlying set $\mathcal{J}(\Phi, \Gamma_{\text{in}})$. It is a function $\boldsymbol{J} : \mathcal{J}(\Phi, \Gamma_{\text{in}}) \to \mathbb{N}$ that for every rational judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$, specifies the number of times it occurs—its *multiplicity* $\boldsymbol{J}(J)$. The cardinality of such a multiset is defined as $|\boldsymbol{J}| = \sum_{J \in \mathcal{J}(\Phi, \Gamma_{\text{in}})} \boldsymbol{J}(J)$; it is the sum of multiplicities of all contained elements. Although, in our representation, this is not completely correct, we use $\mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ to denote the set of all profiles with n judges. We write $\mathcal{J}(\Phi, \Gamma_{\text{in}})^\dagger$ to denote the set of all profiles, for any finite number of judges. By a slight abuse of notation we write $J \in \boldsymbol{J}$, if $J \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$ and $\boldsymbol{J}(J) > 0$. Further, for any (multi)set $S = \{s\}$, with cardinality $|S| = 1$, we may simply write s to refer to set $\{s\}$.

We now introduce auxiliary expressions related to profiles. We define $\boldsymbol{J} \uplus \boldsymbol{J'}$ to be the sum of two profiles, i.e., $(\boldsymbol{J} \uplus \boldsymbol{J'})(J) = \boldsymbol{J}(J) + \boldsymbol{J'}(J)$. For every $\ell \in \mathbb{N}$, we further define $\ell\boldsymbol{J} = \boldsymbol{J} \uplus \cdots \uplus \boldsymbol{J}$ as the sum of ℓ copies of profile $\boldsymbol{J}$. A *compound* profile is a profile that is constructed from several *constituent* profiles. As an example, if $\boldsymbol{J} = \boldsymbol{J}_1 \uplus \boldsymbol{J}_2$, then $\boldsymbol{J}$ is a compound profile; the profiles $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$ are its constituents.

Let $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ be any profile for n judges. The *support judgements*, hereafter simply referred to as the support, of a formula $\varphi \in \Phi$ is a function $N_\varphi^{\boldsymbol{J}} : \mathcal{J}(\Phi, \Gamma_{\text{in}}) \to \mathbb{N}$, defined as:

$$N_\varphi^{\boldsymbol{J}}(J) = \begin{cases} \boldsymbol{J}(J) & \text{if } \varphi \in J \\ 0 & \text{otherwise} \end{cases}$$

Note that function $N_\varphi^{\boldsymbol{J}}$ defines a multiset (containing the judgements that accept formula φ); and $|N_\varphi^{\boldsymbol{J}}|$ equals the number of judges that support formula φ in profile $\boldsymbol{J}$. Its complement with respect to $\boldsymbol{J}$, $\overline{N_\varphi^{\boldsymbol{J}}} = \boldsymbol{J} \setminus N_\varphi^{\boldsymbol{J}}$, denotes the multiset of judgements that reject φ. Further, we define the *majoritarian judgement* $m(\boldsymbol{J})$:

$$m(\boldsymbol{J}) = \left\{ \varphi \in \Phi \mid |N_\varphi^{\boldsymbol{J}}| > \frac{n}{2} \right\}$$

If judgement $m(\boldsymbol{J})$ is feasible, or Γ_{out}-consistent, we say that profile $\boldsymbol{J}$ is *majority-consistent.*

Comparison of profile representations. To conclude the formal description of scenarios, we compare the two (vector and multiset) representations in terms of the total number of profiles. Clearly, for both representations, the total number of profiles depends on the number of judges n, and the number of rational judgements $|\mathcal{J}(\Phi, \Gamma_{\text{in}})|$. Let $g, g' : (n, \mathcal{J}(\Phi, \Gamma_{\text{in}})) \to \mathbb{N}$ be the functions that relate a scenario to the total number of vectors and the total number of multisets (respectively). Of course, for the vector representation, we have: $g(n, \mathcal{J}(\Phi, \Gamma_{\text{in}})) = |\mathcal{J}(\Phi, \Gamma_{\text{in}})|^n$.

For the multiset representation this number is given by the so called *multiset coefficient.* Let S, with $|S| = s$, be an arbitrary set. The *multiset coefficient* for the number of multisets with cardinality k, with underlying set S is defined as:

$$\left(\!\!\binom{s}{k}\!\!\right) = \binom{s+k-1}{k} = \frac{(s+k-1)!}{k!(s-1)!}$$

Thus, for the multiset representation, the function g' is defined as:

$$g'(n, \mathcal{J}(\Phi, \Gamma_{\text{in}})) = \frac{(|\mathcal{J}(\Phi, \Gamma_{\text{in}})| + n - 1)!}{n!(|\mathcal{J}(\Phi, \Gamma_{\text{in}})| - 1)!}$$

Example 2.2. We consider three numbers of feasible judgements: $|\mathcal{J}(\Phi, \Gamma_{\text{in}})| = 4$, $|\mathcal{J}(\Phi, \Gamma_{\text{in}})| = 13$, and $|\mathcal{J}(\Phi, \Gamma_{\text{in}})| = 24$. For three different numbers of judges ($n = 5$, $n = 15$, and $n = 50$) we present the total number of profiles, for both the multiset and the vector representation, in Table 2.1. $\triangle$

Discursive dilemma revisited. Finally, to recapitulate the formal framework we described above, let us reconsider the discursive dilemma (see Chapter 1).

	# multisets	# tuples
4 rational judgements		
5 judges	56	$1.02 \cdot 10^3$
15 judges	$8.16 \cdot 10^2$	$1.07 \cdot 10^9$
50 judges	$2.34 \cdot 10^4$	$1.27 \cdot 10^{30}$
13 rational judgements		
5 judges	$6.19 \cdot 10^3$	$3.71 \cdot 10^5$
15 judges	$1.74 \cdot 10^7$	$5.12 \cdot 10^{16}$
50 judges	$2.16 \cdot 10^{12}$	$4.98 \cdot 10^{55}$
24 rational judgements		
5 judges	$9.83 \cdot 10^4$	$7.90 \cdot 10^6$
15 judges	$1.55 \cdot 10^{10}$	$5.05 \cdot 10^{20}$
50 judges	$5.69 \cdot 10^{18}$	$1.02 \cdot 10^{69}$

Table 2.1: The number of different profiles for the multiset (employed here)and thevector representation (employed by Endriss et al. (2020)). We show thenumber of different profiles, for different numbers of judges ($n \in \{5, 15, 50\}$) anddifferent numbers of feasible judgements ($|\mathcal{J}(\Phi, \Gamma_{\text{in}})| \in \{4, 13, 24\}$).

Example 2.3. In the situation we described in Chapter 1, three judges gave their opinion on the following issues: (i) there was an inadmissible confession, (ii) the (original) jury was biased, and (iii) statement (i) or statement (ii) is true (in this case there is a legal ground to reverse the verdict). To describe the situation as a formal judgement aggregation scenario, we could define $\Phi^+ = \{\varphi_1, \varphi_2, \varphi_1 \vee \varphi_2\}$, where formula φ_1 stands for 'statement (i) is true' (similarly, φ_2 indicates whether statement (ii) is true). The input and output constraint coincide and are trivially satisfied: $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \top$. To represent the opinions of the judges we can no longer refer to the individual judges, we could this in the following way:

	φ_1	φ_2	$\varphi_1 \vee \varphi_2$
#1	Yes	No	Yes
#1	No	Yes	Yes
#1	No	No	No
$m(\boldsymbol{J})$	No	No	Yes

Now, the 'dilemma' we saw was that if we let the judges decide on the individual premises (φ_1 and φ_2), then there is no legal ground to reverse the verdict (according to the majority of the judges). However, if we decide to let the judges evaluate only the conclusion ($\varphi_1 \vee \varphi_2$), then there is reason to reverse the verdict. Put differently, the majoritarian judgement $m(\boldsymbol{J})$ is inconsistent; and if we want to decide what to do in this case we cannot simply use majority aggregation. △

2.2 Rules

A *judgement aggregation procedure* or *rule* is a function, mapping an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{\dagger}$ (for any number of judges) to a set of corresponding collective judgements. In contrast to voting theory, in judgement aggregation there is little research that focuses on the properties of a particular class of aggregation procedures (Lang et al., 2011). In this section we treat three exceptions: quota-based rules, the premise-based rule, and distance-based rules. The majority rule belongs to the class of quota-based rules. The other two classes have been proposed as solutions to the problem posed by the discursive dilemma, which we discussed in the Introduction (Chapter 1). We start with the formal definition of a judgement aggregation procedure.

Procedures in judgement aggregation. We distinguish two types of aggregation procedures. *Resolute rules* always return a single collective judgement, while *irresolute rules* may return multiple collective judgements. To distinguish the two types of procedures, we write f for resolute, and F for irresolute procedures. A resolute rule is a function f mapping an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{\dagger}$, for any number of judges, of individual judgements to a collective judgement that is complete and complement-free. On the other hand, an irresolute rule F maps any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{\dagger}$ to a non-empty set of (possibly multiple) complete and complement-free judgements.

In the judgement aggregation literature it is customary to employ resolute rules (Endriss, 2016). However, as the 'measure of irresoluteness' of a rule, i.e., the average number of collective judgements it returns, is an interesting property by itself (see Chapter 5), we use irresolute rules instead.[1]

Quota-based rules. Quota-based rules associate every issue in the agenda with a quota; whenever the number of judges that accept the issue is (weakly) higher than the quota, the issue is collectively accepted. Quota-based rules are studied by (a.o.) Dietrich and List (2007) and Dietrich (2010). Formally, given a function $q : \Phi \to \{0, 1, \ldots, n+1\}$ which maps every issue to its associated quota, the corresponding rule is defined as:

$$F_q(\boldsymbol{J}) = \{\varphi \in \Phi \mid |N_{\varphi}^{\boldsymbol{J}}| \geq q(\varphi)\}.$$

Note that when $q(\varphi) = 0$, then issue $\varphi \in \Phi$ is always accepted; if $q(\varphi) = n + 1$, then issue $\varphi \in \Phi$ is never accepted.

The *majority rule* is the most familiar example of this class. Formally, the (strict) majority rule is instigated by the function $\varphi \mapsto \lfloor \frac{n}{2} + 1 \rfloor$. In general, whenever function q is constant, F_q is said to be a *uniform quota rule*.

[1] Only in Section 3.1, where we present the theorem of List and Pettit (2002)—which we already encountered in the Introduction (Chapter 1)—we consider resolute rules f.

Premise-based rule. The premise-based rule is studied by (a.o.) Kornhauser and Sager (1993), Dietrich and Mongin (2010), and Mongin (2008). With the *premise-based rule* it is assumed that the agenda Φ can be partitioned into premises Φ^p—which are required to be logically independent—and conclusions Φ^c, i.e., $\Phi = \Phi^p \cup \Phi^c$ and $\Phi^p \cap \Phi^c = \emptyset$.

The judges give their opinion only on the premises, and a premise $\varphi \in \Phi^p$ is accepted if and only if it is accepted by a majority of the judges. A conclusion $\varphi \in \Phi^c$ is accepted if and only if it is entailed by the set of accepted premises. Formally, the premise-based rule is defined as follows:

$$F_p(\boldsymbol{J}) = \Phi^p_{\text{maj}} \cup \{\varphi \in \Phi^c \mid \Phi^p_{\text{maj}} \models \varphi\}, \quad \text{with } \Phi^p_{\text{maj}} = \left\{\varphi \in \Phi^p \mid |N^{\boldsymbol{J}}_\varphi| > \frac{n}{2}\right\}$$

Because the set of premises is logically independent—i.e., there is no subset of premises that is inconsistent—and conclusions are accepted only if they are consistent with the accepted premises, the discursive dilemma is circumvented.

Example 2.4. In the example we considered in the Introduction (Chapter 1, formalised in Example 2.3) it would be natural to designate issues φ_1 and φ_2 as premises, and issue $\varphi_1 \vee \varphi_2$ as conclusion; i.e., $\Phi^p = \{\varphi_1, \varphi_2\}$ and $\Phi^c = \{\varphi_1 \vee \varphi_2\}$. The premise-based rule would return the judgement $J = \{\neg\varphi_1, \neg\varphi_2, \neg(\varphi_1 \vee \varphi_2)\}$, and there would be no reason to revise the verdict. △

Distance-based rules. Distance-based rules have been studied by (a.o.) Lang et al. (2011), Miller and Osherson (2009) and Pigozzi (2006). A distance-based rule uses a distance metric $d(J_{\text{in}}, J)$ to determine the (dis)similarity between any rational judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$ and feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$.[2]

The distance from an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ to any feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ is then defined as an aggregate of the individual distances between the judgements $J_{\text{in}} \in \boldsymbol{J}$ in the profile and the feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$. Formally, given a distance metric $d : \mathcal{J}(\Phi, \Gamma_{\text{in}}) \cup \mathcal{J}(\Phi, \Gamma_{\text{out}}) \times \mathcal{J}(\Phi, \Gamma_{\text{in}}) \cup \mathcal{J}(\Phi, \Gamma_{\text{out}}) \to \mathbb{R}$ and some operation $\odot$, a distance based rule is defined as:

$$F_d(\boldsymbol{J}) = \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\text{argmin}} \bigodot_{J_{\text{in}} \in \boldsymbol{J}} \left(\odot_{i=1}^{\boldsymbol{J}(J_{\text{in}})} d(J_{\text{in}}, J)\right)$$

The double appearance of the operation $\odot$ might be confusing. The expression says that for every judgement $J_{\text{in}} \in \boldsymbol{J}$ we calculate a partial contribution which is based on the multiplicity $\boldsymbol{J}(J_{\text{in}})$; then, we use the same operation to aggregate the partial contributions of every judgement $J_{\text{in}} \in \boldsymbol{J}$. By construction, a distance-based rule always returns a set of feasible judgements.

The Kemeny rule is a distance-based rule that frequently appears in the literature; see, e.g., Nehring and Pivato (2022), De Haan and Slavkovik (2017), and Lang et al.

[2]A distance metric on a set S, with arbitrary $x, y, z \in S$, is a function $d : S \times S \to \mathbb{R}$ that must satisfy the following conditions: (i) $d(x, x) = 0$, (ii) $d(x, y) > 0$ whenever $x \neq y$, (iii) $d(x, y) = d(y, x)$, and (iv) $d(x, y) + d(y, z) \geq d(x, z)$.

(2011). In the literature the Kemeny rule is defined as the rule that minimises the sum of Hamming distances; for an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{\dagger}$ the Kemeny rule is conventionally defined as $F'_{\text{kem}}(\boldsymbol{J}) = \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmin}} \sum_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot H(J_{\text{in}}, J))$. For reasons we discuss in Section 2.4 we define the Kemeny rule in a slightly different, but equivalent, manner. Instead of minimising the sum of Hamming distances, we define the Kemeny rule as the procedure that maximises the sum of agreements (see Section 2.1).

Definition 2.1. For an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ for n judges, the Kemeny rule is (unconventionally) defined as follows:

$$F_{\text{kem}}(\boldsymbol{J}) = \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} \sum_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \operatorname{Agr}(J_{\text{in}}, J))$$

We use the Kemeny rule F_{kem} throughout this thesis. To enhance readability, for any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{\dagger}$ and feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ we define the Kemeny score as follows:

$$S_{\text{kem}}(\boldsymbol{J}, J) = \sum_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \operatorname{Agr}(J_{\text{in}}, J))$$

Example 2.5. For the example we discussed in the Introduction (Chapter 1, formalised in Example 2.3), the Kemeny-Nash rule F_{kem} would return the following three collective judgements: $J = \{\varphi_1, \neg\varphi_2, \varphi_1 \vee \varphi_2\}$, $J' = \{\neg\varphi_1, \varphi_2, \varphi_1 \vee \varphi_2\}$, and $J'' = \{\neg\varphi_1, \neg\varphi_2, \neg(\varphi_1 \vee \varphi_2)\}$. △

2.3 Welfare Economics and the Kemeny Rule

In this section we want to link aggregation procedures to Bergson-Samuelson social welfare functions, hereafter referred to as cardinal social welfare functions.[3] The notion of a social welfare function originates in the field of welfare economics; it is a way to rank economically feasible allocations. For any feasible allocation, a cardinal social welfare function computes the collective welfare on the basis of the individual utilities that are induced by the allocation.[4]

Of course, just as there are different judgement aggregation procedures, there are different ways to compute the collective welfare of a society on the basis of the individual utilities. Three important approaches are: utilitarian, egalitarian and Nash social welfare. As we explain below, the utilitarian and egalitarian approach focus on economic *efficiency* and economic *equity* (or fairness), respectively. Nash social welfare combines efficiency and fairness considerations.

[3]In social choice theory it is more common (and it would have been clearer) to use the term collective utility function for this purpose; the term social welfare function is typically reserved for Arrow social welfare functions.

[4]The utility of an individual agent is a measure for their satisfaction.

The purpose of this section is to show that, under mild assumptions, the Kemeny rule can be interpreted as a utilitarian approach. Moreover, under the same mild assumptions, a small modification to the rule makes it adhere to Nash social welfare; we obtain the *Kemeny-Nash* judgement aggregation procedure.

The remainder of this section is structured as follows. We start by covering the relevant background from welfare economics (Section 2.3.1). Among other things, we give formal definitions of the three welfare functions mentioned above (utilitarian, egalitarian and Nash social welfare). In Section 2.3.2 we use a widely accepted assumption for the preferences of individual judges (Hamming preferences) that enables us to analyse judgement aggregation procedures in the context of welfare economics. This analysis suggests a novel judgement aggregation procedure: the *Kemeny-Nash rule.*

2.3.1 Introduction to Welfare Economics

In welfare economics the goal is to rank economically feasible allocations according to the societal welfare they induce. It is assumed that the societal welfare is *completely* determined by the satisfaction levels of the individuals in the society. Other factors, e.g., the impact on the environment are not taken into account—unless the satisfaction of an individual agent depends on it.

There are two ways to describe the satisfaction level of an individual agent: via a *preference order*, or via a cardinal *utility function.* Let A be the set of feasible allocations and take arbitrary allocations $a, a', a'' \in A$. A preference order is a binary relation $\succeq : A \times A \to \{0, 1\}$ over the set of feasible allocations. This relation is required to satisfy *completeness* (i.e., either $a \succeq a'$ or $a' \succeq a$) and *transitivity* (i.e., if $a \succeq a'$ and $a' \succeq a''$ then $a \succeq a''$). The other method, a cardinal utility function $u : A \to \mathbb{R}$ maps every feasible allocation to a real value, where a higher value indicates a higher level of satisfaction. More precisely, the utility of allocation a is greater than (or equal to) the utility of allocation a' if and only if a is weakly preferred over allocation a'; i.e., $u(a) \geq u(a')$ if and only if $a \succeq a'$. Thus, a cardinal utility function u corresponds to a unique preference order $\succeq$, the converse is not true.

When we consider the satisfaction of the society as a whole, we can distinguish, in a similar way as before, a *cardinal social welfare function* (SWF) from a *social welfare ordering* (SWO). Before we describe these notions properly, we introduce some new notations for collections of utility functions.

Let $N = \{1, \ldots, n\}$ be a set of individual agents, each associated to a utility function $u_i : A \to \mathbb{R}$, where A denotes the set of feasible allocations. Given an allocation $a \in A$, and utility functions u_i for all agents $i \in N$, we define the *utility vector* $\boldsymbol{u}_a = (u_1(a), \ldots, u_n(a)) \in \mathbb{R}^n$. Hereafter, we may simply talk about a utility vector, without explicit reference to the allocation a that induced it; we may write $\boldsymbol{u}$ instead of $\boldsymbol{u}_a$. Further, given an arbitrary utility vector $\boldsymbol{u} \in \mathbb{R}^n$, we let vector $\boldsymbol{u}^{\text{lex}}$

denote the vector in which the components are lexicographically ordered. That is, if index i is smaller than index i'; then $\boldsymbol{u}_i^{\text{lex}} \leq \boldsymbol{u}_{i'}^{\text{lex}}$. Finally we define the arithmetic mean of vector $\boldsymbol{u} \in \mathbb{N}$ as: $\langle \boldsymbol{u} \rangle = \frac{1}{n} \sum_{i \in N} \boldsymbol{u}_i$.

A *cardinal social welfare function* $U : \mathbb{R}^n \to \mathbb{R}$ maps an arbitrary utility vector to a collective utility. On the other hand, a *social welfare ordering* is a binary relation $\succeq^{\text{col}} : \mathbb{R}^n \times \mathbb{R}^n \to \{0, 1\}$ over the set of possible utility vectors. For arbitrary vectors $\boldsymbol{u}, \boldsymbol{u'} \in \mathbb{R}^n$ we have, $U(\boldsymbol{u}) \geq U(\boldsymbol{u'})$ if and only if $\boldsymbol{u} \succeq^{\text{col}} \boldsymbol{u'}$.

We now consider three important social welfare functions: utilitarian, egalitarian and Nash social welfare. For each SWF, we give its formal definition, literature references and a brief explanation of the intuition behind the particular welfare function. To explain the ideas behind the different approaches it can be helpful to think of utility as something material; in particular the concept of money may be a helpful analogy.

Utilitarian social welfare is studied by Sen (1974). The formal definition of the SWF is given below.

Definition 2.2. Given a utility vector $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ for n agents, the utilitarian social welfare function is defined as:

$$U_{\text{util}}(\boldsymbol{u}) = \sum_{i \in N} u_i$$

The utilitarian SWF ranks feasible allocations on the basis of the average utility they induce. If we think of the utility of an agent in terms of money: the utilitarian approach says that the welfare of a society is completely determined by the total amount of money that is distributed among its members. The idea is that we should distribute as much money as possible, in that sense the utilitarian approach maximises *economic efficiency.* Considering the distribution of the money among the individuals: according to utilitarianism this factor does not effect the societal welfare. The function is blind to *economic equity* (or fairness) considerations.

Example 2.6. Let $\boldsymbol{u} = (5, 1, 1)$ and $\boldsymbol{u'} = (3, 3, 1)$ and $\boldsymbol{u''} = (2, 2, 2)$. For the utilitarian approach we have: $U_{\text{util}}(\boldsymbol{u}) = U_{\text{util}}(\boldsymbol{u'}) > U_{\text{util}}(\boldsymbol{u''})$. △

Egalitarion social welfare is studied by Hammond (1976), Sen (1976) and Sen (1974). Below we provide the formal definition of the egalitarian SWF.

Definition 2.3. Given a utility vector $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ for n agents, the egalitarian social welfare function is defined as:

$$U_{\text{egal}}(\boldsymbol{u}) = \min_{i \in N} u_i$$

In an egalitarian society the societal welfare is equal to the welfare of its poorest member. Egalitarian social welfare is solely concerned with economic equity, it does not consider economic efficiency.

Example 2.7. Let $\boldsymbol{u} = (5, 1, 1)$ and $\boldsymbol{u}' = (3, 3, 1)$ and $\boldsymbol{u}'' = (2, 2, 2)$. For the egalitarian approach we have: $U_{\text{egal}}(\boldsymbol{u}'') > U_{\text{egal}}(\boldsymbol{u}') = U_{\text{egal}}(\boldsymbol{u})$. △

Nash social welfare (NSW) is studied in Nash (1950), Caragiannis et al. (2019) and Varian (1974); the formal definition of the SWF is given below.

Definition 2.4. Given a utility vector $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ for n agents, the Nash social welfare function is defined as:

$$U_{\text{nash}}(\boldsymbol{u}) = \prod_{i \in N} u_i$$

By maximising the product of utilities, the Nash social welfare function combines economic efficiency and economic equity. Looking to the same example as before, we see that the Nash social welfare stimulates inequality-reducing trades that do not effect the average utility.

Example 2.8. Let $\boldsymbol{u} = (5, 1, 1)$ and $\boldsymbol{u}' = (3, 3, 1)$ and $\boldsymbol{u}'' = (2, 2, 2)$. For the Nash SWF we have: $U_{\text{nash}}(\boldsymbol{u}'') > U_{\text{nash}}(\boldsymbol{u}') = U_{\text{nash}}(\boldsymbol{u})$. △

To show that this is not just the case for this particular example we prove the following proposition.

Proposition 2.1. *Let $X = (x_1, \ldots, x_n)$, with $x_i \in \mathbb{N}$, and define $N = \{1, \ldots, n\}$. We constrain the values x_i for $i \in N$ in the following way: $\sum_{i \in N} x_i = C$. We further assume that $C \mod n = 0$, then:*

$$\underset{(x_1,\ldots,x_n)\in\mathbb{N}^n}{\operatorname{argmax}} \prod_{i \in N} x_i = (x_1^*, \ldots, x_n^*), \qquad \textit{where } x_i^* = \frac{C}{n} \textit{ for all } i \in N \tag{2.1}$$

Proof. We use Jensen's inequality (Jensen, 1906). Jensen's inequality states that for any real concave function $f : \mathbb{R} \to \mathbb{R}$, with numbers $x_1, \ldots, x_n \in \mathbb{R}$ in the domain of f the following inequality holds:

$$\frac{\sum_{i=1}^n f(x_i)}{n} \leq f\left(\frac{\sum_{i=1}^n x_i}{n}\right)$$

Moreover, equality holds if and only if $x_i = x_j$ for all $i, j \in N$. By substituting the concave logarithmic function $\log(x)$ for $f(x)$ we obtain:

$$\frac{\sum_{i=1}^n \log(x_i)}{n} \leq \log\left(\frac{\sum_{i=1}^n x_i}{n}\right) \tag{2.2}$$

To apply Jensen's inequality we rewrite the original argument in Equation 2.1 as follows:

$$\begin{aligned}
\underset{(x_1,\ldots,x_n)\in\mathbb{N}^n}{\operatorname{argmax}} \prod_{i \in N} x_i &= \underset{(x_1,\ldots,x_n)\in\mathbb{N}^n}{\operatorname{argmax}} \frac{\log\left(\prod_{i \in N} x_i\right)}{n} \\
&= \underset{(x_1,\ldots,x_n)\in\mathbb{N}^n}{\operatorname{argmax}} \frac{\sum_{i \in N} \log(x_i)}{n} \qquad (2.3)
\end{aligned}$$

To obtain the first equality, we first take the logarithm of the argument and divide by constant $n \in \mathbb{N}$. The second equality is trivial. The conclusion, Equation 2.1, now directly follows from Jensen's inequality. $\square$

The proposition shows that under the constraint that $\sum_{i \in N} u_i = C$—if conditions (i) $C \mod n = 0$ and (ii) there exists a feasible allocation $a \in A$ such that $u_i(a) = u_j(a)$ for all $i, j \in N$ are met—maximising Nash social welfare amounts to dividing the total utility equally among all agents.

Without providing a formal proof, if the conditions (i) and (ii) are not met, we can still see from Equation 2.3—as the logarithmic function is a concave function (i.e., $\log(x + \mathrm{d}x) - \log(x) < \log(x) - \log(x - \mathrm{d}x)$ for all $x \in \mathbb{R}_{\geq 0}$)—that for a fixed average utility the Nash social welfare is maximised by putting the individual utilities as close to the average utility as possible. In particular, under the constraint that $\sum_{i \in N} (\boldsymbol{u}_a)_i = C$ for some $C \in \mathbb{N}$ we posit the following equality:

$$\operatorname*{argmax}_{a \in A} \; U_{\text{nash}}(\boldsymbol{u}_a) = \operatorname*{argmax}_{a \in A} \prod_{i \in N} (\boldsymbol{u}_{\boldsymbol{a}})_i = \operatorname*{argmin}_{a \in A} \sum_{i \in N} \left| (\boldsymbol{u}_a)_i - \frac{C}{n} \right| \tag{2.4}$$

As a final remark, note that the three social welfare functions we defined above (utilitarian, egalitarian and Nash) do not depend on the specific order of the (individual) utilities. In particular, for $\odot \in \{\text{util}, \text{egal}, \text{Nash}\}$ and an arbitrary vector $\boldsymbol{u} \in \mathbb{R}^n$, we have:

$$U_\odot(\boldsymbol{u}) = U_\odot(\boldsymbol{u}^{\text{lex}}) \tag{2.5}$$

2.3.2 Judgement Aggregation as Welfare Aggregation

In this section we study judgement aggregation procedures from the perspective of welfare economics. With a social welfare function, economically feasible allocations are ranked according to the societal welfare they induce. On the other hand, an aggregation procedure returns the collective judgements that represent the opinions (preferences or beliefs) of the individuals in the best possible way. It is natural to assume that individual judges have preferences over the set of collective judgements. Moreover, in the literature, there is a commonly accepted way to define such preferences—so called Hamming preferences—on the basis of the individual judgements. On the basis of these preferences we propose utility functions for the individual judges. Showing that the Kemeny rule adheres to utilitarian social welfare is then straightforward.

The remainder of this section is organised as follows. We start with a brief remark on notation, concerning the representation of individual utilities as multisets instead of vectors. We continue by defining individual preference orders, which leads us to the definition of individual utility functions. To conclude we show that the Kemeny rule maximises utilitarian social welfare and propose a novel judgement aggregation procedure that maximises Nash social welfare: the Kemeny-Nash rule.

In welfare economics it is customary to use a non-anonymous framework; the utilities of the individual agents are represented as vectors. In such a framework it makes sense to define a preference order $\succeq_i$ for every agent $i \in N$. In our framework this is at least confusing, but also (slightly) incorrect; we cannot distinguish any two judges other than by their judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$. So, in our framework we define individual preference orders $\succeq_{J_{\text{in}}}$ and utilities $u_{J_{\text{in}}}$ that are parameterised by the judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$, rather than by the judge $i \in N$ that holds it.

It is natural to assume that judges participating in a judgement aggregation scenario have preferences over the set of feasible judgements. Formally, we may assume that a judge with judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$ is endowed with a preference order $\succeq_{J_{\text{in}}} : |\mathcal{J}(\Phi, \Gamma_{\text{out}})| \to \mathbb{N}$. Although it is not the only accepted method; in the judgement aggregation literature, the prevalent way to establish such preference orders is to endow the judges with *Hamming preferences* (Botan et al., 2021, 2023; Baumeister et al., 2015, 2017).

Let $J, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ be arbitrary feasible judgements. A judge $i \in N$ with judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$ is said to have *Hamming preferences* when they (weakly) prefer judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ over judgement $J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ if and only if the Hamming distance between their own judgement J_{in} and judgement J is not larger than the Hamming distance between their own judgement J_{in} and judgement J':

$$J \succeq_{J_{\text{in}}} J' \qquad \text{if and only if} \qquad H(J_{\text{in}}, J) \leq H(J_{\text{in}}, J')$$

Of course, the statement above is equivalent to:

$$J \succeq_{J_{\text{in}}} J' \qquad \text{if and only if} \qquad \text{Agr}(J_{\text{in}}, J) \geq \text{Agr}(J_{\text{in}}, J')$$

As we want to use the preference orders to construct utility functions, the latter representation is more suitable: a higher agreement indicates a more preferred feasible judgement, while a higher Hamming distance indicates a less preferred judgement. We give the formal definition of our assumed preference orders below.

Definition 2.5. Consider a judgement aggregation scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ with n judges and let $J, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ be arbitrary feasible judgements. For a judge $i \in N$ with judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$, we define the following binary relation $\succeq_{J_{\text{in}}} : \mathcal{J}(\Phi, \Gamma_{\text{out}}) \times \mathcal{J}(\Phi, \Gamma_{\text{out}}) \to \{0, 1\}$:

$$J \succeq_{J_{\text{in}}} J' \qquad \text{if and only if} \qquad \text{Agr}(J_{\text{in}}, J) \geq \text{Agr}(J_{\text{in}}, J')$$

It is easy to verify that the defined preference orders satisfy completeness and transitivity.

The preference orders above (Definition 2.5) determine the utility functions up to a positive affine transformation. For any judge with judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$ the utility function $u_{J_{\text{in}}} : \mathcal{J}(\Phi, \Gamma_{\text{out}}) \to \mathbb{R}$ is constrained by:

$$\dot{u}_{J_{\text{in}}}(J) = \alpha \cdot \text{Agr}(J_{\text{in}}, J) + \beta, \quad \text{with} \;\; \alpha \in \mathbb{R}_{>0}, \beta \in \mathbb{R}_{\geq 0}$$

If we set the parameters $\alpha \in \mathbb{R}_{>0}$ and $\beta \in \mathbb{R}_{\geq}0$, we have fixed unique utility functions $u_{J_{\text{in}}}$. The values of these parameters cannot be derived by reason; we stipulate $\alpha = 1$ and $\beta = 0$. That is, we stipulate that if a judge is in complete disagreement with the collective judgement, their utility is zero; they have nothing to be satisfied with. Further we suppose that for every issue $\varphi \in \Phi$ they (judge i) share with the collective judgement, their utility is increased by 1. We give the formal definition of our utility functions below.

Definition 2.6. Given a judgement aggregation scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ with n judges, let $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ be an arbitrary feasible judgement. For judge $i \in N$ with judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$, we define the utility function $u_{J_{\text{in}}} : \mathcal{J}(\Phi, \Gamma_{\text{out}}) \to \{1, \ldots, |\Phi^+|\}$:

$$u_{J_{\text{in}}}(J) = \text{Agr}(J_{\text{in}}, J) \tag{2.6}$$

Note that the individual utilities $u_{J_{\text{in}}}$ are elements of a finite set $\{1, \ldots, |\Phi^+|\}$. Hence, for an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ we can define a multiset of individual utilities $\boldsymbol{u}^{\text{ms}} : \{1, \ldots, |\Phi^+|\} \to \mathbb{N}$ as follows:

$$\boldsymbol{u}_J^{\text{ms}}(u') = \sum_{\substack{J_{\text{in}} \in \boldsymbol{J} \text{ s.t.} \\ \text{Agr}(J_{\text{in}}, J) = u'}} \boldsymbol{J}(J_{\text{in}}) \tag{2.7}$$

That is, the multiplicity of utility $u' \in \{1, \ldots, |\Phi|\}$ equals the sum of multiplicities of all judgements $J_{\text{in}} \in \boldsymbol{J}$ in the profile that are associated with utility u'.

Example 2.9. Let $\boldsymbol{J} = \{\{\neg\varphi_1, \neg\varphi_2\}^1, \{\varphi_1, \neg\varphi_2\}^1, \{\varphi_1, \varphi_2\}^2\}$ be an arbitrary profile and let $J = \{\neg\varphi_1, \varphi_2\}$ for some scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$. For profile $\boldsymbol{J}$ the multiset of utilities is defined as $\boldsymbol{u}_J^{\text{ms}} = \{0^1, 1^3\}$. △

In the previous section we defined the social welfare functions for utility vectors, not for multisets. However, as we mentioned above (Equation 2.5), the SWFs do not depend on the order of the individual utilities. It is straightforward to restate the different social welfare functions for multisets of individual utilities.

Definition 2.7. For an arbitrary multiset $\boldsymbol{u}_J^{\text{ms}} : \{1, \ldots, |\Phi^+|\} \to \mathbb{N}$, the utilitarian social welfare function $U_{\text{util}}^{\text{ms}}$ is defined as:

$$U_{\text{util}}^{\text{ms}}(\boldsymbol{u}_J^{\text{ms}}) = \sum_{u' \in \boldsymbol{u}_J^{\text{ms}}} (\boldsymbol{u}_J^{\text{ms}}(u') \cdot u')$$

Definition 2.8. For an arbitrary multiset $\boldsymbol{u}_J^{\text{ms}} : \{1, \ldots, |\Phi^+|\} \to \mathbb{N}$, the egalitarian social welfare function $U_{\text{egal}}^{\text{ms}}$ is defined as:

$$U_{\text{egal}}^{\text{ms}}(\boldsymbol{u}_J^{\text{ms}}) = \min_{u' \in \boldsymbol{u}_J^{\text{ms}}} u'$$

Definition 2.9. For an arbitrary multiset $\boldsymbol{u}_J^{\text{ms}} : \{1, \ldots, |\Phi^+|\} \to \mathbb{N}$, the Nash social welfare function $U_{\text{nash}}^{\text{ms}}$ is defined as:

$$U_{\text{nash}}^{\text{ms}}(\boldsymbol{u}_J^{\text{ms}}) = \prod_{u' \in \boldsymbol{u}_J^{\text{ms}}} (\boldsymbol{u}_J^{\text{ms}}(u') \cdot u')$$

Finally, we have established the tools we need to derive our conclusion: the Kemeny rule maximises utilitarian social welfare. Given an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ with n judges we have:

$$\begin{aligned} F_{\text{kem}}(\boldsymbol{J}) &= \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} \sum_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \operatorname{Agr}(J_{\text{in}}, J)) \\ &= \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} \sum_{u' \in \boldsymbol{u}_J^{\text{ms}}} \big(\boldsymbol{u}_J^{\text{ms}}(u') \cdot u'\big) \\ &= \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} U_{\text{util}}^{\text{ms}}(\boldsymbol{u}_J^{\text{ms}}) \end{aligned}$$

The first line is the definition of the Kemeny rule (Definition 2.1). The second equality follows from the definition of $\boldsymbol{u}_J^{\text{ms}}$ (Equation 2.7). The final step follows directly from the definition of the utilitarian SWF for multisets of utilities (Definition 2.7).

In the judgement aggregation literature there is no rule that, in a similar way as the Kemeny rule does for utilitarian social welfare, maximises Nash social welfare. In other fields of social choice theory, in particular in fair allocation, maximising Nash social welfare is known to yield allocations that are both efficient and fair (Caragiannis et al., 2019). It would be relevant to examine whether collective judgements that result from maximising Nash social welfare are comparable in terms of fairness and efficiency. Reversing the steps we took above, this time starting from the Nash SWF for multisets (Definition 2.9) we obtain the rule that does this. We introduce the novel judgement aggregation procedure, the *Kemeny-Nash rule*:

$$\underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} U_{\text{nash}}^{\text{ms}}(\boldsymbol{u}_J^{\text{ms}}) = \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} \prod_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \operatorname{Agr}(J_{\text{in}}, J)) = F_{\text{kn}}(\boldsymbol{J}) \tag{2.8}$$

The Kemeny-Nash rule is formally defined in the next section.

2.4 The Kemeny-Nash Rule

In this section we make our first exploration into the Kemeny-Nash rule F_{kn}. We start by providing the formal definition of the Kemeny-Nash rule and reason about the differences between the Kemeny and Kemeny-Nash rule.

Definition 2.10. For an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{out}})^n$ with n judges we define the Kemeny-Nash rule as follows:

$$F_{\text{kn}}(\boldsymbol{J}) = \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} \prod_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \operatorname{Agr}(J_{\text{in}}, J)) \tag{2.9}$$

Similarly to the Kemeny score, we define the Kemeney-Nash score $S_{\text{kn}}(\boldsymbol{J}, J) = \prod_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \text{Agr}(J_{\text{in}}, J))$. In the previous section we argued that this rule maximises Nash social welfare (NSW). In other areas of social choice theory, maximising NSW is known to provide solutions that are both efficient and fair (Caragiannis et al., 2019); that is why it is relevant to study the Kemeny-Nash judgement aggregation procedure.

Before we start our investigation into the properties of the Kemeny-Nash rule, a brief remark on formalities. Given a multiset $\boldsymbol{x}$ we define $\langle \boldsymbol{x} \rangle$ the average of $\boldsymbol{x}$ as:

$$\langle \boldsymbol{x} \rangle = \frac{1}{|\boldsymbol{x}|} \sum_{x \in \boldsymbol{x}} (\boldsymbol{x}(x) \cdot x)$$

Further, for an arbitrary multiset of individual utilities $\boldsymbol{u}_J^{\text{ms}}$ (Equation 2.7) we define the *inequality* $I(\boldsymbol{u}_J^{\text{ms}})$ as the sum of the absolute differences between the individual utilities and the average utility:

$$I(\boldsymbol{u}_J^{\text{ms}}) = \sum_{u' \in \boldsymbol{u}_J^{\text{ms}}} \left(\boldsymbol{u}_J^{\text{ms}}(u') \cdot |u' - \langle \boldsymbol{u}_J^{\text{ms}} \rangle| \right) \tag{2.10}$$

We propose that this is a relevant measure for the inequality of a feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$. Unquestionably, there are other measures that are just as relevant to measure the inequality of a feasible judgement. E.g., the minimum utility; the ruler that is used in the egalitarian SWF. As a justification for our measure, we note that for a multiset $\boldsymbol{u}_J^{\text{ms}} = \{c^n\}$ in which all judges have the same utility—without doubt the most equal distribution conceivable—the inequality $I(\boldsymbol{u}_J^{\text{ms}}) = 0$. Reasoned differently, if $I(\boldsymbol{u}_J^{\text{ms}})$ is large, some judges enjoy a high level of satisfaction that is to be compensated by the judges that have a low utility (in order to keep the average utility fixed). We note that this measure is meant to compare the equity of a rule F and F' for a fixed scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$.

Finally, when a judgement aggregation rule returns multiple collective judgements, exactly one of these judgements is materialised; a tie-breaking procedure has to take place. We do not consider the exact details of such a procedure here. However, we do make the following mild assumption: in case a judgement rule F returns a tied outcome $F(\boldsymbol{J})$, with $|F(\boldsymbol{J})| > 1$, then all returned collective judgements $J \in F(\boldsymbol{J})$ have a non-zero probability of being realised (after some tie-breaking procedure has been executed). That is to say, (the expected value of) an arbitrary property of a rule is affected by all of its returned judgements. Because we do not want to treat tie-breaking rules, expected values or probability distributions in detail, the statement here is imprecise. Our point will be clarified in the next example.

Example 2.10. Consider the judgement aggregation scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$, with the agenda Φ that is based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2, \varphi_1 \wedge \varphi_2\}$, the constraints $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \top$ coincide and the profile $\boldsymbol{J} = \{J_1^1, J_2^1\}$ contains judgements $J_1 =$

$\{\neg\varphi_1, \varphi_2, \neg(\varphi_1 \wedge \varphi_2)\}$ and $J_2 = \{\varphi_1, \neg\varphi_2, \neg(\varphi_1 \wedge \varphi_2)\}$. The Kemeny and Kemeny-Nash scores, as well as the agreement between judgement $J_{\text{in}} \in \{J_1, J_2\}$ and J, for the four different consistent judgements are given by:

Feasible judgement J	$\mathrm{Agr}(J_1, J)$	$\mathrm{Agr}(J_2, J)$	$S_{\text{kem}}(\boldsymbol{J}, J)$	$S_{\text{kn}}(\boldsymbol{J}, J)$
$\{\neg\varphi_1, \neg\varphi_2, \neg(\varphi_1 \wedge \varphi_2)\}$	2	2	4	4
$\{\neg\varphi_1, \varphi_2, \neg(\varphi_1 \wedge \varphi_2)\}$	3	1	4	3
$\{\varphi_1, \neg\varphi_2, \neg(\varphi_1 \wedge \varphi_2)\}$	1	3	4	3
$\{\varphi_1, \varphi_2, (\varphi_1 \wedge \varphi_2)\}$	1	1	2	1

In this example the Kemeny rule returns three collective judgements; for two of these judgements one of the judges has agreement 1 while the other judge has agreement 3 with the collective judgement, for the other collective judgement both judges have agreement 2. The Kemeny-Nash rule returns only the collective judgement for which the agreement of both judges is 2. By using the Kemeny-Nash rule, instead of the Kemeny rule, the efficiency of the returned collective judgements is not affected. However, because we assume that every returned judgement has a non-zero probability of being realised after some tie-break procedure, the (expected) inequality of the Kemeny rule is higher. △

In the above example the Kemeny-Nash rule is certainly more favourable than the Kemeny rule; we gain economic equity without sacrificing economic efficiency. Let us consider another example in which the economic equity is increased by the Kemeny-Nash rule; this time it comes with a sacrifice of economic efficiency. Moreover, the example illustrates a peculiar facet of the Kemeny-Nash rule: a feasible judgement $J = \overline{J_{\text{in}}}$ that is the antipodal of any of the judgements $J_{\text{in}} \in (\boldsymbol{J})$ contained in the profile has a Kemeny-Nash score of 0. We refer to this feature of the Kemeny-Nash rule as the *Zero-Effect*, we provide the formal definition below.

Example 2.11. Consider the judgement aggregation scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$, with the agenda Φ that is based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2\}$, the constraints $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \top$ coincide and the profile $\boldsymbol{J} = \{J_1^2, J_2^1\}$ contains judgements $J_1 = \{\neg\varphi_1, \neg\varphi_2\}$ and $J_2 = \{\varphi_1, \varphi_2\}$. The Kemeny and Kemeny-Nash scores, for the four different consistent judgements are given by:

Feasible judgement J	$\mathrm{Agr}(J_1, J)$	$\mathrm{Agr}(J_2, J)$	$S_{\text{kem}}(\boldsymbol{J}, J)$	$S_{\text{kn}}(\boldsymbol{J}, J)$
$\{\neg\varphi_1, \neg\varphi_2\}$	2	0	4	0
$\{\neg\varphi_1, \varphi_2\}$	1	1	3	1
$\{\varphi_1, \neg\varphi_2\}$	1	1	3	1
$\{\varphi_1, \varphi_2\}$	0	2	2	0

For this profile the Kemeny rule $F_{\text{kem}}(\boldsymbol{J}) = \{\neg\varphi_1, \neg\varphi_2\} = J_{\text{kem}}$ returns a single collective judgement, two of the judges are in complete agreement and have utility 2. The other judge completely disagrees with the collective judgement and has utility 0.

The Kemeny-Nash rule $F_{\mathrm{kn}}(\boldsymbol{J}) = \{\{\neg\varphi_1, \varphi_2\}, \{\varphi_1, \neg\varphi_2\}\}$ returns a tie between two judgements, for both collective judgements, all three judges have utility 1.

In this scenario the average utility induced by the Kemeny rule equals $\langle \boldsymbol{u}^{\mathrm{ms}}_{J_{\mathrm{kem}}} \rangle = 4/3$, while the average utility $\langle \boldsymbol{u}^{\mathrm{ms}}_{J} \rangle = 1$ for the judgements $J \in F_{\mathrm{kn}}(\boldsymbol{J})$. On the positive side, the minimal utility $\min_{u' \in \boldsymbol{u}^{\mathrm{ms}}_J} u' = 1$ for the Kemeny-Nash collective judgements, while it is $\min_{u' \in \boldsymbol{u}^{\mathrm{ms}}_{J_{\mathrm{kem}}}} = 0$ for the Kemeny judgement. Moreover, for the Kemeny rule we have: $I(\boldsymbol{u}^{\mathrm{ms}}_{J_{\mathrm{kem}}}) = 2 \cdot \frac{2}{3} + \frac{4}{3} = \frac{8}{3}$. For the multisets of utilities that correspond to the Kemeny-Nash judgements $J \in F_{\mathrm{kn}}(\boldsymbol{J})$, we have $I(\boldsymbol{u}^{\mathrm{ms}}_J) = 0$. In conclusion, compared to the Kemeny rule, the Kemeny-Nash rule sacrifices economic efficiency on the one side, but it gains economic equity on the other side.

Finally, it is noteworthy that $S_{\mathrm{kn}}(\boldsymbol{J}, J_{\mathrm{kem}}) = 0$; the single judgement that is returned by the Kemeny rule can only be returned by the Kemeny-Nash rule if all feasible judgement have Kemeny-Nash score of 0, in that case all feasible judgements are returned by the Kemeny-Nash rule. The reason that the Kemeny-Nash score $S_{\mathrm{kn}}(\boldsymbol{J}, J_{\mathrm{kem}}) = 0$ because judgement $J_{\mathrm{kem}} = \overline{J_2}$, resulting in a zero multiplication, we term this effect the *Zero-Effect.* $\triangle$

The formal definition of the Zero-Effect is given below.

Definition 2.11. Let $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\mathrm{in}})^n$ be an arbitrary profile, and let $J \in \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$ be any feasible judgement. We define the function $\mathrm{ZE}_{\boldsymbol{J}} : \mathcal{J}(\Phi, \Gamma_{\mathrm{out}}) \to \{0, 1\}$ as follows:

$$\mathrm{ZE}_{\boldsymbol{J}}(J) = \begin{cases} 1, & \text{if } \overline{J} \in \boldsymbol{J} \\ 0, & \text{otherwise} \end{cases}$$

For any judgement $J \in \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$ with $\mathrm{ZE}_{\boldsymbol{J}}(J) = 1$ the Kemeny-Nash score $S_{\mathrm{kn}}(\boldsymbol{J}, J) = 0$; the probability that judgement $J \in F_{\mathrm{kn}}(\boldsymbol{J})$ is selected by the Kemeny-Nash rule is zeroed out.[5]

In the previous example the ZE did not appear to be a (potential) weakness; it led to some efficiency loss but also to a gain in equity. In an early stage of this work we assumed that the Zero-Effect (in some cases) may be a weakness of the Kemeny-Nash rule. For profile $\boldsymbol{J}$, a single judge with judgement $J_{\mathrm{in}} \in \boldsymbol{J}$ is sufficient to preclude the possibility that $\overline{J_{\mathrm{in}}} \in F_{\mathrm{kn}}(\boldsymbol{J})$ is selected by the Kemeny-Nash rule. We reasoned that for scenarios where basically all judges $n' \approx n$ have judgements that are very similar to $\overline{J_{\mathrm{in}}}$; precluding the possibility that judgement $\overline{J_{\mathrm{in}}}$ is selected by the Kemeny-Nash rule might damage the economic efficiency (for that scenario) disproportionately, compared to the gained economic equity.

In an attempt to designate the scenarios for which the ZE might be a weakness, we introduced the Variance-Increasing Zero-Effect. In this stage of the work we

[5] Technically, judgement J can still be selected by the Kemeny-Nash rule; $J \in F_{\mathrm{kn}}(\boldsymbol{J})$ if and only if $S_{\mathrm{kn}}(\boldsymbol{J}, J') = 0$ for all feasible judgements $J' \in \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$, then $F_{\mathrm{kn}}(\boldsymbol{J}) = \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$. Although the judgement is indeed technically selected, from a practical point of view we might as well say that none of the judgements is selected.

believe our previous argument is flawed. We explain our initial idea behind the VIZE. By use of an example we show the problem of our previous train of thought.

As mentioned, in the previous example the ZE did not appear to be a (potential) weakness. However, as the ZE is independent of the number of judges—if a single judge with judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$ completely disagrees with judgement $\overline{J_{\text{in}}} = J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$; it is precluded that judgement J is selected by the Kemeny-Nash rule. It is reasonable to suggest that, in some scenarios, the power that a single judge has to preclude a particular judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ from being selected as collective judgement, should be seen as a deficiency of the Kemeny-Nash rule.

In an attempt to demarcate the scenarios for which the ZE should be considered as a (potential) weakness of the Kemeny-Nash rule we defined the *Variance-Increasing Zero-Effect* (VIZE).

Definition 2.12. For an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ let $J \in F_{\text{kn}}(\boldsymbol{J})$ be any judgement that is returned by the Kemeny-Nash rule. We define the *Variance-Increasing Zero-Effect* (VIZE) as:

$$\text{VIZE}_{\boldsymbol{J}}(J) = \begin{cases} 1 & \text{if } I(\boldsymbol{u}_J^{\text{ms}}) > I(\boldsymbol{u}_{J'}^{\text{ms}}) \text{ for all } J' \in F_{\text{kem}}(\boldsymbol{J}) \\ 0 & \text{otherwise} \end{cases}$$

That is, a judgement $J \in F_{\text{kn}}(\boldsymbol{J})$ is counted as an instance of the VIZE if the sum of distances from the individual utilities to the average utility (multiplied by the multiplicities of the individual utilities) is larger for (the utilities induced by) judgement $J \in F_{\text{kn}}(\boldsymbol{J})$ than for any judgement $J' \in F_{\text{kem}}(\boldsymbol{J})$ returned by the Kemeny rule (See Equation 2.10).

Example 2.12. Consider a profile $\boldsymbol{J} = \{J_{\text{maj}}^{n_{\text{maj}}}, J_{\text{min}}^{n_{\text{min}}}\}$ with $J_{\text{maj}}, J_{\text{min}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$, $J_{\text{maj}}, J_{\text{min}} \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, $J_{\text{min}} = \overline{J_{\text{maj}}}$ and $n_{\text{maj}} \gg n_{\text{min}}$; we let $n = n_{\text{maj}} + n_{\text{min}}$.

Our initial idea was that the efficiency of the Kemeny-Nash rule would be disproportionally affected if judgement J_{maj} cannot be returned by the Kemeny rule: in theory the satisfaction of a large number of judges is damaged to increase the satisfaction of a single (or maybe a few) judges. We now abandon this view.

Yes, the sum of utilities that is lost by the majority of the judges outweighs the sum of utilities that is gained by the minority. But we should realise that this reasoning holds only if the majority of the judges enjoys a significantly higher utility than the judges that make up the minority. The idea of the egalitarian approach is that the will of the minority cannot be disregarded by the will of the majority; and we introduced the Kemeny-Nash rule with the assumption that it is a combination of fairness and efficiency considerations. At this point, we see no reason to assume that VIZE instances indicate a disproportional loss in economic efficiency.

Maybe, in scenarios where the output constraint Γ_{out} is very limiting, and there is no feasible judgement that is close to J_{maj}, the VIZE might be a weak point. However, this had nothing to do with our initial idea, and we consider it improbable that these cases affect (the efficiency of) the Kemeny-Nash rule in a significant way. △

We argued that the potentially worrying VIZE (if it indeed turned out to be a deficiency of the Kemeny-Nash rule) might be mitigated by introducing a small nonzero parameter λ in the Kemeny-Nash rule; we proposed the parameterised Kemeny-Nash rule.

Definition 2.13. The parameterised Kemeny-Nash rule is defined as:

$$F_{\text{kn}}^{\lambda}(\boldsymbol{J}) = \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\text{argmax}} \prod_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \max\{\text{Agr}(J_i, J), \lambda\}), \qquad \text{with} \quad 0 < \lambda \ll 1.$$

Similar to the Kemeny and Kemeny-Nash score, we define the parameterised Kemeny-Nash score $S_{\text{kn}}^{\lambda}(\boldsymbol{J}, J, \lambda) := \prod_{i \in N} \max\{\text{Agr}(J_i, J), \lambda\}$. The Kemeny-Nash rule is studied throughout the rest of this thesis. Although our reasoning above is now abandoned, as we will see in Chapter 3, in which we study desirable normative principles (or *axioms*) for judgement aggregation procedures; from this perspective the parameterised Kemeny-Nash rule is still a relevant addition the the Kemeny-Nash rule.

As a minor remark, when we talk about the Kemeny-Nash rule we explicitly mean the unparameterised variant F_{kn}. In our further examination we often study both the Kemeny-Nash and the parameterised Kemeny-Nash, for readability we write (parameterised) Kemeny-Nash rule in such cases.

To conclude this section we treat a particular scenario in full detail, and discuss how the outcomes that are returned by the Kemeny and (parameterised) Kemeny-Nash rule differ from one another.

Example 2.13. To see how the different rules work we consider a scenario from Endriss et al. (2020). Mathematically, the scenario is defined as follows. We have $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$, where agenda Φ is based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4\}$, and the input and output constraints coincide $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \neg\varphi_1 \vee (\neg\varphi_2 \wedge \neg\varphi_4) \vee \neg\varphi_3$. Further, let the profile $\boldsymbol{J}$ contain two judgements: $\boldsymbol{J} = \{J_1^1, J_2^1\}$. Finally, we stipulate: $J_1 = \{\varphi_1, \varphi_2, \neg\varphi_3 \neg\varphi_4\}$ and $J_2 = \{\neg\varphi_1, \neg\varphi_2, \varphi_3, \neg\varphi_4\}$. Intuitively, this scenario could correspond to the situation where two mechanics judge the cause of a failure. The failure could be caused by a fault in either component 1 or component 3 or by a failure of both components 2 and 4. Then, the relation between the mathematical and intuitive description is that the positive literal φ_i stands for component i works; while $\neg\varphi_i$ indicates that component i does not work.

From the input and output constraints we can reason that there are thirteen feasible judgements, i.e., $|\mathcal{J}(\Phi, \Gamma_{\text{out}})| = 13$. The judgement that all four components are faulty, $J = \{\neg\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4\}$ is feasible. All (four) judgements that state that one component works (while the other three components do not work) are feasible; e.g., judgement $J = \{\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4\}$ is feasible. It is also clear that all judgements that judge two components to function, and two components to be damaged, are feasible; e.g., judgement $J = \{\varphi_1, \varphi_2, \neg\varphi_3, \neg\varphi_4\}$ is feasible. From the four judgements that declare a single component to be faulty, only the two judgements that declare either component 1 or component 3 to be flawed, are feasible. For example, the judgement $J = \{\neg\varphi_1, \varphi_2, \varphi_3, \varphi_4\}$ is feasible. Finally, the judgement that says that all components are working properly is not feasible.

Let $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ be any feasible judgement. Table 2.2 shows the judgement J, and the agreement of J with judgement J_1 and J_2 (respectively): $\text{Agr}(J_1, J)$ and $\text{Agr}(J_2, J)$. Further, Table 2.2 shows the Kemeny and parameterised Kemeny-Nash score of J (for profile $\boldsymbol{J}$): $S_{\text{kem}}(\boldsymbol{J}, J)$ and $S^{\lambda}_{\text{kn}}(\boldsymbol{J}, J, \lambda)$, respectively. The maximum scores are printed in bold face. Note that the Kemeny-Nash score is obtained from the parameterised Kemeny-Nash score, simply by setting λ to 0. That is, $S_{\text{kn}}(\boldsymbol{J}, J) = S^{\lambda}_{\text{kn}}(\boldsymbol{J}, J, 0)$.

Now, let us compare the collective judgements returned by the Kemeny rule with those that are returned by the (parameterised) Kemeny-Nash rule. First, observe that for all $\lambda < 1$ we have: $3\lambda < 6$. In this particular scenario, the collective judgements of the parameterised Kemeny-Nash rule are not influenced by the value of λ, and they coincide with those of the Kemeny-Nash rule: $F_{\text{kn}}(\boldsymbol{J}) = F^{\lambda}_{\text{kn}}(\boldsymbol{J})$. Further, we can see that the set of judgements that is returned by the (parameterised) Kemeny-Nash rule is a proper subset of the set that is returned by the Kemeny rule: $F_{\text{kn}}, F^{\lambda}_{\text{kn}} \subset F_{\text{kem}}$. That is, the efficiency of (the collective judgements that are returned by) the different rules coincide. However, we can also see that for all collective judgements that are returned by the (parameterised) Kemeny-Nash rule, the agreement difference between the two judges is equal to 1. Rephrased in a mathematical precise way, for all $J' \in F^{\lambda}_{\text{kn}}(\boldsymbol{J})$ (alternatively, for all $J' \in F_{\text{kn}}(\boldsymbol{J})$) we have: $|\text{Agr}(J_1, J') - \text{Agr}(J_2, J')| = 1$. While, for some judgements $J'' \in F_{\text{kem}}(\boldsymbol{J})$, selected by the Kemeny rule, this difference is equal to 3: $|\text{Agr}(J_1, J'') - \text{Agr}(J_2, J'')| = 3$.

In conclusion, in this scenario the (parameterised) Kemeny-Nash rule selects judgements that are (compared to the Kemeny rule) equal in terms of efficiency, but better in terms of fairness; as they induce a more equal distribution of satisfaction among the judges. $\triangle$

Feasible judgement J	$\mathrm{Agr}(J_1, J)$	$\mathrm{Agr}(J_2, J)$	$S_{\mathrm{kem}}(\boldsymbol{J}, J)$	$S^{\lambda}_{\mathrm{kn}}(\boldsymbol{J}, J, \lambda)$
$\{\neg\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4\}$	2	3	**5**	**6**
$\{\neg\varphi_1, \neg\varphi_2, \neg\varphi_3, \varphi_4\}$	2	1	3	2
$\{\neg\varphi_1, \neg\varphi_2, \varphi_3, \neg\varphi_4\}$	2	3	**5**	**6**
$\{\neg\varphi_1, \varphi_2, \neg\varphi_3, \neg\varphi_4\}$	4	1	**5**	4
$\{\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4\}$	2	3	**5**	**6**
$\{\neg\varphi_1, \neg\varphi_2, \varphi_3, \varphi_4\}$	2	1	3	2
$\{\neg\varphi_1, \varphi_2, \neg\varphi_3, \varphi_4\}$	1	4	**5**	4
$\{\neg\varphi_1, \varphi_2, \varphi_3, \neg\varphi_4\}$	1	2	3	2
$\{\varphi_1, \neg\varphi_2, \neg\varphi_3, \varphi_4\}$	3	0	3	3λ
$\{\varphi_1, \neg\varphi_2, \varphi_3, \neg\varphi_4\}$	3	2	**5**	**6**
$\{\varphi_1, \varphi_2, \neg\varphi_3, \neg\varphi_4\}$	0	3	3	3λ
$\{\neg\varphi_1, \varphi_2, \varphi_3, \varphi_4\}$	1	2	3	2
$\{\varphi_1, \varphi_2, \neg\varphi_3, \varphi_4\}$	3	2	**5**	**6**

Table 2.2: Analysis of the consistent judgements corresponding to the scenario considered in Example 2.13. (Profile $\boldsymbol{J}$ contains judgements $J_1 = \{\varphi_1, \varphi_2, \neg\varphi_3, \neg\varphi_4\}$ and $J_2 = \{\neg\varphi_1, \neg\varphi_2, \varphi_3, \neg\varphi_4\}$.) For every feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$ we show (from left to right): the judgement J; the agreement of J with judgement J_1, $\mathrm{Agr}(J_1, J)$; the agreement of J with judgement J_2, $\mathrm{Agr}(J_2, J)$; the Kemeny score of J (for profile $\boldsymbol{J}$), $S_{\mathrm{kem}}(\boldsymbol{J}, J)$; and the parameterised Kemeny-Nash score of J, $S^{\lambda}_{\mathrm{kn}}(\boldsymbol{J}, J, \lambda)$. (Note that the Kemeny-Nash score is computed by substituting $\lambda = 0$ in the parameterised Kemeny-Nash score: $S_{\mathrm{kn}}(\boldsymbol{J}, J) = S^{\lambda}_{\mathrm{kn}}(\boldsymbol{J}, J, 0)$.) The maximal Kemeny and parameterised Kemeny-Nash scores are printed bold.

Chapter 3

Theoretical Analysis I: Axiomatics

List and Pettit (2002) showed that the conflict exhibited by the discursive dilemma is not restricted to the field of legal theory. As we mentioned in the Introduction (Chapter 1), they established that it is impossible to aggregate individual judgements made on a set of logically connected issues, in such a way that a group of (seemingly) undemanding normative principles or *axioms* is satisfied. The result is exemplary for the axiomatic approach, which is the subject of this chapter.

The chapter is organised as follows. In Section 3.1 we start by introducing the axiomatic approach. In Section 3.2, we examine the work of Nehring and Pivato (2022), who showed that the Kemeny rule is the only rule that satisfies a particular group of axioms (in that sense, the group of axioms characterises the Kemeny rule).[1] As we will see, from the four axioms that are used, the Kemeny-Nash rule violates all four, while the parameterised Kemeny-Nash rule violates only one of the axioms. Next, in Section 3.3 we treat four more axioms to further study the axiomatic properties of the (parameterised) Kemeny-Nash rule. We conclude the chapter in Section 3.4.

Before we begin, a remark on notation is in order. In this chapter we frequently consider scenarios in which the agenda Φ is based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2\}$ containing two propositional variables. For the sake of readability, we define judgements $J_{00} = \{\neg\varphi_1, \neg\varphi_2\}$, $J_{01} = \{\neg\varphi_1, \varphi_2\}$, $J_{10} = \{\varphi_1, \neg\varphi_2\}$ and $J_{11} = \{\varphi_1, \varphi_2\}$. As a mnemonic, the subscripts may be recognised as indicating whether the corresponding non-negated issue (contained in the pre-agenda) is accepted (1) or rejected (0).

[1] To be precise, Nehring and Pivato (2022) provide two distinct (but logically related) characterisations of the Kemeny rule.

3.1 Introduction

The purpose of this section is to introduce the axiomatic approach, one of the fundamental instruments in social choice theory. We present the renowned theorem by List and Pettit (2002), and distinguish this type of result (impossibility theorems) from another type of result (characterisations).

From the very beginning—the seminal work of Arrow (1951)—the *axiomatic approach* is one of the vital instruments used in social choice theory. An *axiom* is a normative principle, of which we consider it desirable that a collective decision procedure complies with it. By comparing the axiomatic properties of such procedures we can make a more deliberate choice between the different alternatives.

When it comes to the role of the axiomatic approach, the field of judgement aggregation is no exception; also in this area the technique is indispensable. As we saw in the Introduction (Chapter 1), judgement aggregation—as a dedicated research area—started with the work of List and Pettit (2002). We review their result below.

There are substantial differences between the framework of List and Pettit (2002) and our framework. To wit, the former (i) is non-anonymous; profiles are represented as tuples $\boldsymbol{J} = (J_1, \ldots, J_n)$ containing n rational judgements, and (ii) uses resolute aggregation rules f, mapping a profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ to a single feasible judgement (see Section 2.2). Hence, the axioms below are requirements for all *ordered* profiles $\boldsymbol{J} = (J_1, \ldots, J_n) \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, and are defined for *resolute* rules f.[2]

Universal Domain (`UD`): A feasible judgement aggregation rule f should return a collective judgement for every possible n-tuple of rational judges; that is, $f : \mathcal{J}(\Phi, \Gamma_{\text{in}})^n \to \mathcal{J}(\Phi, \Gamma_{\text{out}})$.

Anonymity (`A`): For arbitrary profiles $\boldsymbol{J} = (J_1, \ldots, J_n)$ and $\boldsymbol{J}' = (J'_1, \ldots, J'_n)$ that are permutations of one another, *anonymity* demands that a resolute procedure f returns $f(\boldsymbol{J}) = f(\boldsymbol{J}')$ identical collective judgements.

Systematicity (`S`): Let $\boldsymbol{J} = (J_1, \ldots, J_n)$ and $\boldsymbol{J}' = (J'_1, \ldots, J'_n)$ be arbitrary profiles for n judges, with $\boldsymbol{J}, \boldsymbol{J}' \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, and $\varphi, \vartheta \in \Phi$ any two formulas in the agenda. If, for all judges $i \in N$ we have $\varphi \in J_i$ if and only if $\vartheta \in J'_i$, then *systematicity* requires that for a resolute rule f we have: $\varphi \in f(\boldsymbol{J})$ if and only if $\vartheta \in f(\boldsymbol{J}')$.

Collective Rationality (`CR`): For any profile $\boldsymbol{J} = (J_1, \ldots, J_n) \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ with n judges, *collective rationality* requires that a resolute rule $f(\boldsymbol{J}) \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ returns a collective judgement that is Γ_{out}-consistent.

The first condition (`UD`) prohibits that an aggregation rule is defined on the domain of multisets of rational judgements; the requirement cannot be translated to our

[2] Note that in this representation, a profile $\boldsymbol{J}$ for n judges actually is an element of $\mathcal{J}(\Phi, \Gamma_{\text{in}})^n$. (While, in our framework, the expression $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ is slightly incorrect, see Section 2.1.)

framework. Anonymity (A) requires that all judges are treated in an equal manner. The third requirement, systematicity (S), demands that (i) the (collective) acceptance of a formula φ solely depends on (the pattern of) individual judgements for that formula φ and (ii) the patterns of individual judgements on a formula φ, that result in collective acceptance of φ, may not depend on the formula φ itself. We now have the tools to (formally) state *the* fundamental impossibility result of judgement aggregation.

Theorem 3.1 (List and Pettit, 2002)**.** *For an agenda Φ that contains a subset of issues of the form $\{\varphi_1, \varphi_2, \varphi_1 \vee \varphi_2\}$, with φ_1 and φ_2 being logically independent; there exists no resolute aggregation rule f satisfying (UD), (A), (S) and (CR).*

Since the seminal work of Arrow (1951), much research has been done to investigate what rules (aggregation rules and otherwise) satisfy what axioms. We can distinguish two types of results: characterisations and impossibility theorems.

Given a group of axioms, *impossibility theorems*—such as the result of List and Pettit (2002)—establish that it is impossible to (simultaneously) fulfil all given axioms. Impossibilities are negative (but relevant) results. On the other hand, *characterisation theorems* characterise a class of judgement aggregation rules by the given axioms; a rule satisfies the axioms if and only if it belongs to the demarcated class. That is, given a set of axioms, an impossibility theorem tells us that we can never find a procedure that fulfils all our wishes, which may help us to find acceptable compromises to make the demands satisfiable; a characterisation narrows down our choice (if we want to adhere to the axioms).

3.2 Characterisation of the Kemeny Rule

Nehring and Pivato (2022) have characterised the Kemeny judgement aggregation rule. They proved that a judgement aggregation rule F satisfies the axioms reinforcement, continuity and ensemble supermajority efficiency if and only if F is the Kemeny rule. Further, they show that judgement consistency and continuity together imply reinforcement—allowing for an alternative characterisation. For each axiom, we prove that the Kemeny rule satisfies it, and examine the adherence of the (parameterised) Kemeny-rule with the axiom. We show that the Kemeny-Nash rule satisfies none of the axioms, while the parameterised variant satisfies continuity, reinforcement and judgement consistency. To start, we provide a formal statements of the results.

Theorem 3.2 (Nehring and Pivato, 2022)**.** *Rule F satisfies continuity (C), reinforcement (R) and ensemble supermajority efficiency (ESME) if and only if F is the Kemeny rule.*

As mentioned, the authors show that (JC) and (C) together imply (R); this allows for an alternative characterisation.

Corollary 3.2.1. *Rule F satisfies continuity (`C`), judgement consistency (`JC`) and ensemble supermajority efficiency (`ESME`) if and only if F is the Kemeny rule.*

In the remainder of this section we study the different axioms ((`C`), (`R`), (`JC`) and (`ESME`)) one-by-one.

3.2.1 Continuity

The first axiom, *continuity*, aims to formalise a stability requirement. A rule satisfies the axiom if a collective judgement cannot be revoked by a small disruption in the associated profile.

Continuity (`C`): For an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ with n judges and any rational judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$, an aggregation rule F satisfies *continuity* if there exists an integer $\ell \in \mathbb{N}$ such that any judgement $J \in F(\boldsymbol{J})$ that is returned for profile $\boldsymbol{J}$, is also returned $J \in F(\ell\boldsymbol{J} \uplus \{J_{\text{in}}^1\})$ for profile $\ell\boldsymbol{J} \uplus \{J_{\text{in}}^1\}$.

In the definition above the profile $\{J_{\text{in}}^1\}$ represents an infinitesimal disruption to the profile $\ell\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{\ell n}$; as ℓ increases the single judge (with judgement J_{in}) is outnumbered by the total number of judges. The principle requires that it is possible to diminish the impact of the single judge to the point that no collective judgement that was returned for profile $\boldsymbol{J}$, is not returned for profile $\ell\boldsymbol{J} \uplus \{J_{\text{in}}^1\}$. Formulated differently, the axiom demands that when a large population is mixed with a much smaller population, the collective judgement is essentially determined by the judgements in the large population.

With nonconstructive proof we show that the Kemeny rule satisfies continuity.

Proposition 3.1. *The Kemeny rule F_{kem} satisfies (`C`).*

Proof. Take an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ (for n judges) and $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$ a rational judgement. Let $\boldsymbol{J'} = \{J_{\text{in}}^1\}$ denote the profile, for one judge, containing judgement J_{in}. Suppose there exists a feasible judgement $J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that $S_{\text{kem}}(\boldsymbol{J} \uplus \boldsymbol{J'}, J') > S_{\text{kem}}(\boldsymbol{J} \uplus \boldsymbol{J'}, J)$, for some judgement $J \in F_{\text{kem}}(\boldsymbol{J})$ that is returned by the Kemeny rule for profile $\boldsymbol{J}$. Then, it must be that $S_{\text{kem}}(\boldsymbol{J}, J) > S_{\text{kem}}(\boldsymbol{J}, J')$. Now, for all freasible judgements $\tilde{J} \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ we have:

$$S_{\text{kem}}(\ell\boldsymbol{J} \uplus \boldsymbol{J'}, \tilde{J}) = S_{\text{kem}}(\boldsymbol{J} \uplus \boldsymbol{J'}, \tilde{J}) + S_{\text{kem}}\Big((\ell - 1)\boldsymbol{J}, \tilde{J}\Big) \tag{3.1}$$

So, with every copy of $\boldsymbol{J}$ we add, we close the finite gap between $S_{\text{kem}}(\boldsymbol{J} \uplus \boldsymbol{J'}, J')$ and $S_{\text{kem}}(\boldsymbol{J} \uplus \boldsymbol{J'}, J)$.

But then, for any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ and feasible judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$, there must exist an $\ell \in \mathbb{N}$ such that $S_{\text{kem}}(\ell\boldsymbol{J} \uplus \{J_{\text{in}}^1\}, J) > S_{\text{kem}}(\ell\boldsymbol{J} \uplus \{J_{\text{in}}^1\}, J')$. $\square$

Because of the zero-effect, the Kemeny-Nash rule does not satisfy (`C`); we use a counterexample to prove this.

Proposition 3.2. *The Kemeny-Nash rule F_{kn} does not satisfy (C).*

Proof. The proof uses a counterexample. The essential feature of the example is that judgement $J_{\text{in}} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$ is antipodal to a collective judgement $J \in F_{\text{kn}}(\boldsymbol{J})$ selected by the Kemeny-Nash rule, for profile $\boldsymbol{J}$.

Consider the scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ with agenda Φ based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2\}$ containing two propositional variables, the input and output constraint coincide $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \top$, and are trivially satisfied; the profile $\boldsymbol{J} = \{J_{11}^2\}$ contains two judgements. Clearly, the Kemeny-Nash rule selects $F_{\text{kn}}(\boldsymbol{J}) = J_{11}$ as collective judgement.

Now, consider the feasible judgement $J_{\text{in}} = J_{00} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})$; for an arbitrary positive integer $\ell \in \mathbb{N}$ we have $S_{\text{kn}}(\ell\boldsymbol{J} \uplus \{J_{\text{in}}^1\}, J_{11}) = 0 < \ell = S_{\text{kn}}(\ell\boldsymbol{J} \uplus \{J_{\text{in}}^1\}, J_{10})$. That is, $J_{11} \in F_{\text{kn}}(\boldsymbol{J})$ while for all $\ell \in \mathbb{N}$ we have $J_{11} \notin F_{\text{kn}}(\ell\boldsymbol{J} \uplus \{J_{\text{in}}^1\})$. □

Mirroring the argument of Proposition 3.1 we can show that the parameterised Kemeny-Nash rule F_{kn}^{λ} satisfies (C); in the proof below, we do not reiterate the line of thought from the aforementioned argument in full detail.

Proposition 3.3. *The parameterised Kemeny-Nash rule F_{kn}^{λ} satisfies (C).*

Proof. The proof is analogous to the argument employed in the proof of Proposition 3.1. The differences being: (i) for a profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ and feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, the parameterised Kemeny-Nash $S_{\text{kn}}^{\lambda}(\boldsymbol{J}, J, \lambda)$ scores are substituted for the Kemeny scores $S_{\text{kem}}(\boldsymbol{J}, J)$, and (2) the sum in Equation 3.1 is replaced by a product.

For an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, and $0 < \lambda \ll 1$, for the parameterised Kemeny-Nash score we have: $S_{\text{kn}}^{\lambda}(\boldsymbol{J}, J, \lambda) > 0$, by definition. The validity of the argument still holds. □

3.2.2 Reinforcement

The *reinforcement* axiom is a requirement on the collective outcomes, returned for a compound profile. The axiom demands that if the constituent profiles share a collective judgement, then, for the compound profile, rule F returns exactly the judgements that are shared by the constituents.

Reinforcement (R): For arbitrary profiles $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ and $\boldsymbol{J}' \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{n'}$ with n and n' judges (respectively), an aggregation rule F satisfies *reinforcement* if $F(\boldsymbol{J}) \cap F(\boldsymbol{J}') \neq \emptyset$ implies that $F(\boldsymbol{J} \uplus \boldsymbol{J}') = F(\boldsymbol{J}) \cap F(\boldsymbol{J}')$.

Intuitively, it is clear that the Kemeny rule satisfies (R): the Kemeny score of a compound profile is maximal when the Kemeny score of its components are maximal. For the sake of completeness we provide a formal proof.

Proposition 3.4. *The Kemeny rule F_{kem} satisfies (R).*

Proof. Take arbitrary profiles $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ and $\boldsymbol{J'} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{n'}$, for n and n' judges, and let $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ be any feasible judgement. By the definition of the Kemeny rule, for the compound profile $\boldsymbol{J} \uplus \boldsymbol{J'}$ we have:

$$F_{\text{kem}}(\boldsymbol{J} \uplus \boldsymbol{J'}) = \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} \left(S_{\text{kem}}(\boldsymbol{J}, J) + S_{\text{kem}}(\boldsymbol{J'}, J) \right)$$

If $F_{\text{kem}}(\boldsymbol{J}) \cap F_{\text{kem}}(\boldsymbol{J'}) \neq \emptyset$—i.e., there exists a feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that $J \in \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} S_{\text{kem}}(\boldsymbol{J}, J)$ and $J \in \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\operatorname{argmax}} S_{\text{kem}}(\boldsymbol{J'}, J)$—then the requirement is satisfied. □

With a counterexample, we show that the Kemeny-Nash rule does not satisfy (R).

Proposition 3.5. *The Kemeny-Nash rule F_{kn} does not satisfy (R).*

Proof. The proof uses a counterexample. The idea is that for every feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, the antipodal judgement $\overline{J}$ is contained in (at least) one of the constituents of the compound profile $\boldsymbol{J} \uplus \boldsymbol{J'}$.

Let agenda Φ be based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2\}$ containing two propositional variables, and let the input and output constraint $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \top$ coincide. Consider the profiles $\boldsymbol{J} = \{J^1_{00}, J^1_{01}, J^1_{10}, J^1_{11}\}$ and $\boldsymbol{J'} = \{J^1_{11}\}$. It is easy to verify that $F_{\text{kn}}(\boldsymbol{J}) = \{J_{00}, J_{01}, J_{10}, J_{11}\}$ and $F_{\text{kn}}(\boldsymbol{J'}) = J_{11}$. Thus, we have $F_{\text{kn}}(\boldsymbol{J}) \cap F_{\text{kn}}(\boldsymbol{J'}) = J_{11} \neq \emptyset$. However, $S_{\text{kn}}(\boldsymbol{J} \uplus \boldsymbol{J'}, J) = S_{\text{kn}}(\boldsymbol{J}, J) \cdot S_{\text{kn}}(\boldsymbol{J'}, J) = 0$, for every feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$. That is, $F_{\text{kn}}(\boldsymbol{J} \uplus \boldsymbol{J'}) \neq F_{\text{kn}}(\boldsymbol{J}) \cap F_{\text{kn}}(\boldsymbol{J'})$. □

With a direct argument we show that the parameterised Kemeny-Nash rule F^{λ}_{kn} satisfies (R).

Proposition 3.6. *The parameterised Kemeny-Nash rule F^{λ}_{kn} satisfies (R).*

Proof. For arbitrary profiles $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ and $\boldsymbol{J'} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{n'}$ with n and n' judges, and feasible judgements $J, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ we have:

$$S^{\lambda}_{\text{kn}}(\boldsymbol{J} \uplus \boldsymbol{J'}, J, \lambda) = S^{\lambda}_{\text{kn}}(\boldsymbol{J}, J, \lambda) \cdot S^{\lambda}_{\text{kn}}(\boldsymbol{J'}, J, \lambda), \quad \text{for all } 0 < \lambda \ll 1$$

For all $0 < \lambda \ll 1$, the parameterised Kemeny-Nash score $S^{\lambda}_{\text{kn}}(\boldsymbol{J}, J, \lambda) > 0$ is positive by definition. Consequently, in correspondence to the argument we presented in Proposition 3.4, if $F^{\lambda}_{\text{kn}}(\boldsymbol{J}) \cap F^{\lambda}_{\text{kn}}(\boldsymbol{J'}) \neq \emptyset$ then: $F^{\lambda}_{\text{kn}}(\boldsymbol{J} \uplus \boldsymbol{J'}) = F^{\lambda}_{\text{kn}}(\boldsymbol{J}) \cap F^{\lambda}_{\text{kn}}(\boldsymbol{J'})$. □

3.2.3 Judgement Consistency

Before considering the final axiom of Theorem 3.2 we make a detour and consider the *judgement consistency* axiom. Judgement consistency is another demand for the collective judgements that are returned for compound profiles. The formal definition is given below.

Judgement Consistency (JC): Take arbitrary profiles $\boldsymbol{J_1}, \boldsymbol{J_1'} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ and $\boldsymbol{J_2}, \boldsymbol{J_2'} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{n'}$, respectively for n and n' judges, and feasible judgements $J, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$. A judgement aggregation rule F satisfies *judgement consistency* if the conjunction of the following premises excludes the possibility that judgement $J' \in F(\boldsymbol{J}_1' \uplus \boldsymbol{J}_2')$ is returned for the compound $\boldsymbol{J}_1' \uplus \boldsymbol{J}_2'$. There are four conditions:

$$J' \in F(\boldsymbol{J_1} \uplus \boldsymbol{J_2}) \tag{3.2}$$
$$J \in F(\boldsymbol{J_1'} \uplus \boldsymbol{J_2}) \tag{3.3}$$
$$J' \notin F(\boldsymbol{J_1'} \uplus \boldsymbol{J_2}) \tag{3.4}$$
$$J \in F(\boldsymbol{J_1} \uplus \boldsymbol{J_2'}) \tag{3.5}$$

Rule F satisfies judgement consistency if the conjunction of the premises implies:

$$J' \notin F(\boldsymbol{J_1'} \uplus \boldsymbol{J_2'}) \tag{3.6}$$

Taken together, the former three premises express that the change from $\boldsymbol{J_1}$ to $\boldsymbol{J_1'}$, in the compound profile with $\boldsymbol{J_2}$, effects the rule to no longer select judgement J'. The latter premise indicates that in the compound of $\boldsymbol{J}_1$ and $\boldsymbol{J}_2'$, judgement J is weakly preferred over judgement J'. If all these conditions are (simultaneously) satisfied, the axiom requires that—in the compound with profile $\boldsymbol{J}_2'$—the shift from $\boldsymbol{J_1}$ to $\boldsymbol{J_1'}$ again excludes judgement J' from being returned as collective judgement. That is, in a compound profile, the effect of a change in one of its constituents should not depend on the other constituent.

We use the contrapositive to show that the Kemeny rule satisfies (JC)

Proposition 3.7. *The Kemeny rule F_{kem} satisfies (JC).*

Proof. We prove the contrapositive. We show that if the implication (Equation 3.6) is *not* true, then it cannot be the case that all premises (Equations 3.2-3.5) are true.

From the conjunction of Equations 3.2-3.4, and some algebra, we obtain:

$$S_{\text{kem}}(\boldsymbol{J_1}, J') - S_{\text{kem}}(\boldsymbol{J_1'}, J') > S_{\text{kem}}(\boldsymbol{J_1}, J) - S_{\text{kem}}(\boldsymbol{J_1'}, J)$$

Similarly, from the conjunction of Equation 3.5 and the negation of Equation 3.6, we derive:

$$S_{\text{kem}}(\boldsymbol{J_1}, J) - S_{\text{kem}}(\boldsymbol{J_1'}, J) \geq S_{\text{kem}}(\boldsymbol{J_1}, J') - S_{\text{kem}}(\boldsymbol{J_1'}, J')$$

The derived inequalities are contradictory. □

By providing a counterexample, we show the the Kemeny-Nash rule does not satisfy (JC).

Proposition 3.8. *The Kemeny-Nash rule does not satisfy (JC).*

Proof. Proof by counterexample. For the agenda Φ that is based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2\}$ containing two propositional variables and $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \top$ trivial constraints we consider four profiles; two profiles $\boldsymbol{J_1}, \boldsymbol{J_1'} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^2$ with two judges, and two profiles $\boldsymbol{J_2}, \boldsymbol{J_2'} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^3$ for tree judges. The profiles are defined as follows:

$$\boldsymbol{J_1} = \{J_{01}, J_{11}\} \qquad \boldsymbol{J_2} = \{J_{00}, J_{10}^2\}$$
$$\boldsymbol{J_1'} = \{J_{00}, J_{01}\} \qquad \boldsymbol{J_2'} = \{J_{00}, J_{10}, J_{11}\}$$

We consider feasible judgements $J, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, stipulated as $J = J_{00}$ and $J' = J_{11}$. It can be verified that the premises (Equations 3.2-3.5) are satisfied:

$$J' \in F_{\text{kn}}(\boldsymbol{J_1} \uplus \boldsymbol{J_2})$$
$$J \in F_{\text{kn}}(\boldsymbol{J_1'} \uplus \boldsymbol{J_2})$$
$$J' \notin F_{\text{kn}}(\boldsymbol{J_1'} \uplus \boldsymbol{J_2})$$
$$J \in F_{\text{kn}}(\boldsymbol{J_1} \uplus \boldsymbol{J_2'})$$

However, for profile $\boldsymbol{J_1'} \uplus \boldsymbol{J_2'}$, judgement $J' \in F_{\text{kn}}(\boldsymbol{J_1'} \uplus \boldsymbol{J_2'})$ is returned by the Kemeny-Nash rule. □

With an algebraic derivation we prove that the parameterised Kemeny-Nash rule satisfies (JC).

Proposition 3.9. *The parameterised Kemeny-Nash rule F_{kn}^{λ} satisfies (JC)*

Proof. Let $\boldsymbol{J}_1, \boldsymbol{J}_1' \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ and $\boldsymbol{J}_2, \boldsymbol{J}_2' \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^{n'}$ be arbitrary profiles (for n and n' judges, respectively) and let $J, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ be any feasible judgements. We show that if the profiles $\boldsymbol{J}_1, \boldsymbol{J}_1', \boldsymbol{J}_2, \boldsymbol{J}_2'$ and feasible judgements J, J' are such that all conditions of the (JC) axiom are met (Equations 3.2-3.5), we can derive the consequence (Equation 3.6).

From the first condition (Equation 3.2) we can conclude:

$$\frac{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_2, J', \lambda)}{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_2, J, \lambda)} \geq \frac{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_1, J, \lambda)}{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_1, J', \lambda)}$$

While, from the second and third conditions (Equations 3.3 and3.4) we can derive:

$$\frac{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_1', J, \lambda)}{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_1', J', \lambda)} > \frac{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_2, J', \lambda)}{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_2, J, \lambda)}$$

Combining the two expressions we obtain:

$$\frac{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_1', J, \lambda)}{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_1', J', \lambda)} > \frac{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_1, J, \lambda)}{S_{\text{kn}}^{\lambda}(\boldsymbol{J}_1, J', \lambda)}$$

Similarly, from the fourth condition (Equation 3.5) we get:

$$\frac{S^{\lambda}_{\text{kn}}(\boldsymbol{J}_1, J, \lambda)}{S^{\lambda}_{\text{kn}}(\boldsymbol{J}_1, J', \lambda)} \geq \frac{S^{\lambda}_{\text{kn}}(\boldsymbol{J}'_2, J', \lambda)}{S^{\lambda}_{\text{kn}}(\boldsymbol{J}'_2, J, \lambda)}$$

The conclusion follows:

$$S^{\lambda}_{\text{kn}}(\boldsymbol{J}'_1, J, \lambda) \cdot S^{\lambda}_{\text{kn}}(\boldsymbol{J}'_2, J, \lambda) > S^{\lambda}_{\text{kn}}(\boldsymbol{J}'_1, J', \lambda) \cdot S^{\lambda}_{\text{kn}}(\boldsymbol{J}'_2, J', \lambda)$$

That is: $J' \notin F^{\lambda}_{\text{kn}}(\boldsymbol{J}'_1 \uplus \boldsymbol{J}'_2)$.

For an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ and feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, the Kemeny-Nash score $S^{\lambda}_{\text{kn}}(\boldsymbol{J}, J, \lambda) > 0$, for all $0 < \lambda < 1$; guaranteeing the soundness of the above expressions. $\square$

3.2.4 Ensemble Supermajority Efficiency

The final axiom, ensemble supermajority efficiency (`ESME`) is connected to the majority preservation (`MP`) axiom. The latter axiom is more common in the literature, and we already encountered it in the previous section. Significantly, the connection between the two axioms is that (`ESME`) is logically stronger than (`MP`). As the Kemeny-Nash and parameterised Kemeny-Nash rules do not satisfy (`MP`) (Propositions 3.11 and 3.12), the rules do not satisfy (`ESME`) either; for this reason we do not examine the axiom to the last detail.

Let $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ be an arbitrary profile for n judges, and $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ any feasible judgement. For $0 \leq n' \leq n$ we define the *n'-majority* $\gamma^{\boldsymbol{J}}$ as the number of issues that is supported in at least n' judgements $J_{\text{in}} \in \boldsymbol{J}$, that are contained in the profile:

$$\gamma^{\boldsymbol{J}}(J, n') = |\{\varphi \in J \mid |N^{\boldsymbol{J}}_{\varphi}| \geq n'\}|$$

Further we say that feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ is supermajority efficient (for profile $\boldsymbol{J}$) if there is no judgement $J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that $\gamma^{\boldsymbol{J}}(J', n') \geq \gamma^{\boldsymbol{J}}(J, n')$ for all $0 \leq n' \leq n$, with strict inequality for some $n' \leq n$. Let $\text{SME}(\boldsymbol{J})$ denote the set of all supermajority efficient judgements $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$; then a judgement aggregation rule is said to be supermajority efficient if it always returns a set of supermajority efficient judgements.

Supermajority Efficiency (`SME`): For an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, rule F is *supermajority effecient* if all collective judgements $J \in F(\boldsymbol{J}) \subseteq \text{SME}(\boldsymbol{J})$ are supermajority efficient.

Note that if profile $\boldsymbol{J}$ is majority-consistent, then $\text{SME}(\boldsymbol{J}) = m(\boldsymbol{J})$. Thus, if a judgement aggregation rule F does not satisfy (`MP`), then F does not satisfy (`SME`) either. That is, the Kemeny-Nash and parameterised Kemeny-Nash rule do not satisfy (`SME`).

Nehring and Pivato (2022) provide a formal definition for an *ensemble* of judgement aggregation scenarios. The ESME requirement is a restriction on the collective judgements that are returned for such ensembles. Specifically, for an arbitrary ensemble—which may contain any judgement aggregation scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ as (one of) its constituents—the axiom requires that a rule F is SME for (i) all the constituent of the ensemble, as well as for (2) the ensemble, considered as a judgement aggregation scenario on its own, as a whole. From what we said above, it is clear that condition (i) is not met—regardless of the precise definition of an ensemble of judgement aggregation scenarios—we establish that neither the Kemeny-Nash rule, nor the parameterised Kemeny-Nash rule, satisfies (`ESME`)

3.3 Axiomatic Properties of the Kemeny-Nash Rule

To further study the axiomatic properties of the (parameterised) Kemeny-Nash rule we examine four more axioms. We consider two axioms that are extensively appearing in the judgement aggregation literature; *majority preservation* and a version of *neutrality*. As we hypothesised (Section 2.3) that the (parameterised) Kemeny-Nash rule—in contrast to the Kemeny rule—produces collective outcomes that are sensitive to equity considerations, we also inspect two equity principles; the *Sen-Hammond equity* and *Pigou-Dalton* principles, both of which are frequently studied in the context of egalitarian welfare economics. In the judgement aggregation literature, both equity principles are studied by Botan et al. (2023).

Majority preservation. The first axiom we consider is *majority preservation*. It requires that rule F maps any majority-consistent profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, with $m(\boldsymbol{J}) \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, to the singleton set $F(\boldsymbol{J}) = \{m(\boldsymbol{J})\}$ that comprises the majoritarian judgement.

Majority Preservation (`MP`): For any majority-consistent profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, with $m(\boldsymbol{J}) \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, *majority preservation* demands a rule F to (exclusively) return $F(\boldsymbol{J}) = m(\boldsymbol{J})$ the majoritarian judgement.

Traditionally, majority preservation is considered to be an essential requirement. Judgement aggregation was initiated by the problems that arise when the majoritarian judgement is inconsistent—not to take issue with majority voting itself. When the majoritarian judgement is consistent, it is widely accepted that a judgement aggregation rule should return that, and only that, judgement.[3] Because of the ZE, the (parameterised) Kemeny-Nash rule does not comply with majority preservation. This can be seen as a serious deficiency.

We provide a direct argument to prove that the Kemeny rule F_{kem} satisfies (`MP`).

Proposition 3.10. *The Kemeny rule F_{kem} satisfies (`MP`).*

[3] This position is disputed by Botan et al. (2021).

Proof. Let $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, with $m(\boldsymbol{J}) \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, be an arbitrary majority-consistent profile with n judges. For an arbitrary feasible judgement $J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ and issue $\varphi \in \Phi$, the Kemeny score for profile $\boldsymbol{J}$ is given by $S_{\text{kem}}(\boldsymbol{J}, J) = \sum_{\varphi \in J} |N_{\varphi}^{\boldsymbol{J}}|$. By definition, for every $\varphi \in m(\boldsymbol{J})$, we have $|N_{\varphi}^{\boldsymbol{J}}| > |N_{\sim\varphi}^{\boldsymbol{J}}|$—and the majoritarian judgement $m(\boldsymbol{J})$ maximises the Kemeny score. □

Because of the ZE, the Kemeny-Nash rule does not comply with (`MP`); we provide a counterexample below.

Proposition 3.11. *The Kemeny-Nash rule F_{kn} does not satisfy (`MP`).*

Proof. Consider the scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ with agenda Φ based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2\}$ containing two propositional variables, trivial $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \top$ (input and output) constraints, and profile $\boldsymbol{J} = \{J_{00}^2, J_{11}^1\}$. Clearly, the majoritarian judgement $m(\boldsymbol{J}) = J_{00} \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ is feasible. However, $S_{\text{kn}}(\boldsymbol{J}, J_{01}) > S_{\text{kn}}(\boldsymbol{J}, m(\boldsymbol{J}))$; profile $\boldsymbol{J}$ witnesses that the Kemeny-Nash rule F_{kn} violates (`MP`). □

For the parameterised Kemeny-Nash rule the proof is more intricate; whether a scenario witnesses that F_{kn}^{λ} does not satisfy (`MP`) depends on the value of λ. In the proof below we introduce a family of scenarios $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})[m]$, with $m \in \mathbb{N}$ an arbitrary integer, such that (`MP`) is violated if $\lambda \leq {}^{(m-1)^2}/_{m^2}$. Put differently, to satisfy (`MP`), we must have $\lambda > {}^{(m-1)^2}/_{m^2}$, clearly violating the constraint $\lambda \ll 1$.

Proposition 3.12. *The parameterised Kemeny-Nash rule F_{kn}^{λ} does not satisfy (`MP`).*

Proof. Given any integer $m \in \mathbb{N}$, scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})[m]$ is defined as follows. The agenda Φ is based on the pre-agenda $\Phi^+ = \{\varphi_1, \ldots, \varphi_m\}$ containing m propositional variables, the constraints $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \top$ are trivially satisfied, and the profile $\boldsymbol{J} = \{J_-^2, J_+^1\}$ contains judgements $J_- = \{\neg\varphi_1, \ldots, \neg\varphi_m\}$ and $J_+ = \{\varphi_1, \ldots, \varphi_m\}$. Clearly, profile $\boldsymbol{J}$ is majority-consistent, and the majoritarian judgement is given by $m(\boldsymbol{J}) = J_- \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$.

Now, for judgement $J_1 = \{\varphi_1, \neg\varphi_2, \ldots, \neg\varphi_m\} \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ the parameterised Kemeny-Nash score is given by $S_{\text{kn}}^{\lambda}(\boldsymbol{J}, J_1, \lambda) = (m-1)^2$, for all $0 < \lambda \ll 1$. For the majoritarian judgement $m(\boldsymbol{J}) \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ we have: $S_{\text{kn}}^{\lambda}(\boldsymbol{J}, m(\boldsymbol{J}), \lambda) = m^2 \cdot \lambda$. Ergo, for an arbitrary integer $m \in \mathbb{N}$ and some scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})[m]$, to satisfy (`MP`), we must have $\lambda > {}^{(m-1)^2}/_{m^2}$—this is not allowed under the constraint $\lambda \ll 1$. □

Neutrality. Neutrality is an axiom that is widely accepted, but in many papers a precise definition is omitted (Slavkovik, 2014).[4] Broadly, the axiom says that an aggregation rule should treat the issues in the agenda in an equitable manner. A version called *issue-neutrality* has been formalised for resolute rules f by Grandi and Endriss (2013) in the following way. For any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ with n

[4] As an example, see Lang et al. (2011).

judges and $\varphi, \vartheta \in \Phi$ any two issues, if $N^{\boldsymbol{J}}_{\varphi} = N^{\boldsymbol{J}}_{\vartheta}$ then issue-neutrality requires that $\varphi \in f(\boldsymbol{J})$ if and only if $\vartheta \in f(\boldsymbol{J})$.

Issue-neutrality can be generalised for irresolute rules in multiple ways; Slavkovik (2014) formalised the strongest possible variant which we adopt here. Their axiom, *strong issue-neutrality*, requires that for any profile in which (any) two issues have identical support, any collective judgement that is returned by F should either include both issues, or none of them.

Strong Issue-Neutrality (SN): Irresolute rule F satisfies *strong issue-neutrality* if for all profiles $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ and all $\varphi, \vartheta \in \Phi$ issues it holds that: $N^{\boldsymbol{J}}_{\varphi} = N^{\boldsymbol{J}}_{\vartheta}$ implies $\varphi \in F(\boldsymbol{J})$ if and only if $\vartheta \in F(\boldsymbol{J})$, for all $J \in F(\boldsymbol{J})$.

By considering again Example 2.13, which we used to illustrate the mechanisms of the different rules, we show that none of the rules satisfies (SN). That is, all proofs are based on the same counterexample.

Proposition 3.13. *The Kemeny rule F_{kem} does not satisfy (SN).*

Proof. Consider again the scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ that was introduced in Example 2.13. For profile $\boldsymbol{J}$ we have: $N^{\boldsymbol{J}}_{\varphi_1} = N^{\boldsymbol{J}}_{\varphi_2} = \{J^1_1\} = \{\varphi_1, \varphi_2, \neg\varphi_3, \neg\varphi_4\}^1$, while judgement $J = \{\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4\}^1 \in F_{\text{kem}}(\boldsymbol{J})$ is returned by the Kemeny rule. $\square$

Proposition 3.14. *The Kemeny-Nash rule F_{kn} does not satisfy (SN).*

Proof. Consider the scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ from Example 2.13; for this profile $\boldsymbol{J}$ we have: $N^{\boldsymbol{J}}_{\varphi_1} = N^{\boldsymbol{J}}_{\varphi_2}$. However, judgement $J = \{\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4\}^1 \in F_{\text{kn}}(\boldsymbol{J})$ is returned by the Kemeny-Nash rule. $\square$

Proposition 3.15. *The parameterised Kemeny-Nash rule F^{λ}_{kn} does not satisfy (SN).*

Proof. Consider the scenario from Example 2.13. Again, for this profile $\boldsymbol{J}$ we have: $N^{\boldsymbol{J}}_{\varphi_1} = N^{\boldsymbol{J}}_{\varphi_2} = \{J^1_1\}$. However, judgement $J = \{\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4\}^1 \in F^{\lambda}_{\text{kn}}(\boldsymbol{J})$ is returned by the parameterised Kemeny-Nash rule, for any value $0 < \lambda \ll 1$ of the parameter. $\square$

The Sen-Hammond Equity Principle. The *Sen-Hammond Equity Principle* was originally formulated in the field of welfare economics. In essence, the principle requires that if there is a pair i, j of agents and feasible allocations X, Y—such that (i) the satisfaction of agent i is strictly higher than that of agent j (for both allocations X and Y) and (ii) agent j strictly prefers allocation Y over allocation X—then, the collective welfare of allocation Y should be (weakly) higher than the welfare of allocation X.

Botan et al. (2023) have adapted the principle for judgement aggregation. We give the formal definition of their principle below.

Sen-Hammond Equity (SHE): Let $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ be an arbitrary profile for n judges, and let J, J' be complete and complement-free judgements. A judgement aggregation rule F satisfies the *Sen-Hammond equity principle* if the following is true. Whenever there exist judges $i, j \in N$ such that:

$$\text{Agr}(J_i, J) > \text{Agr}(J_i, J') > \text{Agr}(J_j, J') > \text{Agr}(J_j, J) \tag{3.7}$$

And, further, for all other judges $i' \in N \setminus \{i, j\}$ it holds:

$$\text{Agr}(J_{i'}, J) = \text{Agr}(J_{i'}, J') \tag{3.8}$$

Then:

$$J \in F(\boldsymbol{J}) \qquad \text{implies} \qquad J' \in F(\boldsymbol{J}) \tag{3.9}$$

That is, comparing complete and complement-free judgements J and J': if there are judges $i, j \in N$ such that (i) judge j is worse off than judge i, and (ii) judge j strictly prefers judgement J' over J, while (iii) the remaining judges $N \setminus \{i, j\}$ are indifferent; then, if judgement $J \in F(\boldsymbol{J})$ is returned, judgement $J' \in F(\boldsymbol{J})$ should be returned as well.

To begin, by use of a counterexample, we show that the Kemeny rule F_{kem} does not satisfy (SHE).

Proposition 3.16. *The Kemeny rule F_{kem} does not satisfy (SHE).*

Proof. We consider a counterexample. Consider the scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ with agenda Φ based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6\}$ containing six propositional variables; the (input and output) constraints coincide and are trivially true, $\Gamma_{\text{in}} = \Gamma_{\text{out}} = \top$. The profile $\boldsymbol{J} = \{J_1^1, J_2^1\}$ with two judges, contains judgements $J_1 = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6\}$ and $J_2 = \{\neg\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4, \neg\varphi_5, \varphi_6\}$. Now, consider $J, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ feasible judgements $J = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6\}$, and $J' = \{\varphi_1, \varphi_2, \varphi_3, \neg\varphi_4, \neg\varphi_5, \neg\varphi_6\}$.

The premises of the Sen-Hammond principle (Equations 3.7 and 3.8) are satisfied: $\text{Agr}(J_1, J) > \text{Agr}(J_1, J') > \text{Agr}(J_2, J') > \text{Agr}(J_2, J)$, and $N \setminus \{1, 2\} = \emptyset$. Thus, we are done if we show that judgement $J \in F_{\text{kem}}(\boldsymbol{J})$ is selected while judgement $J' \notin F_{\text{kem}}(\boldsymbol{J})$ is not selected. As the Kemeny rule F_{kem} selects exactly the feasible judgements that accept the sixth propositional variable—i.e., for any $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, we have $J^* \in F_{\text{kem}}(\boldsymbol{J})$ if and only if $\varphi_6 \in J^*$—this is indeed the case. □

By use of a counterexample we show that the Kemeny-Nash rule F_{kn} does not satisfy (SHE).

Proposition 3.17. *The Kemeny-Nash rule F_{kn} does not satisfy (SHE).*

Proof. We provide a counterexample. Consider the scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ with the agenda Φ based on the pre-agenda $\Phi^+ = \{\varphi_1, \varphi_2, \varphi_3 \varphi_4, \varphi_5 \varphi_6, \varphi_7\}$ containing seven propositional variables. The input and output constraint $\Gamma_{\text{in}} = \Gamma_{\text{out}} = c_1 \wedge c_2 \wedge c_3$ coincide, and are defined as the conjunction of the following clauses:

$$
\begin{aligned}
c_1 &= (\varphi_1 \vee \varphi_2 \vee \varphi_3) \vee (\neg\varphi_1 \wedge \neg\varphi_2 \wedge \neg\varphi_3 \wedge \neg\varphi_4 \wedge \neg\varphi_5)\\
c_2 &= \neg\varphi_1 \vee \varphi_4 \vee \neg\varphi_6\\
c_3 &= \neg\varphi_1 \vee \varphi_5 \vee \neg\varphi_6
\end{aligned}
$$

The profile $\boldsymbol{J} = \{J_1^1, J_2^1\}$ contains two judgements:

$$
\begin{aligned}
J_1 &= \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7\}\\
J_2 &= \{\neg\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4, \neg\varphi_5, \varphi_6, \varphi_7\}
\end{aligned}
$$

In addition, we define the feasible judgements $J, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ as follows:

$$
\begin{aligned}
J &= J_1\\
J' &= \{\varphi_1, \varphi_2, \varphi_3 \neg\varphi_4, \neg\varphi_5 \neg\varphi_6, \varphi_7\}
\end{aligned}
$$

Now, we have: $\text{Agr}(J_1, J) > \text{Agr}(J_1, J') > \text{Agr}(J_2, J') > \text{Agr}(J_2, J)$ and profile $\boldsymbol{J}$ satisfies the conditions for (SHE). However, it can be verified that $F(\boldsymbol{J}) = \{J_1, J_2\}$.[5] That is, judgement $J \in F(\boldsymbol{J})$ is selected by the Kemeny-Nash rule while judgement $J' \notin F(\boldsymbol{J})$ is not, which is prohibited by the (SHE) requirement. □

We use the foregoing counterexample again, to show the parameterised Kemeny-Nash rule F_{kn}^{λ} does not satisfy (SHE).

Proposition 3.18. *The parameterised Kemeny-Nash judgement aggregation rule F_{kn}^{λ} does not satisfy (SHE).*

Proof. We use the scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ that is defined in the proof of Proposition 3.17. We established that given profile $\boldsymbol{J}$, (SHE) demands that if judgement $J \in F(\boldsymbol{J})$ is selected by rule F, then the rule should return judgement $J' \in F(\boldsymbol{J})$ as well. We found that the Kemeny-Nash rule F_{kn} does not meet this demand; the Kemeny-Nash score is maximised by judgements J_1 and J_2, with $S_{\text{kn}}(\boldsymbol{J}, J_1) = S_{\text{kn}}(\boldsymbol{J}, J_2) = 14$. We now show that introducing the parameter

[5] We omit the derivation of the collective judgements. They can be obtained by iterating through the possible Kemeny-Nash scores $S_{\text{kn}}(\boldsymbol{J}, J^*)$ in the following way. We start by asking whether there is a feasible judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that $\text{Agr}(J_i, J^*) = 7$, for some judge $i \in \{1, 2\}$. As the judgements J_1 and J_2 differ on five issues; the other agent, denote it as J_{-i}, has $\text{Agr}(J_{-i}, J^*) \in \{0, 1, 2\}$. We observe that both judgements J_1 and J_2 have a Kemeny-Nash score of $7 \cdot 2 = 14$ (which is maximal under the constraint that $\text{Agr}(J_i, J^*) = 7$ for some judge $i \in \{1, 2\}$). Subsequently, we look for a judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that (i) $\text{Agr}(J_i, J^*) = 6$ for some judge $i \in \{1, 2\}$, and (ii) $S_{\text{kn}}(\boldsymbol{J}, J^*) \geq 14$. Using the same steps as before, we conclude that such a judgement J^* does not exist. After all possible configurations of individual agreements are exhausted, we conclude that $F(\boldsymbol{J}) = \{J_1, J_2\}$.

$0 < \lambda \ll 1$ does not effect the set of collective judgements; i.e., we show that $F_{\text{kn}}(\boldsymbol{J}) = F^{\lambda}_{\text{kn}}(\boldsymbol{J})$, for all $0 < \lambda \ll 1$.

For all judgements $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, with $J^* \notin \{\overline{J_1}, \overline{J_2}\}$, that are not antipodal to any of the judgements contained in profile $\boldsymbol{J}$ we have: $S_{\text{kem}}(\boldsymbol{J}, J^*) = S^{\lambda}_{\text{kn}}(\boldsymbol{J}, J^*, \lambda)$, for all $0 < \lambda \ll 1$. Moreover, for a judgement $\tilde{J} \in \{\overline{J_1}, \overline{J_2}\}$ that is antipodal to a judgement in the profile we have: $S^{\lambda}_{\text{kn}}(\boldsymbol{J}, \tilde{J}, \lambda) \leq 7\lambda < 14$, for all $0 < \lambda \ll 1$. $\square$

Pigou-Dalton principle. The *Pigou-Dalton principle* (`PD`) was originally formulated in the context of welfare economics, it is studied by (a.o.) Dubey (2016) and Hara et al. (2008). In welfare economics, (`PD`) is a requirement on the possibility of inequality reducing trades. In particular, it demands that the collective welfare of an allocation a—in which an inequality-reducing transfer between two agents is possible—is weakly lower than that of allocation a' (obtained after the transfer).

Below we consider the formulation by Botan et al. (2023), which is adapted to the framework of judgement aggregation.

Pigou-Dalton principle (`PD`): Let $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ be an arbitrary profile with n judges, including judges $i, j \in N$, and consider a pair J, J' of complete and complement-free judgements. The principle restricts the behaviour of a rule F on any profile $\boldsymbol{J}$ that satisfies the following conditions:

$$\text{Agr}(J_i, J) > \text{Agr}(J_i, J') \geq \text{Agr}(J_j, J') > \text{Agr}(J_j, J) \tag{3.10}$$

$$\text{Agr}(J_i, J) + \text{Agr}(J_j, J) = \text{Agr}(J_i, J') + \text{Agr}(J_j, J') \tag{3.11}$$

$$\text{Agr}(J_{i'}, J) = \text{Agr}(J_{i'}, J') \quad \text{for all } i' \in N \setminus \{i, j\} \tag{3.12}$$

For any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ that satisfies these conditions, the Pigou-Dalton principle requires that $J \in F(\boldsymbol{J})$ implies $J' \in F(\boldsymbol{J})$.

Clearly, the Kemeny rule F_{kem} satisfies (`PD`); it is required that the Kemeny score of judgement J (before the transfer) is identical to the score of judgement J' (after the transfer). For the sake of completeness, we provide a formal proof.

Proposition 3.19. *The Kemeny rule F_{kem} satisfies (`PD`).*

Proof. Let $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$ be a profile with n judges, including judges i and j. If Equations 3.10-3.12 are satisfied for feasible judgements $J, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, we have:

$$\begin{aligned} S_{\text{kem}}(\boldsymbol{J}, J) &= S_{\text{kem}}(\boldsymbol{J}, J') + \text{Agr}(J_i, J) + \text{Agr}(J_j, J) - \text{Agr}(J_i, J') - \text{Agr}(J_j, J') \\ &= S_{\text{kem}}(\boldsymbol{J}, J') \end{aligned}$$

That is, if judgement $J \in F_{\text{kem}}(\boldsymbol{J})$ is returned by the Kemeny rule, then judgement $J' \in F_{\text{kem}}(\boldsymbol{J})$ is also returned. $\square$

By using a proof by contradiction we show that the Kemeny-Nash rule F_{kn} satisfies (`PD`).

Proposition 3.20. *The Kemeny-Nash rule F_{kn} satisfies (*`PD`*).*

Proof. Given a profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\mathrm{in}})^n$ with n judges, including judges $i, j \in N$, and feasible judgements $J, J' \in \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$, such that Equations 3.10-3.12 are fulfilled, we derive $S_{\mathrm{kn}}(\boldsymbol{J}, J') \geq S_{\mathrm{kn}}(\boldsymbol{J}, J)$.

For the sake of contradiction, assume that $S_{\mathrm{kn}}(\boldsymbol{J}, J) > S_{\mathrm{kn}}(\boldsymbol{J}, J')$. Using Equation 3.12 we obtain:

$$\mathrm{Agr}(J_i, J) \cdot \mathrm{Agr}(J_j, J) > \mathrm{Agr}(J_i, J') \cdot \mathrm{Agr}(J_j, J')$$

Which we can rewrite as:

$$\log \mathrm{Agr}(J_i, J) + \log \mathrm{Agr}(J_j, J) > \log \mathrm{Agr}(J_i, J') + \log \mathrm{Agr}(J_j, J')$$

Let $T = \mathrm{Agr}(J_i, J) + \mathrm{Agr}(J_j, J)$. We define $dA = \mathrm{Agr}(J_i, J) - \frac{1}{2}T > 0$, and $dA' = \mathrm{Agr}(J_i, J') - \frac{1}{2}T > 0$; by Equation 3.10 we have $dA > dA'$. Using Equation 3.10 we obtain:

$$\log\left(\frac{1}{2}T + dA\right) + \log\left(\frac{1}{2}T - dA\right) > \log\left(\frac{1}{2}T + dA'\right) + \log\left(\frac{1}{2}T - dA'\right)$$

From which we derive:

$$\log(-(dA)^2) > \log(-(dA')^2)$$

As $dA > dA'$, the above expression is a contradiction. We conclude that $S_{\mathrm{kn}}(\boldsymbol{J}, J') \geq S_{\mathrm{kn}}(\boldsymbol{J}, J)$; i.e., if judgement $J \in F_{\mathrm{kn}}(\boldsymbol{J})$ is returned by the Kemeny-Nash rule, then judgement $J' \in F_{\mathrm{kn}}(\boldsymbol{J})$. □

We show that under the restriction that $0 < \lambda \leq \frac{1}{2}$, the parameterised Kemeny-Nash rule satisfies (`PD`). As λ should be a (very) small non-zero parameter, we are satisfied with this result, and do not consider the cases that $\lambda > \frac{1}{2}$.

Proposition 3.21. *Under the restriction of $0 < \lambda \leq \frac{1}{2}$, the parameterised Kemeny-Nash rule $F_{\mathrm{kn}}^{\lambda}$ satisfies (*`PD`*).*

Proof. We show that the proposition is true for any scenario $(\Phi, \Gamma_{\mathrm{in}}, \Gamma_{\mathrm{out}}, \boldsymbol{J})$ in which the agenda is based on the pre-agenda that contains m propositional variables, the input and output constraint are trivially true (i.e., $\Gamma_{\mathrm{in}} = \Gamma_{\mathrm{out}} = \top$, and the profile contains two antipodal judgements. From the argument it is clear that allowing other agendas, constraints, or profiles does not effect the validity of the argument.

Take arbitrary $m \in \mathbb{N}$. Consider the scenario $(\Phi, \Gamma_{\mathrm{in}}, \Gamma_{\mathrm{out}}, \boldsymbol{J})$ in which the agenda Φ is based on the pre-agenda $\Phi^+ = \{\varphi_1, \ldots, \varphi_m\}$ containing m propositional variables; the input and output constraint $\Gamma_{\mathrm{in}} = \Gamma_{\mathrm{out}} = \top$ are trivially satisfied. Further let the profile $\boldsymbol{J} = \{J_{\mathrm{in}}^1, \overline{J_{\mathrm{in}}}^1$ contain two judgements that are antipodal.

If (PD) is violated there must be judgements $J, J' \subseteq \phi$ such that:

$$\mathrm{Agr}(overline J_{\mathrm{in}}, J) > \mathrm{Agr}(overline J_{\mathrm{in}}, J') \geq \mathrm{Agr}(J_{\mathrm{in}}, J') > \mathrm{Agr}(J_{\mathrm{in}}, J)$$

Because the Kemeny-Nash rule satisfies (PD) (Proposition 3.20), we must have $\mathrm{Agr}(J_{\mathrm{in}}, J) = 0$; thus, $J = overline J_{\mathrm{in}}$. Let us try to construct the judgement J' that satisfies the conditions to violate (PD). To minimise the parameterised Kemeny-Nash score $S^{\lambda}_{\mathrm{kn}}(\boldsymbol{J}, J')$ we minimise the product $\mathrm{Agr}(overline J_{\mathrm{in}}, J') \cdot \mathrm{Agr}(J_{\mathrm{in}}, J')$; i.e., $\mathrm{Agr}(J_i, J') = \mathrm{Agr}(J_i, J) - 1$. Any judgement J' that is obtained from judgement $overline J_{\mathrm{in}}$ by 'flipping' one accepted judgement (i.e., any judgement with $|J' \cap overline J_{\mathrm{in}}| = m - 1$) satisfies this condition.

Now, for the parameterised Kemeny-Nash scores we have:

$$S^{\lambda}_{\mathrm{kn}}(\boldsymbol{J}, J) = m \cdot \lambda$$
$$S^{\lambda}_{\mathrm{kn}}(\boldsymbol{J}, J') = m - 1$$

That is, for $\lambda > \frac{m-1}{m}$—we have $J \in F^{\lambda}_{\mathrm{kn}}(\boldsymbol{J})$, while $J' \notin F^{\lambda}_{\mathrm{kn}}(\boldsymbol{J})$—we cannot construct a judgement J' such that(PD) is violated.

The reader can verify that introducing non-trivial constraints—or allow more than two judgements in the profile, or taking non-antipodal judgements that are the relevant judgements in the profile (there may only be two judgements for which the agreements differ)—does not effect the validity of the proof □

3.4 Concluding Remarks

In this chapter we studied the axiomatic properties of the (parameterised) Kemeny-Nash rule.

Nehring and Pivato (2022) showed that the Kemeny rule is characterised by the following set of axioms: {(C), (R),(ESME)}. They further showed that (JC) and (C) together imply (R), allowing an alternative characterisation. We considered the four different axioms one-by-one. We showed the the Kemeny-Nash rule satisfies none of the axioms, while the parameterised variant satisfies (C), (R) and (JC). The results indicate that the axiomatic properties of the parameterised Kemeny-Nash rule are significantly better than those of the Kemeny-Nash rule.

Further, we saw that neither variants of the Kemeny-Nash rule satisfied (MP), an axiom that is conventionally seen as a fundamental requirement. We studied two fairness axioms that are widely used in the literature ((SHE) and (PD)), on the basis of these axioms we cannot distinguish the Kemeny rule from the Kemeny-Nash rule (or the parameterised Kemeny-Nash rule); this is a negative result.

Chapter 4

Theoretical Analysis II: Computational Complexity

After the seminal work of Arrow (1951) social choice theory was mainly concerned with studying the axiomatic properties of procedures (both existing and newly introduced) for different kinds of collective decision problems. Initially researchers neglected whether the procedures they studied (or introduced) were feasible in practice; they overlooked that the computational effort that was needed to use the procedures was often restrictive, and sometimes even prohibitive (Brandt et al., 2016). This changed over the course of the last fifty years, in which social choice theorists became increasingly aware of the computational resources needed to apply a procedure in practice. Around the turn of the century this shift culminated in the emergence of a new research area: *computational social choice*.[1]

In computational social choice, an essential aspect of a collective decision problem procedure, is how computationally challenging it is to determine the outcome—given an instance of the collective decision problem—also referred to as the *outcome determination problem.* More precisely, for a particular aggregation rule F, the outcome determination problem is to compute a collective judgement $J \in F(\boldsymbol{J})$, given an input scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$. In this chapter we study the outcome determination problem for the Kemeny, the Kemeny-Nash and the parameterised Kemeny-Nash rule.

In Section 4.1 we elaborate on the formal definition of the outcome determination problem in judgement aggregation. Section 4.2 introduces basic machinery from computational complexity theory that we use to obtain our results. Finally, in Section 4.3, we present our results.

We assume the reader has some acquaintance with the very basic notions of complexity theory; e.g., the complexity classes P and NP, a deterministic Turing machine as a model for computation and a polynomial-time (many-one) reduction. For more information, we recommend the textbook by Arora and Barak (2009).

[1]The term 'computational social choice', as a means to delineate a particular research area, was first used in 2006 (Brandt et al., 2016).

4.1 Outcome Determination in Judgement Aggregation

In this section we take a closer look at the outcome determination problem in judgement aggregation. In particular, we define two decision problems (i.e., problems that have a yes or no answer) that encompass an upper and lower bound on the complexity of the outcome determination problem.

The content of this section heavily relies on the work of Endriss et al. (2020). In particular, Endriss et al. (2020) studied the different formalisations of the outcome determination problem that have been studied in the literature. From the corresponding decision problems, they singled out the most general variant, as well as the most restricted variant. The former problem is used to establish an upper bound, while the latter is used to derive a lower bound on the complexity of the outcome determination problem. The decision problems with which we conclude this section, are the ones that were singled out by the authors.

For a fixed aggregation rule F, the outcome determination problem is (informally) defined as: 'Given an input scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$, also referred to as an *instance*, compute a collective judgement $J \in F(\boldsymbol{J})$.'[2] Clearly, this problem cannot be answered with 'yes' or 'no'; on the face of it, it is not a decision problem.

However, for the purpose of studying the complexity, we may formulate a decision problem that can be used to solve the original problem, if we were allowed to make multiple queries. Then, the complexity of the original problem is upper bounded by the complexity of querying the formulated decision problem multiple times (as many times as is necessary to solve the original problem).

Example 4.1. Suppose we have a scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ with $|\Phi^+| = m$ issues in the pre-agenda. To compute a collective judgement, we could iterate over all complete and complement-free judgements J and query: 'Is judgement $J \in F(\boldsymbol{J})$ a collective judgement?' With at most 2^m queries we would find (or rather, stumble upon) a valid solution. $\triangle$

The approach above works, but is uninformed; the large number of queries (exponential in the size of the input instance) leads to an upper bound that is not tight. Our aim then, is to formulate a more informed decision problem that, in a similar way as above, can be used to solve the original problem. Moreover, to describe the complexity of a particular problem, it is customary to derive both an upper and a lower bound on the complexity; for the lower bound, we introduce a second decision problem.

We denote the problem involving the upper bound as $\textsc{OutDet}(F)$. Important is that, given any judgement aggregation scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$, we can construct a collective judgement $J \in F(\boldsymbol{J})$ by solving $\textsc{OutDet}(F)$ a polynomial number of times.[3] As we indicated in Chapter 2, the judgement aggregation framework we have

[2] This problem has been formalised in various ways; for the different formal definitions we refer to Endriss et al. (2020).

[3] Polynomial in the size of the input scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$.

been using, is the most general framework that is known in the literature; problem $\textsc{OutDet}(F)$ should work for any scenario that can be described in this framework.

The problem that is used to derive a lower bound is denoted as $\textsc{OutDet}^*(F)$. For this problem, we require that there is no scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ for which we can construct a collective judgement $J \in F(\boldsymbol{J})$ by solving $\textsc{OutDet}^*(F)$ a polynomial number of times.[4] In order to establish a lower bound on the complexity of the outcome determination problem we should use the most restricted judgement aggregation framework, rather than the most general one.

The most restricted framework that appears in the literature is studied by (a.o.) Grandi (2012) and Grandi and Endriss (2013). In this framework the agenda $\Phi \subseteq \mathcal{L}$ only contains literals; further, there is a single constraint (i.e., $\Gamma_{\text{in}} = \Gamma_{\text{out}}$) that only contains variables that appear in the agenda Φ. The input space of $\textsc{OutDet}^*(F)$ consists of all scenarios that meet these restrictions.

The minimal conditions we outlined above do not determine the decision problems directly; in the judgement aggregation literature, multiple variants have been proposed. Endriss et al. (2020) have singled out the formulations that yield solutions that are universally valid. That is, for the lower bound $\textsc{OutDet}^*(F)$, they identified the most restricted variant of the outcome determination problem, in the judgement aggregation literature. Conversely, for the upper bound $\textsc{OutDet}(F)$, the most general decision problem was specified. We provide the formal definitions below.

$\textsc{OutDet}(F)$
Instance: An agenda Φ, constraints Γ_{in} and Γ_{out}, a profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, and subsets $L, L_1, \ldots, L_u \subseteq \Phi$ of the agenda, for $u \geq 0$.
Question: Is there a judgement set $J^* \in F(\boldsymbol{J})$ such that $L \subseteq J^*$ and $L_i \nsubseteq J^*$ for each $i \in \{1, \ldots, u\}$?

$\textsc{OutDet}^*(F)$
Instance: An agenda Φ, constraints Γ_{in} and Γ_{out}, a profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, and a formula $\varphi^* \in \Phi$ from the agenda.
Question: Is there a judgement set $J^* \in F(\boldsymbol{J})$ such that $\varphi^* \in J^*$?

Note that $\textsc{OutDet}(F)$ can, indeed, be used to solve the original problem of computing a collective judgement. Given any scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ we construct a collective judgement $J^* \in F(\boldsymbol{J})$ as follows. We fix an arbitrary ordering $\varphi_1, \ldots, \varphi_m \in \Phi^+$ of the issues in the pre-agenda. Starting from scratch, for the first issue φ_1, we compute whether there is a collective judgement $J^* \in F(\boldsymbol{J})$ such that $\{\varphi_1\} \subseteq J^*$ (i.e., we set $L = \{\varphi_1\}$ and $u = 0$). If the answer is 'yes', our next step is to compute whether there is a collective judgement $J^* \in F(\boldsymbol{J})$ such

[4]If we allow the existence of a scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J})$ that can be solved with $\textsc{OutDet}^*(F)$, by making a polynomial number of queries, then we would have to establish that there is no other scenario $(\Phi', \Gamma'_{\text{in}}, \Gamma'_{\text{out}}, \boldsymbol{J}')$ for which it is strictly easier to compute a collective judgement $J \in F(\boldsymbol{J}')$.

that $\{\varphi_1, \varphi_2\} \subseteq J^*$; otherwise, if 'no', we continue by evaluating whether there is a collective judgement $J^* \in F(\boldsymbol{J})$ such that $\{\sim\varphi_1\} \subseteq J^*$. Continuing this way, we either arrive at a complete collective judgement $J^* \in F(\boldsymbol{J})$, or decide that such a judgement does not exist, in at most $2m = |\Phi|$ queries. That is, we can solve the original problem, by solving the decision problem $\textsc{OutDet}(F)$ a polynomial number of times; the latter problem correctly encompasses the complexity of the former problem.

Considering the problem for the lower bound, $\textsc{OutDet}^*(F)$, we note that it cannot be used to construct a collective judgement. That is, the problem indeed encompasses a lower bound on the complexity of the outcome determination problem. On the other hand, it can be recognised that $\textsc{OutDet}^*(F)$ is a special case of $\textsc{OutDet}(F)$. Although $\textsc{OutDet}^*(F)$ cannot be used to solve the original problem entirely, it is reasonable to suggest that it captures an important aspect of the sought complexity, which increases the likelihood that we arrive at a tight lower bound.

4.2 Background

Computational complexity theory is the formal study of the resources required to solve computational problems. In this section we describe the computational complexity theory tools that we use to obtain our results (which are presented in the next section). We start with the definitions of two auxiliary notions, those of a truth assignment and of an alphabet. To continue, we recapitulate fundamental notions in complexity theory, e.g., complexity classes, completeness and polynomial-time many-one reductions. We assume the reader to have seen these concepts before. Finally, we describe relevant complexity classes, and corresponding complete problems.

Let Σ denote a finite set of symbols, or *alphabet*. A string (or *instance*) over alphabet Σ is a concatenation of alphabet symbols. We use Σ^n and Σ^* to denote the set of all strings with length n and all strings of finite length, respectively. A formal language L over an alphabet Σ is a subset $L \subseteq \Sigma^*$ of strings. In the previous section we described decision problems informally as problems that are answered with 'yes' or 'no'. Using formal notation, a decision problem is defined as a language $L \subseteq \Sigma^*$, over some alphabet Σ, specifying the positive instances.

For a propositional formula φ, we denote the set of all variables that appear in the formula as $\text{var}(\varphi)$. By a slight abuse of notation, for a set of propositional formulas S, we use $\text{var}(S)$ to denote the set of all variables $\text{var}(S) = \bigcup_{\varphi \in S} \text{var}(\varphi)$, that appear in any of the formulas. A (partial) *truth assignment* $\alpha : \text{var}(\varphi) \to \{0, 1\}$ maps variables to truth values; $\alpha(\varphi)$ is 0 if φ is false, and 1 if φ is true. With $\varphi[\alpha]$ we denote the formula that is obtained after instantiating the variables that are assigned by α with their assigned truth value $\alpha(\varphi)$. Again, by a slight abuse of notation, for any assignment $\alpha : \text{var} \to \{0, 1\}$ that is complete—i.e., all variables in $\text{var}(\varphi)$ are designated by α—we let $\varphi[\alpha]$ denote the truth value of formula φ.

In computational complexity we distinguish different complexity classes. A *complexity class* C is a collection of computational problems that are relatable in terms of complexity, for some bounded resource (such as time or memory space). We only consider complexity classes of decision problems. Typical examples are the classes P and NP, which contain the decision problems that can be decided and verified by a polynomial-time deterministic Turing machine, respectively.

A problem P is said to be *hard* for class C if, in some mathematical well-defined way, there is no problem Q in the class C that is harder than problem P. For decision problems hardness is established via *many-one reductions*. A many-one reduction maps instances of some decision problem L_1 to instances of another decision problem L_2, such that the reduced instance is in language L_2 if and only if the original instance is in language L_1. Of particular importance are reductions that are polynomial-time computable.

Formally, a *polynomial-time many-one reduction* from problem L_1 to problem L_2 is a polynomial-time computable function $f : \Sigma^* \to \Sigma^*$ such that $x \in L_1$ if and only if $f(x) \in L_2$. If problem L_1 is (polynomial-time) *reducible* to problem L_2—denoted as $L_1 \leq_p L_2$—having an efficient algorithm to decide language L_2 implies that we can at least efficiently decide all instances of language L_1 (and possibly other instances as well). In other words, we establish that problem L_2 is at least as hard as problem L_1. All reductions we describe are polynomial-time many-one reductions (and hereafter we often refer to them simply as reductions).

A problem is *complete* for class C if it is both contained in class C, and hard for class C. The satisfiability problem SAT, asking whether a Boolean formula is satisfiable, is a classic example of an NP-complete problem.

SAT
Instance: A Boolean formula φ in conjunctive normal form.
Question: Is there a truth assignment that satisfies φ?

Regarding theoretic results, we distinguish *membership results* from *hardness results*. A membership result, stating that a particular problem P is in a class C, establishes an upper bound on the complexity of a problem. On the other hand, hardness results, stating that a particular problem P is hard for class C, entail a lower bound on the complexity of problem P.

Besides the classes P and NP, two classes that will be relevant are Θ_2^p and Δ_2^p. The class Δ_2^p contains all decision problems that can be decided by a polynomial-time deterministic Turing machine that has access to an NP oracle—the oracle can solve any instance of the particular NP problem in a single time step. The class Θ_2^p is a subset of Δ_2^p, for which the number of oracle queries is restricted to be logarithmic in the size of the input. That is, Θ_2^p contains the problems that can be solved by a polynomial-time deterministic Turing machine that can query an NP oracle a logarithmic number of times. Accordingly, the class Θ_2^p may also be denoted as $\mathrm{P}^{\mathrm{NP}[\log]}$, while Δ_2^p can be denoted as P^{NP}.

The following problem is known to be complete for Θ_2^p (Endriss et al., 2020).[5]

MAX-MODEL **Instance:** A satisfiable propositional formula φ, and a variable $x^* \in \mathrm{var}(\varphi)$. **Question:** Is there a model α of φ that sets x^* to true, such that there is no other model of φ that sets more variables in $\mathrm{var}(\varphi)$ to true than α?

To conclude this section, we give another problem that is used in the derivation of our results.

AGR-K **Instance:** An agenda Φ, constraints Γ_{in} and Γ_{out}, a profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, and an integer $k \in \mathbb{N}$. **Question:** Is there a feasible judgement set $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that $\sum_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \mathrm{Agr}(J_{\text{in}}, J^*)) \geq k$?

Clearly this problem is in NP; a certificate can be checked in polynomial time.

4.3 Results

Here we present our computational complexity results for the outcome determination problem for the Kemeny, for the Kemeny-Nash, and for the parameterised Kemeny-Nash rule. For the Kemeny rule, we recapitulate the proof by Endriss et al. (2020), showing that the outcome determination problem is Θ_2^p-complete. For both the Kemeny-Nash and parameterised Kemeny-Nash rule we establish that the outcome determination problem is in Δ_2^p and hard for Θ_2^p. That is, for the Kemeny-Nash and the parameterised Kemeny-Nash rule, the derived lower and upper bound on the complexity of the outcome determination problem do not coincide, and we have no guarantee that the established bounds are tight.

4.3.1 Outcome Determination for the Kemeny Rule

To start, we reiterate the proof by Endriss et al. (2020), and show that the outcome determination problem for the Kemeny rule is in Θ_2^p.

[5] In the literature MAX-MODEL is commonly formulated slightly different, accepting any propositional formula φ as valid input (see, e.g., Krentel (1988), Wagner (1990), and Chen and Toda (1995)). With a simple reduction we can show that the problem, as we define it here, is not easier than its more common formulation; which guarantees that our formulation is Θ_2^p-complete.

Let MAX-MODEL* denote the common formulation of the problem, which differs from our formulation only in that it accepts any propositional formula φ as input. We show that MAX-MODEL* is polynomial-time reducible to MAX-MODEL. For any propositional formula φ, we define φ' as the disjunction of φ and the conjunction of the negations of all variables that appear in φ. Thus, we define: $\varphi' = \varphi \vee \bigwedge_{x \in \mathrm{var}(\varphi)} \neg x$. As φ' is satisfiable by construction, (φ', x^*) is a well-formed instance for MAX-MODEL. Now, any truth assignment that sets at least one variable to true is a model for φ' if and only if it is a model for φ; the reduction works.

Theorem 4.1 (Endriss et al., 2020). *The outcome determination problem for the Kemeny rule F_{kem} is in Θ_2^p.*

Proof. Using the general judgement aggregation framework (with arbitrary constraints), we show that $\textsc{OutDet}(F_{\text{kem}})$ is in Θ_2^p (see Section 4.1). We do this by describing a polynomial-time algorithm that has access to an NP oracle, and solves the problem with at most $\mathcal{O}(\log|\Phi^+| + \log n)$ queries.

By querying the oracle, the algorithm first checks if there exists a feasible judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that the Kemeny score $S_{\text{kem}}(\boldsymbol{J}, J) \geq k$ exceeds a given value k. This is an NP problem; by selecting an NP-complete problem for the oracle, we can solve any instance of the problem in a single time step. For an arbitrary feasible judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ and any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, the Kemeny score is upper bounded by $S_{\text{kem}}(\boldsymbol{J}, J^*) \leq n \cdot |\Phi^+|$. Using binary search we can determine the maximal Kemeny score $k_{\max}$ with $\mathcal{O}(\log|\Phi^+| + \log n)$ queries.

We now solve the original decision problem with one more query: 'Is there a judgement set $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that $S_{\text{kem}}(\boldsymbol{J}, J^*) = k_{\max}$, $L \subseteq J^*$, and $L_i \notin J^*$ for all $i \in \{1, \ldots, u\}$?' This problem is in NP, and any instance can be solved with a single oracle query. The answer to this final question is 'yes' if and only if such a feasible judgement J^* exists. The algorithm runs in polynomial time, queries the oracle at most $\mathcal{O}(\log|\Phi^+| + \log n)$ times, and correctly solves the problem. Thereby we establish that the problem is in Θ_2^p. □

To continue, we present the proof by Endriss et al. (2020), showing that the outcome determination problem for the Kemeny rule F_{kem} is hard for Θ_2^p.

Theorem 4.2 (Endriss et al.,2020). *The outcome determination problem for the Kemeny rule F_{kem} is hard for Θ_2^p.*

Proof. Using the restricted judgement aggregation framework—in which the input and output constraint coincide, and solely contain propositional variables that appear in the agenda—we show that $\textsc{OutDet}^*(F_{\text{kem}})$ is hard for Θ_2^p (see Section 4.1). We construct a polynomial-time reduction from $\textsc{Max-Model}$ (see Section 4.2) to $\textsc{OutDet}^*(F_{\text{kem}})$.

Let (ψ, x^*), with $\text{var}(\psi) = \{x_1, \ldots, x_n\}$, be any instance of $\textsc{Max-Model}$. Without loss of generality, we let $x^* = x_1$. We use the following reduction to obtain an instance $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ of the outcome determination problem $\textsc{OutDet}^*(F_{\text{kem}})$ for the Kemeny rule.

To begin, we introduce $n(n+1)$ propositional variables $z_{i,j}$, for all $1 \leq i \leq n$ and $1 \leq j \leq n+1$. We define the agenda Φ as follows:

$$\Phi = \{x_i, \neg x_i, z_{i,j}, \neg z_{i,j} \mid 1 \leq i \leq n, 1 \leq j \leq n+1\}$$

Next, we define the following input and output constraints:

$$\Gamma_{\text{in}} = \Gamma_{\text{out}} = \bigvee_{1 \leq i \leq n} \left(\bigwedge_{1 \leq j \leq n+1} z_{i,j} \right) \vee \left(\psi \wedge \bigwedge_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n+1}} \neg z_{i,j} \right)$$

$\boldsymbol{J}$	J_1	J_2	$\cdots$	J_{n-1}	J_n	$m(\boldsymbol{J})$
x_1	0	1	1	$\cdots$	1	1
x_2	1	0	1	$\cdots$	1	1
$\vdots$	$\vdots$		$\ddots$		$\vdots$	$\vdots$
x_{n-1}	1	$\cdots$	1	0	1	1
x_n	1	$\cdots$	1	1	0	1
$z_{1,1}$	1	0	0	$\cdots$	1	0
$z_{1,1}$	0	1	0	$\cdots$	1	0
$\vdots$	$\vdots$		$\ddots$		$\vdots$	$\vdots$
$z_{n-1,1}$	0	$\cdots$	0	1	0	0
$z_{n,1}$	0	$\cdots$	0	0	1	0
$\vdots$			$\vdots$			$\vdots$
$z_{1,n+1}$	1	0	0	$\cdots$	0	0
$z_{2,n+1}$	0	1	0	$\cdots$	0	0
$\vdots$	$\vdots$		$\ddots$		$\vdots$	$\vdots$
$z_{n-1,n+1}$	0	$\cdots$	0	1	0	0
$z_{n,n+1}$	0	$\cdots$	0	0	1	0

Figure 4.1: Construction of the profile $\boldsymbol{J}$ in the proof of Theorem 4.2. Figure taken from Endriss et al. (2020).

We define the profile containing $|\boldsymbol{J}| = n$ judgements as illustrated in Figure 4.1. Lastly, we let $\varphi^* = x_1$.

We explain the idea behind the reduction below. First, we note that for any instance $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ that is produced with the reduction, and any feasible judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, the Kemeny score $S_{\text{kem}}(\boldsymbol{J}, J^*)$ solely depends on the intersection $J^* \cap m(\boldsymbol{J})$. In particular, for any two issues $\varphi, \varphi' \in m(\boldsymbol{J})$ in the majoritarian judgement we have $|N^{\boldsymbol{J}}_{\varphi}| = |N^{\boldsymbol{J}}_{\varphi'}| = n - 1$, and we can express the Kemeny score as follows:

$$S_{\text{kem}}(\boldsymbol{J}, J^*) = |J^* \cap m(\boldsymbol{J})| \cdot (n-1) + |J^* \setminus m(\boldsymbol{J})|$$

Consequently, for feasible judgements $J^*, J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, we have $S_{\text{kem}}(\boldsymbol{J}, J^*) \geq S_{\text{kem}}(\boldsymbol{J}, J')$ if and only if $|J^* \cap m(\boldsymbol{J})| \geq |J' \cap m(\boldsymbol{J})|$. That is, the Kemeny rule selects the feasible judgements that maximise the intersection with the majoritarian judgement $m(\boldsymbol{J})$.

Now, the idea behind the reduction is as follows. Again, let (ψ, x^*), with $\text{var}(\psi) = \{x_1, \ldots, x_n\}$, be an arbitrary instance of MAX-MODEL, and let $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ denote the reduced instance. If the collective judgement does not satisfy ψ, then it must accept $n+1$ variables $\bigwedge_{1 \leq j \leq n+1} z_{i,j}$, for some $1 \leq i \leq n$. That is, any collective judgement that does not satisfy ψ must deviate from the majoritarian judgement $m(\boldsymbol{J})$ on at least $n+1$ issues. On the other hand, any feasible judgement that satisfies ψ can deviate from the majoritarian judgement $m(\boldsymbol{J})$ on at most n issues—the number of variables that appear in ψ.

Thus, for any feasible judgement that is consistent with ψ, the Kemeny score is strictly higher than it is for any judgement that is not consistent with ψ. Further, the more variables $x_i \in \text{var}(\psi)$ that are collectively accepted, the higher the Kemeny score. Intuitively it is clear that the collective judgements corresponds to a maximal models (i.e., maximal number of variables is set to true) for ψ.

To make this precise, we show that for any instance (ψ, x^*) of Max-Model, the instance is a positive instance if and only if the reduced instance $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ is a positive instance of $\text{OutDet}^*(F_{\text{kem}})$. Concretely, there exists a model $\alpha : \text{var}(\psi) \to \{0,1\}$ that sets (i) a maximal number of variables to true, and (ii) sets variable x_1 to true, if and only if there exists a collective judgement $J^* \in F_{\text{kem}}(\boldsymbol{J})$ that accepts issue $\varphi^* \in J^*$.

To begin, we prove that for any positive instance (ψ, x^*) of Max-Model, with $\text{var}(\psi) = \{x_1, \dots, x_n\}$ and $\varphi^* = x_1$, the reduced instance $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ is in $\text{OutDet}^*(F_{\text{kem}})$. Suppose there is a truth assignment $\alpha : \text{var}(\psi) \to \{0,1\}$ that (i) sets a maximal number of variables to true, and (ii) sets x_1 to true. Consider the following feasible judgement: $J^* = \{x_i \mid 1 \leq i \leq n, \alpha(x_i) = 1\} \cup \{\neg x_i \mid 1 \leq i \leq n, \alpha(x_i) = 0\} \cup \{\neg z_{i,j} \mid 1 \leq i \leq n, 1 \leq j \leq n+1\}$. We show that judgement $J^* \in F_{\text{kem}}(\boldsymbol{J})$ is a collective judgement.

For the sake of contradiction, suppose there exists a feasible judgement $J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that $|J' \cap m(\boldsymbol{J})| > |J^* \cap m(\boldsymbol{J})|$. Since judgement $J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ is feasible, we have either (i) $z_{i,j} \in J'$, for some $1 \leq i \leq n$, and for all $1 \leq j \leq n+1$, or (ii) $J' \cup \psi$ is satisfiable and $\neg z_{i,j} \in J'$, for all $1 \leq i \leq n$, and all $1 \leq j \leq n+1$. As ψ is satisfiable by definition, we can rule out the former option: if (i), judgement J' must deviate from $m(\boldsymbol{J})$ on at least $n+1$ issues, while (ii) implies that judgement J' deviates from $m(\boldsymbol{J})$ on at most n issues. Thus, $J' \cup \psi$ is satisfiable.

Now, consider the assignment $\alpha' : \text{var}(\psi) \to \{0,1\}$ that is defined as follows: $\alpha'(x_i) = 1$ if $x_i \in J'$, and $\alpha'(x_i) = 0$ otherwise, for all $1 \leq i \leq n$. Assignment α' satisfies ψ and sets more variables to true than assignment α; this contradicts our assumption that α sets a maximal number of variables to true.

We conclude that the existence of a model α—that sets a maximal number of variables in $\text{var}(\psi)$ to true and that sets x_1 to true—implies that there is a collective judgement $J^* \in F_{\text{kem}}(\boldsymbol{J})$ that accepts issue $\varphi^* \in J^*$.

We now show that for any instance $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$, if there exists a collective judgement $J^* \in F_{\text{kem}}(\boldsymbol{J})$ that accepts issue φ^*, then there is a model $\alpha : \text{var}(\psi) \to \{0,1\}$ that sets a maximal number of variables to true and sets x_1 to true. Assume that there exists a collective judgement $J^* \in F_{\text{kem}}(\boldsymbol{J})$ that accepts issue $\varphi^* \in J^*$. Examine the truth assignment $\alpha : \text{var}(\psi) \to \{0,1\}$ that is defined as: $\alpha(x_i) = 1$ if $x_i \in J^*$ and $\alpha(x_i) = 0$ otherwise, for all $1 \leq i \leq n$. Since judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ is feasible, it must be that either (i) $z_{i,j} \in J^*$, for some $1 \leq i \leq n$, and for all

$1 \leq j \leq n+1$, or (ii) $J^* \cup \psi$ is satisfiable and $\neg z_{i,j} \in J^*$, for all $1 \leq i \leq n$, and all $1 \leq j \leq n+1$. As ψ is satisfiable by assumption, the former option can be ruled out: if (i), judgement J^* must deviate from $m(\boldsymbol{J})$ on at least $n+1$ issues, while (ii) implies that judgement J^* deviates from $m(\boldsymbol{J})$ on at most n issues. Thus, we establish that α satisfies ψ. Further, because $x_1 = \varphi^* \in J^*$, we know that α sets x_1 to true. We demonstrate that there is no assignment $\alpha' : \mathrm{var}(\psi) \to \{0,1\}$ that sets more variables to true than α.

To get a contradiction, we assume that there exists a truth assignment $\alpha' : \mathrm{var}(\psi) \to \{0,1\}$ that sets more variables to true than α. Consider the following feasible judgement: $J' = \{x_i \mid 1 \leq i \leq n, \alpha'(x_i) = 1\} \cup \{\neg x_i \mid 1 \leq i \leq n, \alpha'(x_i) = 0\} \cup \{\neg z_{i,j} \mid 1 \leq i \leq n, 1 \leq j \leq n+1\}$. For feasible judgement $J' \in \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$ we have $|J' \cap m(\boldsymbol{J})| > |J^* \cap m(\boldsymbol{J})|$, which is contradictory to our assumption that $J^* \in F_{\mathrm{kem}}(\boldsymbol{J})$ is a collective judgement.

We conclude that α is a model for ψ that sets a maximal number of variables to true and that sets x_1 to true. $\square$

4.3.2 Outcome Determination for the Kemeny-Nash Rule

Continuing with the results for the Kemeny-Nash rule F_{kn}, we show that the outcome determination problem is in Δ_2^p.

Theorem 4.3. *The outcome determination problem for the Kemeny-Nash rule F_{kn} is in Δ_2^p.*

Proof. Using the general judgement aggregation framework (with arbitrary constraints), we show that $\textsc{OutDet}(F_{\mathrm{kn}})$ is in Δ_2^p (see Section 4.1). The proof resembles the proof of Theorem 4.1. We describe a polynomial-time algorithm that uses at most $\mathcal{O}(n \cdot \log |\Phi^+|)$ queries to an NP oracle, and that correctly solves the problem.

By querying the oracle, we first check if there exists a feasible judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$ such that the Kemeny-Nash score $S_{\mathrm{kn}}(\boldsymbol{J}, J) \geq k$ exceeds a given value k. This is an NP problem; by selecting an NP-complete problem for the oracle, we can solve any instance in a single time step. For an arbitrary feasible judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$ and any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\mathrm{in}})^n$, the Kemeny-Nash score is upper bounded by $S_{\mathrm{kn}}(\boldsymbol{J}, J^*) \leq |\Phi^+|^n$. Using binary search, we can determine the maximal Kemeny-Nash score $k_{\max}$ with $\mathcal{O}(n \cdot \log |\Phi^+|)$ queries.

We now solve the original problem with one more query: 'Is there a judgement set $J^* \in \mathcal{J}(\Phi, \Gamma_{\mathrm{out}})$ such that $S_{\mathrm{kn}}(\boldsymbol{J}, J^*) = k_{\max}$, $L \subseteq J^*$, and $L_i \notin J^*$ for all $i \in \{1, \ldots, u\}$?' This is an NP problem; any instance is solved with a single oracle query. The answer to this final question is 'yes' if and only if such a feasible judgement J^* exists. Thus, the algorithm runs in polynomial time, queries the oracle at most $\mathcal{O}(n \cdot \log |\Phi^+|)$ times, and correctly solves the problem. We conclude that the problem is in Δ_2^p. $\square$

Next, we show that the outcome determination problem for the Kemeny-Nash rule F_{kn} is hard for Θ_2^p.

Theorem 4.4. *The outcome determination problem for the Kemeny-Nash rule F_{kn} is hard for Θ_2^p.*

Proof. Using the restricted judgement aggregation framework—in which the input and output constraint coincide, and solely contain propositional variables that appear in the agenda—we show that $\textsc{OutDet}^*(F_{\text{kn}})$ is hard for Θ_2^p (see Section 4.1). We provide a polynomial-time reduction from $\textsc{Max-Model}$ (see Section 4.2) to $\textsc{OutDet}^*(F_{\text{kn}})$. The reduction is equivalent to the reduction that we used for the Kemeny rule, in the proof of Theorem 4.2. In particular, we show that for any instance $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ that is produced with the reduction, the outcomes of the Kemeny and the Kemeny-Nash rule coincide; i.e., $F_{\text{kem}}(\boldsymbol{J}) = F_{\text{kn}}(\boldsymbol{J})$.

Let $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ be any judgement aggregation scenario that is produced with the reduction from the proof of Theorem 4.2. We show that any feasible judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ is collectively accepted $J^* \in F_{\text{kem}}(\boldsymbol{J})$ by the Kemeny rule if and only if the judgement $J^* \in F_{\text{kn}}(\boldsymbol{J})$ is collectively accepted by the Kemeny-Nash rule.

For the forward direction, let judgement $J^* \in F_{\text{kem}}(\boldsymbol{J})$ be any collective judgement that is selected by the Kemeny rule, further let $J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ be an arbitrary feasible judgement. By the definition of the Kemeny rule we have:

$$\sum_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \text{Agr}(J_{\text{in}}, J^*)) \geq \sum_{J_{\text{in}} \in \boldsymbol{J}} \big(\boldsymbol{J}(J_{\text{in}}) \cdot \text{Agr}(J_{\text{in}}, J')\big)$$

That is, the sum of individual agreements is maximised by judgement J^*. Moreover, for any two judgements $J_{\text{in}}, J'_{\text{in}} \in \boldsymbol{J}$ contained in the profile, it holds that $|\text{Agr}(J_{\text{in}}, J^*) - \text{Agr}(J_{\text{in}}, J^*)| \leq 1$. Thus, even if we neglect the output constraint Γ_{out}, it is impossible to distribute the total agreement more equally over the individual judgements. But then, by Equation 2.4, judgement $J^* \in \underset{J \in \mathcal{J}(\Phi, \Gamma_{\text{out}})}{\text{argmax}} \prod_{J_{\text{in}} \in \boldsymbol{J}} S_{\text{kn}}(\boldsymbol{J}, J)$ is collectively accepted by the Kemeny-Nash rule.

For the other direction, assume that judgement $J^* \in F_{\text{kn}}(\boldsymbol{J})$ is collectively accepted by the Kemeny-Nash rule. Let $J' \in F_{\text{kem}}(\boldsymbol{J})$ be an arbitrary feasible judgement that is accepted by the Kemeny rule. For the sake of contradiction, assume that $\sum_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \text{Agr}(J_{\text{in}}, J^*)) < \sum_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \text{Agr}(J_{\text{in}}, J'))$. From the argument above it is clear that $\prod_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \text{Agr}(J_{\text{in}}, J')) > \prod_{J_{\text{in}} \in \boldsymbol{J}} (\boldsymbol{J}(J_{\text{in}}) \cdot \text{Agr}(J_{\text{in}}, J^*))$, contradicting our assumption that $J^* \in F_{\text{kn}}(\boldsymbol{J})$ is collectively accepted by the Kemeny-Nash rule.

We have shown that for any judgement aggregation scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ that is produced with the reduction, presented in the proof of Theorem 4.2, it holds that $F_{\text{kem}}(\boldsymbol{J}) = F_{\text{kn}}(\boldsymbol{J})$ the outcomes of the Kemeny and Kemeny-Nash rule coincide. From Theorem 4.2, it follows that the outcome determination problem $\textsc{OutDet}^*(F_{\text{kn}})$ for the Kemeny-Nash rule is hard for Θ_2^p. $\square$

4.3.3 Outcome Determination for the Parameterised Kemeny-Nash Rule

To conclude, we present the results for the complexity of the outcome determination problem in judgement aggregation for the parameterised Kemeny-Nash rule F_{kn}^{λ}. We start by showing that the problem is contained in Δ_2^p.

Theorem 4.5. *The outcome determination problem for the parameterised Kemeny-Nash rule F_{kn}^{λ} is in Δ_2^p.*

Proof. Using the general judgement aggregation framework (with arbitrary constraints), we show that $\textsc{OutDet}(F_{\text{kn}}^{\lambda})$ is in Δ_2^p (see Section 4.1). The proof takes after the proofs of Theorems 4.1 and 4.3. We describe a polynomial-time algorithm that uses at most $\mathcal{O}\left(n(\log|\Phi^+| - \log\lambda)\right)$ queries to an NP oracle, and that correctly solves the problem.

By querying the oracle, we check if there is a feasible judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that the parameterised Kemeny-Nash score $S_{\text{kn}}^{\lambda}(\boldsymbol{J}, J) \geq k$ exceeds a given value k. This is an NP problem; by selecting an NP-complete problem for the oracle, we can solve any instance in unit time. For an arbitrary feasible judgement $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ and any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma_{\text{in}})^n$, the parameterised Kemeny-Nash score is upper bounded by $S_{\text{kn}}^{\lambda}(\boldsymbol{J}, J^*) \leq |\Phi^+|^n$. If none of the judgements in profile $\boldsymbol{J}$ is antipodal to judgement J^*, i.e., if $\overline{J^*} \notin \boldsymbol{J}$, then the parameterised Kemeny-Nash score is an integer value. Otherwise, the score is divisible by λ^n. This gives a total of $(|\Phi^+|^n + 1) + \left(\lfloor\lambda^{-1}|\Phi^+|\rfloor^n\right)$ possible values for $k_{\max}$. Consequently, using binary search, the value of $k_{\max}$ can be determined with $\mathcal{O}\left(n(\log|\Phi^+| - \log\lambda)\right)$ oracle queries.

Again, we solve the original problem with one more query: 'Is there a judgement set $J^* \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$ such that $S_{\text{kn}}^{\lambda}(\boldsymbol{J}, J^*) = k_{\max}$, $L \subseteq J^*$, and $L_i \notin J^*$ for all $i \in \{1, \dots, u\}$?' This is an NP problem that can be solved with a single oracle query. The answer is 'yes' if and only if such a feasible judgement J^* exists. Thus, the algorithm runs in polynomial time, queries the oracle at most $\mathcal{O}\left(n(\log|\Phi^+| - \log\lambda)\right)$ times, and correctly solves the problem. Therefore the problem is in Δ_2^p. $\square$

Regarding the lower bound on the complexity of the outcome determination problem, we show that for the parameterised Kemeny-Nash rule this problem is hard for Θ_2^p.

Theorem 4.6. *The outcome determination problem for the parameterised Kemeny-Nash rule F_{kn}^{λ} is hard for Θ_2^p.*

Proof. Using the restricted judgement aggregation framework (input and output constraint coincide, and solely contain propositional variables that appear in the agenda), we show that $\textsc{OutDet}^*(F_{\text{kn}}^{\lambda})$ is hard for Θ_2^p (see Section 4.1). We provide a polynomial-time reduction from $\textsc{Max-Model}$ (see Section 4.2) to $\textsc{OutDet}^*(F_{\text{kn}}^{\lambda})$. The reduction we provide is the one that we already used for the Kemeny rule, in the proof of Theorem 4.2. As we argued above—in the proof of Theorem 4.4—for any instance that is produced with this reduction, the outcomes of the Kemeny rule coincide with the outcomes of the Kemeny-Nash rule. Here we show that, for any such instance, the outcomes of the Kemeny-Nash rule and the outcomes of the parameterised Kemeny-Nash coincide, too. Thereby we establish that the outcome determination problem for the parameterised Kemeny-Nash rule is hard for Θ_2^p.

Let $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ be any judgement aggregation scenario that can be produced with the reduction presented in the proof of Theorem 4.2; we show that $F_{\text{kn}}(\boldsymbol{J}) = F_{\text{kn}}^{\lambda}(\boldsymbol{J})$. First note that for any feasible judgement $J' \in \mathcal{J}(\Phi, \Gamma_{\text{out}})$, the parameterised Kemeny-Nash score $S_{\text{kn}}^{\lambda}(\boldsymbol{J}, J') \neq S_{\text{kn}}(\boldsymbol{J}, J')$ is effected by λ, if and only if the judgement $\overline{J'} \in \boldsymbol{J}$ is antipodal to one of the judgements in the profile.

Now, let $J_{\text{in}} \in \boldsymbol{J}$ be any judgement that is contained in the profile. It is easy to verify that for any judgement $J'_{\text{in}} \in \boldsymbol{J}$, contained in the profile, it holds that $\text{Agr}(J'_{\text{in}}, \overline{J_{\text{in}}}) \leq 2(n+2)$—equality holds for any $J'_{\text{in}} \neq J_{\text{in}}$. On the other hand, we know that for any judgement $J^* \in F_{\text{kn}}(\boldsymbol{J})$ that is selected by the Kemeny-Nash rule it holds that $\text{Agr}(J'_{\text{in}}, J^*) \geq (n-1)(n+1)$. Thus, independent of the value of $0 < \lambda \ll 1$, for all $n > 3$,[6] and for all judgements $J'_{\text{in}} \in \boldsymbol{J}$ contained in the profile, we have: $\text{Agr}(J'_{\text{in}}, J^*) > \text{Agr}(J'_{\text{in}}, \overline{J_{\text{in}}})$. That is, for $n > 3$, the outcomes of the Kemeny-Nash rule are not effected by introducing the parameter $0 < \lambda \ll 1$.

We demonstrated that for any judgement aggregation scenario $(\Phi, \Gamma_{\text{in}}, \Gamma_{\text{out}}, \boldsymbol{J}, \varphi^*)$ with $n > 3$, that can be constructed with the reduction in the proof of Theorem 4.2, it holds that $F_{\text{kn}}(\boldsymbol{J}) = F_{\text{kn}}^{\lambda}(\boldsymbol{J})$, the outcomes of the Kemeny-Nash and the parameterised Kemeny-Nash rule coincide. From Theorems 4.2 and 4.4, it follows that the outcome determination problem $\textsc{OutDet}^*(F_{\text{kn}}^{\lambda})$ for the parameterised Kemeny-Nash rule is hard for Θ_2^p. $\square$

[6] Computational complexity studies the asymptotic behaviour of computational problems; it is not a problem that the proof does not work for $n \leq 3$.

Chapter 5

Experimental Analysis

In the previous two chapters our approach was theoretical. In this chapter we study the empirical properties of the (parameterised) Kemeny-Nash rule. The experimental approach may help us to uncover general, albeit not universal, patterns in the outcomes of the Kemeny-Nash rule; in particular, in comparison to the outcomes of the Kemeny rule. Such patterns point to properties that are satisfied most of the time, but not always. Of course, as social choice theory procedures are meant to be applied in real-world scenarios, such properties are relevant.

Thus, with our experimental analysis we aim to establish general relations between the Kemeny and (parameterised) Kemeny-Nash rule. Further, we hope to get an idea of how these relations depend on the scenario under investigation. In a nutshell, the idea is to compare the outcomes of the Kemeny and (parameterised) Kemeny-Nash rule on a wide range of different judgement aggregation scenarios $(\Phi, \Gamma, \boldsymbol{J})$.[1]

The remainder of this chapter is structured as follows. In Section 5.1 we present the partial scenarios—consisting of an agenda Φ and constraint Γ—that have been investigated. Subsequently we give an overview of the relevant parts of our implementation (Section 5.2). We continue in Section 5.3 by describing (the theory behind) our evaluation procedure. Section 5.4 contains all information that would enable one to reproduce the results (at least in theory). We list relevant machine specifications, software requirements and specify the parameters that are used in the different experiments. In Section 5.5 we present the results, which are discussed in Section 5.6.

Our implementation is based on the `jaggpy` library,[2] and publicly available on `GitHub`.[3]

[1] For all investigated scenarios the input and output constraint coincide, so we do not distinguish an input and output constraint.

[2] `https://pypi.org/project/jaggpy/`

[3] `https://github.com/paulinebaanders/ThesisAI-JA`

5.1 Partial Scenarios

We used ten partial scenarios, consisting of an agenda Φ and a single constraint Γ. (In the rest of this chapter—when it is clear from the context that we consider a partial scenario—we may refer to it simply as a scenario.) For each partial scenario we specify the pre-agenda Φ^+, the constraint Γ and provide the following figures:

- $2^{|\Phi^+|}$: Total number of binary valuations over the agenda issues (here a valuation must be complete but may be inconsistent).
- $|\mathcal{J}(\Phi, \Gamma)|$: Number of consistent judgements. (As our partial scenarios only have a single constraint, this figure equals the number of rational judgements, and equals the number of consistent judgements.)
- $|\{J \mid J, \overline{J} \in \mathcal{J}(\Phi, \Gamma)\}|$: Number of *consistent-antipodal* judgements J. Judgement $J \in \mathcal{J}(\Phi, \Gamma)$ is consistent-antipodal if and only if its antipodal $\overline{J} \in \mathcal{J}(\Phi, \Gamma)$ is also consistent.

Further, for each scenario, we sketch a real-world situation in which the judgement aggregation scenario may be relevant.

Our first partial scenario corresponds to the situation we saw in the Introduction (Chapter 1), where a group of judges decides the liability of a defendant. The Juridical Verdict Contract (JVC) scenario can be summarised as follows:

Juridical Verdict Contract (JVC)	
Pre-agenda:	$\Phi^+ = \{\varphi_1, \varphi_2, \varphi_1 \wedge \varphi_2\}$
Constraint:	$\Gamma = \top$
# Total bin:	$2^{\|\Phi^+\|} = 8$
# Consistent:	$\|\mathcal{J}(\Phi, \Gamma)\| = 4$
# Antipodal:	$\|\{J \mid J, \overline{J} \in \mathcal{J}(\Phi, \Gamma)\}\| = 2$

Here φ_1 may stand for 'a valid contract was in place', φ_2 for 'there was a breach' and $\varphi_1 \wedge \varphi_2$ for 'the defendant is liable'.

Our next partial scenario corresponds to the (complete) scenario we examined in Example 2.13, which was also used to prove that none of our rules satisfy `(SN)` (Propositions 3.13-3.15). The Car Mechanics (CM) partial scenario is taken from Endriss et al. (2020), and may arise when a group of mechanics judge which elements of a car are broken. The scenario may be summarised as follows:

Car Mechanics (CM)	
Pre-agenda:	$\Phi^+ = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4\}$
Constraint:	$\Gamma = \neg\varphi_1 \vee (\neg\varphi_2 \wedge \neg\varphi_4) \vee \neg\varphi_3$
# Total bin:	$2^{\|\Phi^+\|} = 16$
# Consistent:	$\|\mathcal{J}(\Phi, \Gamma)\| = 13$
# Antipodal:	$\|\{J \mid J, \overline{J} \in \mathcal{J}(\Phi, \Gamma)\}\| = 10$

In this scenario the propositional formula φ_i may correspond to 'element i works'. Then, the constraint says that a consistent judgement must either identify element 1, element 3, or both element 2 and element 4 as broken.

The Self-Driving Car (SDC) scenario may occur in a self-driving car that is processing information coming from different sensors. It may be summarised as follows:

Self-Driving Car (SDC)	
Pre-agenda:	$\Phi^+ = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_1 \rightarrow (\varphi_3 \wedge \varphi_4), \varphi_2 \vee \varphi_3, \varphi_1 \wedge \varphi_2\}$
Constraint:	$\Gamma = \top$
# Total bin:	$2^{\lvert\Phi^+\rvert} = 128$
# Consistent:	$\lvert\mathcal{J}(\Phi, \Gamma)\rvert = 16$
# Antipodal:	$\lvert\{J \mid J, \overline{J} \in \mathcal{J}(\Phi, \Gamma)\}\rvert = 2$

Here the variable φ_1 may express 'swerve', φ_2 may mean 'high probability of accidental death', φ_3 may signify 'car damage' and φ_4 may mean 'driver in danger'. Then, the other formulas in the pre-agenda express statements that may be relevant for the car to compute whether it should swerve or not. For example, the formula $\varphi_1 \rightarrow (\varphi_3 \wedge \varphi_4)$ states 'swerve implies car damage and driver in danger'.

Our next partial scenario is taken from Lang et al. (2011). The Government Regulation (GR) scenario may be used by a government that aims to establish rules related to the COVID-19 disease. It can be represented as follows:

Government Regulation (GR)	
Pre-agenda:	$\Phi^+ = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_1 \rightarrow (\varphi_2 \vee \varphi_3), \varphi_1 \rightarrow (\varphi_4 \vee \varphi_5)\}$
Constraint:	$\Gamma = \top$
# Total bin:	$2^{\lvert\Phi^+\rvert} = 128$
# Consistent:	$\lvert\mathcal{J}(\Phi, \Gamma)\rvert = 32$
# Antipodal:	$\lvert\{J \mid J, \overline{J} \in \mathcal{J}(\Phi, \Gamma)\}\rvert = 2$

Here the literals φ_1, φ_2, φ_3, φ_4, and φ_5 may be interpreted as 'positive test', 'infected', 'apparatus faulty', 'verify test' and 'quarantine', respectively. Then, formula $\varphi_1 \rightarrow (\varphi_2 \vee \varphi_3)$ denotes that 'if positive test, then infected or apparatus faulty' (one could deny this statement by assuming that there is a human error involved). Finally, formula $\varphi_1 \rightarrow (\varphi_4 \vee \varphi_5)$ states that 'a positive test implies verification or quarantine'.

The following partial scenarios encode (strict) preference aggregation with three (PA3) and four (PA4) alternatives, respectively. In general, a preference aggregation scenario with m alternatives can be encoded in judgement aggregation as follows:

Preference Aggregation m (PAm)	
Pre-agenda:	$\Phi^+(m) = \{\varphi_{ij} \mid i,j \in \{1,\ldots,m\}, i<j\}$
Constraint:	$\bigwedge_{\substack{i,j,k\in\{1,\ldots,m\}\\ i<j<k}} ((\varphi_{ij} \vee \neg\varphi_{ik} \vee \varphi_{jk}) \wedge (\neg\varphi_{ij} \vee \neg\varphi_{jk} \vee \varphi_{ik}))$
# Total bin:	$2^{\lvert\Phi^+(m)\rvert} = 2^{m(m-1)/2}$
# Consistent:	$\lvert\mathcal{J}(\Phi,\Gamma,m)\rvert = m!$
# Antipodal:	$\lvert\{J \mid J, \overline{J} \in \mathcal{J}(\Phi,\Gamma,m)\}\rvert = m!$

In this representation a variable φ_{ij} could indicate that alternative i is strictly preferred over alternative j. To see that the number of consistent judgements equals $m!$, we have to realise that every permutation (of the m alternatives) corresponds to exactly one consistent judgement. The number of consistent-antipodal judgements equals the number of consistent judgements: for every permutation, also the permutation in reverse order corresponds to exactly one consistent judgement.

Finally we consider the pre-agenda that consists of m positive, logically independent, literals; i.e., $\Phi^+ = \{\varphi_1,\ldots,\varphi_m\}$. We consider pre-agendas with three (PL3), four (PL4), six (PL6) and seven (PL7) positive literals. We can represent the Positive Literals m (PLm) scenario, for arbitrary m, as follows:

Positive Literals m (PLm)	
Pre-agenda:	$\Phi^+ = \{\varphi_1,\ldots,\varphi_m\}$
Constraint:	$\Gamma = \top$
# Total bin:	$2^{\lvert\Phi^+\rvert} = 2^m$
# Consistent:	$\lvert\mathcal{J}(\Phi,\Gamma)\rvert = 2^m$
# Antipodal:	$\lvert\{J \mid J, \overline{J} \in \mathcal{J}(\Phi,\Gamma)\}\rvert = 2^m$

In this case φ_i, for any $i \in \{1,\ldots,m\}$, may correspond to any statement; the sole requirement is that φ_i is logically independent of φ_j, for all $j \in \{1,\ldots,m\} \setminus \{i\}$.

To conclude, in Table 5.1 we provide an overview of the partial scenarios (including relevant statistics) that might be helpful for reference.

5.2 Implementation

In this section we discuss the relevant parts of our implementation. The design of our code is to facilitate a comparison between (the returned collective judgements of) the Kemeny and Kemeny-Nash rule—on a wide range of different judgement aggregation scenarios. In the implementation we can distinguish four components. To start, we briefly outline these components.

The first component is based on the `jaggpy` library, it contains a *scenario object and solver methods*. Together they deal with all the formal requirements of judgement aggregation, including the translation from propositional logic to a formal computer language. Second, we have the encodings of the *aggregation rules*;

	$\lvert\Phi^+\rvert$	$\lvert\mathcal{J}(\Phi,\Gamma)\rvert$	$\frac{\lvert\mathcal{J}(\Phi,\Gamma)\rvert}{2^{\lvert\Phi^+\rvert}}$	$\frac{\lvert\{J \mid J,\overline{J} \in \mathcal{J}(\Phi,\Gamma)\}\rvert}{\lvert\mathcal{J}(\Phi,\Gamma)\rvert}$
Juridical Verdict Contract (JVC)	3	4	$1/2$	$1/2$
Car Mechanics (CM)	4	13	$13/16$	$10/13$
Self-Driving Car (SDC)	7	16	$1/8$	$1/8$
Government Regulation (GR)	7	32	$2/8$	$1/16$
Preference Aggregation 3 (PA3)	3	4	$3/4$	1
Preference Aggregation 4 (PA4)	6	24	$3/8$	1
Positive Literals 3 (PL3)	3	8	1	1
Positive Literals 4 (PL4)	4	16	1	1
Positive Literals 6 (PL6)	6	64	1	1
Positive Literals 7 (PL7)	7	128	1	1

Table 5.1: Overview of partial scenarios (characterised by an agenda Φ and constraint Γ). $|\mathcal{J}(\Phi,\Gamma)|$ symbolises the number of consistent judgements, while $|\Phi^+|$ indicates the number of issues in the pre-agenda. The fraction $|\mathcal{J}(\Phi,\Gamma)|/2^{|\Phi^+|}$ stands for the ratio between the total number of binary valuations of the pre-agenda issues (may be inconsistent) and the number of consistent judgements. Finally, $|\{J \mid J,\overline{J} \in \mathcal{J}(\Phi,\Gamma)\}|/|\mathcal{J}(\Phi,\Gamma)|$ represents the proportion between consistent-antipodal judgements (judgement J is consistent-antipodal if $J,\overline{J} \in \mathcal{J}(\Phi,\Gamma)$) and consistent judgements. (See Section 5.1 for more detailed information about the scenarios.)

these encodings are used by the solver methods to compute the outcome of a particular judgement aggregation scenario. The *profile iteration* component deals with the generation of (many) admissible profiles that are tested for a particular partial scenario. Finally, the fourth component consists of procedures we use for the *evaluation* of the collective judgements that are selected by the different rules.

In the remainder of this section we discuss the former three components: the scenario object and solver methods, the encodings of the aggregation rules and the procedure to iterate through profiles.[4] We conclude the section with a side-note on profile representation; in our implementation profiles are represented as multisets.

Scenario object and solver methods. This part is based on the `jaggpy` library. Minor modifications to the original code included fixing some bugs, and restructuring the code to better fit our purpose (of testing many different profiles corresponding to one partial scenario).

The user specifies a partial scenario—that is a pre-agenda Φ^+ and (input and output) constraints Γ_{in} and Γ_{out}, in a `.jagg` file. The propositional formulas that appear in the (pre-)agenda and constraints may contain logical connectives ($\vee, \wedge, \rightarrow, \neg$). The constraints may not include variables that do not occur in the (pre-)agenda.

With the `jaggpy` library we can construct two different solver objects: a Brute

[4]The evaluation procedures are interesting from a theoretical point of view, rather than from an operational point of view; we review the procedures in a dedicated section (Section 5.3).

Force (BF) solver, or an ASP solver. We did preliminary experiments to compare the time efficiency of the different methods. For our purpose, the BF solver was shown to be abundantly faster than the ASP solver. All experiments hereafter are performed with the BF solver. For the results of the preliminary experiments, and a more detailed explanation, see Section 5.5.5. Briefly, in large part, the efficiency of the BF solver can be ascribed to pre-computations that are executed once for every partial scenario; this time is negligible when, for every partial scenario, we test many (often hundreds of thousands) different profiles. The solver methods can be used to compute the set of collective judgements of the scenario, for various judgement aggregation rules.

Aggregation rules. Although the solver objects of the `jaggpy` library include implementations for some of our rules; in our implementation they were implemented anew, to better suit our modified `jaggpy` code. Like the original `jaggpy` library, our encodings of the rules depend on the solver method that is used; most of our rules are implemented twice (both for the BF solver and for the ASP solver).

For the BF solver we implemented the Kemeny and MaxHam anew, and added an implementation for the Kemeny-Nash, parameterised Kemeny-Nash and MaxEq rules.[5] In the ASP solver we implemented a new version of the Kemeny rule, and added implementations for the Kemeny-Nash, and parameterised Kemeny-Nash rules. For the ASP solver, the rules are implemented twice; once using the method of optimisation (ASP-OPT), where we use the ASP build-in aggregate `#maximize`. The other version (ASP-SAT) employs the method of saturation, which was introduced by Eiter and Gottlob (1995).

Profile iteration. For every partial scenario we want to investigate a large number of (admissible) profiles for n agents. Thus, given a partial scenario (Φ, Γ) and a number of judges n, we require a large multiset P, with $\mathcal{J}(\Phi, \Gamma)^n$ as underlying set.[6] Subsequently, we compute the collective judgements (returned by the different rules) for every profile $\boldsymbol{J} \in P$ one-by-one. That is, we *iterate* over the profiles in P. The iteration itself is self-explanatory; here we explain how we construct the multiset of profiles P.

If feasible—see Section 5.4 for the exact conditions—we iterate over all admissible profiles; i.e., we set $P = \mathcal{J}(\Phi, \Gamma)^n$. Otherwise, we randomly draw $|P|$ profiles from $\mathcal{J}(\Phi, \Gamma)^n$ (with replacement). Because we sample with replacement P may indeed contain duplicates. Although it is not entirely correct, for the sake of readability, we may write $P \subseteq \mathcal{J}(\Phi, \Gamma)^n$.

A weak point regarding our profile iteration procedure is that all profiles are ascribed equal probability. Especially when judgement aggregation is used to reach consensus about a factual matter—as opposed to choose an alternative based on individual preferences—this procedure is quite problematic. Often consensus should

[5] In Section 5.3 we define—and explain why we implemented—the MaxHam and MaxEq rule.

[6] As explained below, P may contain duplicates, hence it is a multiset (and not a set).

be reached among a group of experts—judges, mechanics, sensors in a self-driving car, professors judging the work of a student, etc.—it is not realistic that judgements in such a group are randomly distributed.

Side-note on profile representation. In our implementation profiles are represented as multisets. Although the tuple representation is more common in the judgement aggregation literature, we choose the multiset representation because this makes the total number of profiles much smaller (see Section 2.1). That is, by using this representation the number of profiles, contained in a full iteration is greatly diminished.

5.3 Evaluation

In this section we explain the metrics and procedures we have used to evaluate the computed collective judgements. The qualitative and quantitative analysis form the main part of our experimental procedure; they are discussed in Section 5.3.1 and Section 5.3.2, respectively. In both parts our goal is to classify the Kemeny-Nash rule, according to its fairness and efficiency properties, and determine how these properties are effected by the λ parameter. To do this, we compare the Kemeny-Nash rule to the Kemeny rule (an efficient procedure *pur sang*, see Section 2.3), and to two rules that count as downright fair (the MaxHam rule and the MaxEq rule, which we introduce below).

Section 5.3.3 is devoted to our analysis of the Variance-Increasing Zero-Effect (VIZE). When we introduced the VIZE, in Section 2.4, we suggested that it has a negative influence on the quality of the collective judgements of the Kemeny-Nash rule. We argued that introducing the λ parameter, with $0 < \lambda \ll 1$, would alleviate the extent of this effect (and its negative consequences). We measure the extent of the VIZE, with particular emphasis on the role of the λ parameter. In combination with the results of the qualitative analysis this allows us to test our hypotheses.

We also studied the (average) number of collective outcomes that is returned per profile, for the four different rules. This is a relevant figure; when a judgement aggregation rule returns multiple collective judgements, the individuals are faced with the (non-trivial) decision of determining the final collective judgement (singular). This part of the evaluation procedure is self-explanatory, and not further discussed in this section.

5.3.1 Qualitative Analysis

This section is devided in two parts: the ideas behind (which should be the justification of) our qualitative metrics, and the formal definitions of these metrics.

Motivation

Here we review the ideas behind our approach. In outline we measure the quality of an outcome in terms of efficiency and fairness. In economic theory, utilitarian and egalitarian approaches—respectively, prioritising efficiency and fairness—are often contrasted (Botan et al., 2021; Moulin, 1988). We already saw that maximising average utility is a pre-eminently utilitarian approach (Section 2.3). Hence, under the assumption of Hamming preferences, the Kemeny rule is a utilitarian rule. In that way, we could assess the efficiency of an arbitrary rule F by comparing the average utilities of F and F_{kem}. We searched for similar egalitarian benchmarks, and corresponding rules, to assess the *fairness* of a judgement aggregation rule. We found two benchmarks that can be used to assess the fairness of a judgement aggregation rule: lowest agreement and maximal agreement difference.

We introduced the Kemeny-Nash rule to apply Nash Social Welfare (NSW) maximisation to judgement aggregation. In other fields, both within the scope of social choice theory as well as outside of it, maximising NSW is known to yield solutions that are both efficient and fair (Caragiannis et al., 2019; Varian, 1974). This supports our choice to evaluate outcomes in terms of efficiency and fairness.

In economic theory, fairness is associated with egalitarian approaches, while efficiency is linked to utilitarian approaches. In our discussion of social welfare functions (Section 2.3) we saw that maximising utilitarian social welfare amounts to maximising the average utility. Under the assumption of Hamming preferences—which is widely accepted in judgement aggregation (see Section 2.3)—the Kemeny rule maximises utilitarian social welfare. Now, we have two important ingredients: a criterion for efficiency (average agreement) and a rule that optimises it (Kemeny rule). It is natural to ask ourselves whether we can find similar measures (and corresponding rules) to judge the fairness of a rule.

While maximising utilitarian social welfare amounts to maximising the average agreement, maximising egalitarian social welfare means maximising the minimal utility (Endriss, 2010). So, again assuming Hamming preferences, maximising the minimal agreement is tantamount to maximising egalitarian social welfare. In the literature the rule that does this is known as the MaxHam (or MaxHamming) rule, and studied by (among others) Botan et al. (2021), Endriss et al. (2020), and Lang et al. (2011).

Definition 5.1. Given any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$ with n judges, the MaxHam judgement aggregation rule (F_{mh}) is defined as:[7]

$$F_{\text{mh}}(\boldsymbol{J}) = \operatorname*{argmax}_{J \in \mathcal{J}(\Phi,\Gamma)} \min_{J' \in \boldsymbol{J}} \text{Agr}(J', J)$$

The definition states that, for any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$, the MaxHam rule picks out the consistent judgements $J \in \mathcal{J}(\Phi, \Gamma)$ that maximise the minimal agreement,

[7] When the input and output constraints do not coincide, i.e., if $\Gamma_{\text{in}} \neq \Gamma_{\text{out}}$, then Γ_{out} should be substituted for Γ.

$\min_{J' \in \boldsymbol{J}} \mathrm{Agr}(J', J)$. Just as we can use the average agreement to judge the efficiency of a rule, we can use the lowest agreement to judge the fairness of a rule.

However, Botan et al. (2021) argue that there are settings for which maximising the lowest agreement is not a good safeguard for fair solutions. In particular, in settings where not only the absolute utility (agreement) an individual enjoys is relevant, but also how it compares to the utility (agreement) of other individuals. That is, it is relevant whether individuals in a society *envy* each other. The authors argue that, for these settings, the maximal utility difference—assuming Hamming preferences, equal to the maximal agreement difference—is a suitable criterion to measure fairness. The criterion gives rise to the definition of the MaxEq rule, which was newly intorduced to the judgement aggregation literature. The rule is formally defined below.

Definition 5.2 (Botan et al., 2021)**.** Given any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$ with n judges, the MaxEq judgement aggregation rule is defined as:[8]

$$F_{\mathrm{mh}}(\boldsymbol{J}) = \underset{J \in \mathcal{J}(\Phi,\Gamma)}{\operatorname{argmin}} \left(\max_{J'' \in \boldsymbol{J}} \mathrm{Agr}(J'', J) - \min_{J' \in \boldsymbol{J}} \mathrm{Agr}(J', J) \right).$$

The definition says that the rule selects the consistent judgements $J \in \mathcal{J}(\Phi, \Gamma)$ that minimise the difference between the maximal agreement, $\max_{J'' \in \boldsymbol{J}} \mathrm{Agr}(J'', J)$, and the minimal agreement, $\min_{J' \in \boldsymbol{J}} \mathrm{Agr}(J', J)$.

We found three measures: average agreement, minimal agreement and maximal agreement difference. We will now procede to formally define the metrics that we based on these measures.

Definitions of the Metrics

We now translate our ideas from above into precise mathematical language. Given any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$ with n judges, it is straightforward to define the average agreement, lowest agreement, and maximal agreement difference—for a particular collective judgement $J \in \mathcal{J}(\Phi, \Gamma)$. But our metrics (for arbitrary rule F) are to be computed on a (large) multiset of different profiles, each corresponding to a (possibly non-singular) set of collective judgements (selected by rule F). We want to generalise the straightforward definitions for one profile, and a single collective judgement, to definitions for a multiset of profiles, each corresponding to a set of collective judgements. Moreover, to facilitate the comparison of results obtained from different experiments, we would like to somehow standardise our metrics.

We begin with the straightforward definitions, for the different criteria, based on a single profile and a single collective judgement. We define a weighted average to generalise the definitions to be applicable to a multiset of profiles. Finally, we standardise the generalised definition to obtain our final metrics.

[8]When the input and output constraints do not coincide, i.e., if $\Gamma_{\mathrm{in}} \neq \Gamma_{\mathrm{out}}$, then Γ_{out} should be substituted for Γ.

To repeat, as relevant criteria, we found the average agreement, lowest agreement and maximal agreement difference. The former criterion measures the efficiency of a rule, while the latter two evaluate the fairness of a rule. For an arbitrary profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$, with n agents, and consistent judgement $J \in \mathcal{J}(\Phi, \Gamma)$ we define the following measures. For the average agreement we define $\hat{S}_{\text{AVG}}$ as follows:

$$\hat{S}_{\text{AVG}}(\boldsymbol{J}, J) = \frac{1}{n} \cdot \sum_{J' \in \boldsymbol{J}} \big(\boldsymbol{J}(J') \cdot \mathrm{Agr}(J', J)\big) \tag{5.1}$$

Similarly, for the lowest agreement we define $\hat{S}_{\text{LOW}}$:

$$\hat{S}_{\text{LOW}}(\boldsymbol{J}, J) = \min_{J' \in \boldsymbol{J}} \mathrm{Agr}(J', J) \tag{5.2}$$

Finally, for the maximal agreement difference, we define:

$$\hat{S}_{\text{MD}}(\boldsymbol{J}, J) = \max_{J'' \in \boldsymbol{J}} \mathrm{Agr}(J'', J) - \min_{J' \in \boldsymbol{J}} \mathrm{Agr}(J', J) \tag{5.3}$$

We now want to generalise the definitions above, to be computable on a multiset of profiles, for a particular judgement aggregation rule. Let $P \subseteq \mathcal{J}(\Phi, \Gamma)^n$ be an arbitrary multiset of profiles for n judges, and F an arbitrary judgement aggregation rule. For $\odot \in \{\text{AVG}, \text{LOW}, \text{MD}\}$ we define the (weighted average) score $S_\odot$:[9]

$$S_\odot(F, P) = \frac{1}{|P|} \cdot \sum_{\boldsymbol{J} \in P} \left(\frac{P(\boldsymbol{J})}{|F(\boldsymbol{J})|} \cdot \sum_{J \in F(\boldsymbol{J})} \hat{S}_\odot(\boldsymbol{J}, J) \right) \tag{5.4}$$

That is, for every profile $\boldsymbol{J} \in P$, we compute $\hat{S}_\odot(\boldsymbol{J}, J)$ for all corresponding collective judgements $J \in F(\boldsymbol{J})$, and take the average. Subsequently, we take the average over the different profiles.

Finally, to be able to compare the results obtained from different experiments, we standardise the scores S_{AVG}, S_{LOW}, and S_{MD} in the following way. The scores are optimised by the Kemeny (F_{kem}), MaxHam (F_{mh}) and MaxEq (F_{me}) rule, respectively. To standardise the scores, we divide the score of rule F by the score of the rule that optimises it. The formal definitions of our (final) qualitative metrics are given below.

Definition 5.3. For any given judgement aggregation rule F and any given multiset of profiles $P \subseteq \mathcal{J}(\Phi, \Gamma)^n$ with n judges, the average agreement metric is defined as follows:

$$\mathrm{Agr}_{\text{AVG}}(F, P) = \frac{S_{\text{AVG}}(F, P)}{S_{\text{AVG}}(F_{\text{kem}}, P)}$$

See Equations (5.4) and (5.1) for the definition of S_{AVG}.

[9] $S_\odot$ is a *weighted* average in the sense that the weight of a particular collective judgement $J \in \mathcal{J}(\Phi, \Gamma)$, for a profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$, depends on the cardinality $P(\boldsymbol{J})$ and on the number of collective judgements that are selected for profile $\boldsymbol{J}$.

Definition 5.4. For any given judgement aggregation rule F and any given multiset of profiles $P \subseteq \mathcal{J}(\Phi, \Gamma)^n$ with n judges, the lowest agreement metric is defined as follows:

$$\mathrm{Agr}_{\mathrm{LOW}}(F, P) = \frac{S_{\mathrm{LOW}}(F, P)}{S_{\mathrm{LOW}}(F_{\mathrm{mh}}, P)}$$

See Equations (5.4) and (5.2) for the definition of S_{LOW}.

Definition 5.5. For any given judgement aggregation rule F and any given multiset of profiles $P \subseteq \mathcal{J}(\Phi, \Gamma)^n$ with n judges, the maximal agreement difference metric is defined as follows:

$$\mathrm{Agr}_{\mathrm{MD}}(F, P) = \frac{S_{\mathrm{MD}}(F, P)}{S_{\mathrm{MD}}(F_{\mathrm{me}}, P)}$$

See Equations (5.4) and (5.3) for the definition of S_{MD}.

Although the metrics formally depend on the multiset of profiles $P \subseteq \mathcal{J}(\Phi, \Gamma)^n$ that is tested, the idea is (of course) that we take P large enough, so that the results do not depend on the particular multiset of profiles that has been tested. For this reason, the explicit dependence on P, the multiset of tested profiles, is often omitted.

5.3.2 Quantitative Analysis

Our idea for this part of the experimental analysis is to formulate a criterion for the distance between different judgement aggregation rules. In particular, it would be interesting to have a measure for the distance between the Kemeny-Nash rule and the egalitarian rules on one side, and the Kemeny-Nash and Kemeny rule on the other side. It would also be interesting to compare these distances to the distance between the egalitarian rules and the Kemeny rule. The quantitative analysis provides a different way—next to the qualitative analysis—to determine the place of the Kemeny-Nash rule on the spectrum of egalitarian and utilitarian approaches. Finally, it would be interesting to see if the results of the quantitative and qualitative analysis are well matched. That is, whether rules that select judgements with similar qualities are also rules that are nearby.

The remainder of this section is structured as follows. Our metric is based on the set-theoretic concept of *symmetric difference*. Our exposition starts with an introduction of this concept. Next, we explain how the concept can be used as a metric for the distance between different judgement aggregation rules. To conclude, we outline our expectations for the results of the quantitative analysis.

In set theory, the symmetric difference between two sets, say A and B, is denoted as $A \triangle B$; it contains the elements that are in exactly one of the two sets, A or B. That is, the symmetric difference between A and B equals the union (of A and B) minus the intersection (of A and B). Formally, the symmetric difference may be defined as:

$$A \triangle B = (A \cup B) \setminus (A \cap B)$$

If A and B are subsets of some other set S, i.e., $A, B \subseteq S$; then the cardinality of the symmetric difference—$|A \triangle B|$—can be seen as a measure for the distance between the two subsets, see the example below.[10]

Example 5.1. Let $S = \{1, \ldots, 5\}$ and suppose we have three subsets, $A, B, C \subseteq S$: $A = \{1, 2\}$, $B = \{1, 5\}$ and $C = \{2, 3, 4\}$. Then we have $|A \triangle B| = 2$, $|A \triangle C| = 3$ and $|B \triangle C| = 5$. We can see that the (cardinality) of the symmetric difference of two (sub)sets depend on the cardinality of the individual (sub)sets, and the number of elements they have in common. Roughly, larger sets must have more elements in common (than smaller sets) for them to be at the same distance. This makes sense for a measure of the distance between two sets. △

To use the notion of symmetric difference to study judgement aggregation rules, we focus on the set of consistent judgements, $\mathcal{J}(\Phi, \Gamma)$. By definition a judgement aggregation rule is a function that, for any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$, returns a subset of consistent judgements: $F(\boldsymbol{J}) \subseteq \mathcal{J}(\Phi, \Gamma)$. Thus, for arbitrary rules F, F', and any profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$ we have: $F, F' \subseteq \mathcal{J}(\Phi, \Gamma)$. Now, if we are to determine the distance between the rules F, F'—on the basis of a single (arbitrary) profile $\boldsymbol{J}$—we can employ the notion of symmetric difference as follows:

$$\hat{D}(F, F', \boldsymbol{J}) = |F(\boldsymbol{J}) \triangle F'(\boldsymbol{J})| \tag{5.5}$$

Where $\hat{D}(F, F', \boldsymbol{J})$ is a measure for the distance between the two rules.[11]

However, in our analysis we want to determine the distance between two rules on the basis of a (large) multiset of profiles, $P \subseteq \mathcal{J}(\Phi, \Gamma)^n$. It is straightforward to modify the definition of $\hat{D}$ accordingly. We simply take the average of the profiles contained in P:

$$D(F, F', \boldsymbol{J}) = \frac{1}{|P|} \cdot \sum_{\boldsymbol{J} \in P} \big(P(\boldsymbol{J}) \cdot |F(\boldsymbol{J}) \triangle F'(\boldsymbol{J})| \big) \tag{5.6}$$

Finally, to facilitate the similarity between results obtained from different partial scenarios, we normalise the above definition; we divide by the cardinality of the set of consistent judgements. The formal definition of our symmetric difference metric, denoted as $d_{\triangle}$, is given below.

[10] This measure is in fact a (formal) metric (on S). The formal definition of a metric, as well as the proof that the symmetric difference is such a metric, is beyond the scope of this thesis; for details we refer to the book by Halmos (1974), Chapter 8.

[11] We emphasise that, given profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$, this metric is defined on the set of consistent judgements $\mathcal{J}(\Phi, \Gamma)$; i.e., the set $F(\boldsymbol{J}) \triangle F'(\boldsymbol{J})$ contains consistent judgements: $F(\boldsymbol{J}) \triangle F'(\boldsymbol{J}) \subseteq \mathcal{J}(\Phi, \Gamma)$. It does not contain agenda issues: $F(\boldsymbol{J}) \triangle F'(\boldsymbol{J}) \nsubseteq \Phi$. To define the metric on the set of consistent judgements has the advantage that we do not need to pay special attention to the number of judgements that is returned by the different rules; we compute the symmetric difference directly over the set of returned judgements (and not over the elements that are contained in these sets). On the negative side, the metric does not differentiate between any two distinct judgements (see Example 5.2 below).

Definition 5.6. The symmetric difference between two given judgement aggregation rules F and F', relative to a given multiset of profiles $P \subseteq \mathcal{J}(\Phi, \Gamma)^n$ with n judges, is defined as follows:[12]

$$d_{\triangle}(F, F', P) = \frac{1}{|P| \cdot |\mathcal{J}(\Phi, \Gamma)|} \sum_{\boldsymbol{J} \in P} \big(P(\boldsymbol{J}) \cdot |F(\boldsymbol{J}) \triangle F'(\boldsymbol{J})|\big)$$

Because $|F(\boldsymbol{J}) \triangle F'(\boldsymbol{J})| \leq |\mathcal{J}(\Phi, \Gamma)|$, this metric is guaranteed to lie between 0 and 1.

As we mentioned above, the symmetric difference of two sets depends on the cardinality of the individual sets and the cardinality of their intersection; it does not depend on the (dis)similarity between the elements that are not included in both sets. In our application to judgement aggregation, this means that we do not account for the similarity (or agreement) between any two *distinct* collective judgements, see the example below. This can be seen as a frailty of our metric.

Example 5.2. Take the PL3 (Positive Literals 3, see Section 5.1) partial scenario; and let F, F', and F'' be arbitrary judgement aggregation rules. Further, suppose that for some profile $\boldsymbol{J} \in \mathcal{J}(\Phi, \Gamma)^n$ we have:

$$\begin{aligned} F(\boldsymbol{J}) &= \{J\}, &\text{with } J &= \{\varphi_1, \varphi_2, \varphi_3\} \\ F'(\boldsymbol{J}) &= \{J'\}, &\text{with } J' &= \{\varphi_1, \varphi_2, \neg\varphi_3\} \\ F''(\boldsymbol{J}) &= \{J''\}, &\text{with } J'' &= \{\neg\varphi_1, \neg\varphi_2, \neg\varphi_3\} \end{aligned}$$

That is, for profile $\boldsymbol{J}$ the rules each return a single collective judgement, and no two rules return the same collective judgement. Judging from the returned judgements on profile $\boldsymbol{J}$, it would be reasonable to suggest that the distance between rule F and F' is smaller than the distance between F and F''. After all, if we compare the agreement between the different pairs of collective outcomes, we have: $\mathrm{Agr}(J, J') < \mathrm{Agr}(J, J'')$. However, according to our metric, the two distances are equal. Since, for any J, J, J'', with $J \neq J'$ and $J \neq J''$, we have: $|\{J\} \triangle \{J'\}| = |\{J\} \triangle \{J''\}|$. That is, the metric does not differentiate between any two non-identical collective judgements. $\triangle$

[12] We note that a potential weakness of this measure is that we do not take the total number of judgements that are returned by rules F and F' (respectively, $|F(\boldsymbol{J})|$ and $|F'(\boldsymbol{J})|$) into account.

To see the point, assume that rules F and F' always return 4 collective judgements (i.e., $|F(\boldsymbol{J})| = |F'(\boldsymbol{J})| = 4$), and the cardinality of the intersection $|F(\boldsymbol{J}) \cap F'(\boldsymbol{J})| = 4$ is also always 4. Further, suppose that for that rules $\tilde{F}$ and $\tilde{F'}$ always return 2 collective judgements (i.e., $|\tilde{F}(\boldsymbol{J})| = |\tilde{F'}(\boldsymbol{J})| = 2$), and the cardinality of the intersection $|\tilde{F}(\boldsymbol{J}) \cap \tilde{F'}(\boldsymbol{J})| = 2$ is also always 2. Everything else being equal, the symmetric difference between F and F' is twice as large as the symmetric difference between $\tilde{F}$ and $\tilde{F'}$; the results are not comparable.

Initially we were not aware of this weakness. However, our results showed that we never measure the symmetric difference between two rules that (on average) are very irresolute. Because of that, it is reasonable to suggest that our results are not too much effected by this weakness, and we left the definition as presented here.

To conclude, we lay down our expectations for the quantitative analysis. From what is said above, it is clear that we expect the distance between the Kemeny-Nash and the Kemeny rule, as well as the distance between the Kemeny-Nash and the egalitarian rules, to be smaller than the distance between the egalitarian rules and the Kemeny rule. Whether the results of the quantative analysis and the qualitative analysis will be congruent—in the sense that rules with solutions that are qualitatively comparable are near to each other—is hard to predict. It could be the case that, because the symmetric difference metric does not differentiate between any two distinct elements (i.e., consistent judgements), the results of the quantitative and qualitative analysis are not well aligned. In other words; two *distinct* consistent judgements, $J, J' \in \mathcal{J}(\Phi, \Gamma)$ with $J \neq J'$, might be very similar in a qualitative way—but this similarity is not accounted for by our quantitative metric—possibly resulting in a discrepancy between the results of our qualitative and quantitative analysis.

5.3.3 Variance-Increasing Zero-Effect

As explained in Section 2.4—in an earlier stage of this work—we assumed that in some cases the ZE might damage the economic efficiency of the Kemeny-Nash rule disproportionally, compared to the gained economic equity. In an attempt to set the 'bad' instances of the ZE apart from other instances, we introduced the VIZE (Definition 2.12). In our experimental analysis of the ZE we measured the extent of the VIZE. We extended Definition 2.12, to define it on a multiset of profiles, in the following way.

Definition 5.7. For a judgement aggregation rule F, with $F \in \{F_{\text{kn}}, F^{\lambda}_{\text{kn}}\}$, and a multiset of profiles $P \subseteq \mathcal{J}(\Phi, \Gamma)^n$ with n judges, the VIZE is defined as follows.

$$\text{VIZE}(F, P) = \frac{1}{|P|} \sum_{\boldsymbol{J} \in P} \left(\frac{1}{|F(\boldsymbol{J})|} \sum_{J \in F(\boldsymbol{J})} \text{VIZE}(J) \right)$$

With $\text{VIZE}(J) \in \{0, 1\}$ as in Definition 2.12.

5.4 Experimental Setup

We now give a full specification of our experimental setup. We include computational resources, software requirements and parameter settings. We begin with a note on terminology.

Note on terminology. Here we refer to an *experiment* as the composite of a partial scenario (Φ, Γ) and a specified number of judges n. For ease of exposition, we refer to an experiment by the name of the partial scenario (see Section 5.1) with an explicit specification of the number of judges. For example, we refer to the experiment with partial scenario JVC and $n = 5$ judges as JVC($n = 5$).

Computational resources. All experiments were done on an Intel Comet Lake 1.6GHz machine with 7.4GB RAM. For our implementation we used `Python 3` (v. 3.8.0).[13] For visualisation we made use of the `matplotlib` (v. 3.5.1) library.[14] To process propositional formulas (from the input file) we employed the `pyparsing` (v. 3.0.7) and `nnf` (v. 0.3.0) library.[15,16] To construct the generator for profile iteration, we used the `more_itertools` (v. 8.12.0) library.[17] Finally, for the ASP implementations, to solve the ASP program we used the ASP solver `Clingo`,[18] accessed through the `clingo` (v. 5.5.1) library.[19]

Parameter settings. We tested ten partial scenarios, each with three different numbers of judges; a total of 30 experiments. In each experiment we examined nine different settings of the λ parameter:

$$\lambda \in \{0.00, 0.01, 0.05, 0.10, 0.15, 0.25, 0.35, 0.45, 0.55\}$$

The growing increments were chosen after test experiments showed that the influence of a fixed increment $\Delta\lambda$ decreases, as λ increases.

The total number of profiles, for a single partial scenario, becomes highly intractable as the number of judges grows. For example, consider the CM partial scenario with $|\mathcal{J}(\Phi, \Gamma)| = 13$ consistent judgements (see Section 5.1). If we test this partial scenario for $n = 50$ judges, CM($n = 50$), the number of corresponding profiles equals $|\mathcal{J}(\Phi, \Gamma)^{50}| \approx 2.2 \cdot 10^{12}$. Whenever the total number of profiles exceeded $2.5 \cdot 10^5$, we did not do a full iteration, but used the sampling procedure instead. In the rule, the number of samples that were taken for a particular experiment depended on the total number of profiles. Because most of our metrics do not depend on the value of λ, whenever we observed that results that should not depend on λ were still varying, we deviated from this rule (and took more samples). In Table 5.2 we show the total number of different profiles and the number of profiles that were tested (for the different number of judges $n \in \{5, 15, 50\}$) for every partial scenario. When all profiles where tested, 'n/a' is indicated for the number of samples.

5.5 Results

In this section we present the obtained results. We focus on the general (average) patterns in the data, with particular emphasis on the role of the λ parameter. The exposition here is straightforward, the results are discussed in more detail in the next section. We (roughly) treat the results in order of importance: we begin with

[13] `https://www.python.org/`
[14] `https://pypi.org/project/matplotlib/`
[15] `https://pypi.org/project/pyparsing/`
[16] `https://pypi.org/project/nnf/`
[17] `https://pypi.org/project/more-itertools/`
[18] `https://potassco.org/clingo/`
[19] `https://pypi.org/project/clingo/`

	Profiles	Samples
JVC		
5 judges	56	n/a
15 judges	$8.2 \cdot 10^2$	n/a
50 judges	$2.3 \cdot 10^4$	n/a
CM		
5 judges	$6.2 \cdot 10^3$	n/a
15 judges	$1.7 \cdot 10^7$	$2.5 \cdot 10^5$
50 judges	$2.2 \cdot 10^{12}$	$2.5 \cdot 10^5$
SDC		
5 judges	$1.6 \cdot 10^4$	n/a
15 judges	$1.6 \cdot 10^8$	$2.5 \cdot 10^5$
50 judges	$2.1 \cdot 10^{14}$	$2.5 \cdot 10^5$
GR		
5 judges	$3.8 \cdot 10^5$	$2.5 \cdot 10^5$
15 judges	$5.1 \cdot 10^{11}$	$2.5 \cdot 10^5$
50 judges	$2.3 \cdot 10^{22}$	$1.0 \cdot 10^6$
PA3		
5 judges	$2.5 \cdot 10^2$	n/a
15 judges	$1.6 \cdot 10^4$	n/a
50 judges	$3.5 \cdot 10^6$	$1.0 \cdot 10^6$

	Profiles	Samples
PA4		
5 judges	$9.8 \cdot 10^4$	n/a
15 judges	$1.6 \cdot 10^{10}$	$2.5 \cdot 10^5$
50 judges	$5.7 \cdot 10^{18}$	$5.0 \cdot 10^5$
PL3		
5 judges	$7.9 \cdot 10^2$	n/a
15 judges	$1.7 \cdot 10^5$	n/a
50 judges	$2.6 \cdot 10^8$	$5.0 \cdot 10^5$
PL4		
5 judges	$1.6 \cdot 10^4$	n/a
15 judges	$1.6 \cdot 10^8$	$2.5 \cdot 10^5$
50 judges	$2.1 \cdot 10^{14}$	$1.0 \cdot 10^6$
PL6		
5 judges	$1.0 \cdot 10^7$	$2.5 \cdot 10^5$
15 judges	$4.4 \cdot 10^{15}$	$5.0 \cdot 10^5$
50 judges	$3.7 \cdot 10^{32}$	$1.0 \cdot 10^6$
PL7		
5 judges	$3.1 \cdot 10^8$	$2.5 \cdot 10^5$
15 judges	$6.8 \cdot 10^{19}$	$5.0 \cdot 10^5$
50 judges	$3.8 \cdot 10^{44}$	$1.0 \cdot 10^6$

Table 5.2: Summary of parameter settings, including the total number of profiles that corresponded to a partial scenario. 'n/a' indicates that the full iteration procedure was used.

the results of the qualitative analysis, followed by the results from the quantitative analysis. We continue with the results of the Variance-Increasing Zero-Effect (VIZE). Finally, we present some results that are concerned with the time performance of the implementation.

5.5.1 Qualitative Analysis

Broadly, the results of the qualitative analysis can be separated into two groups: experiments for which the quality of the solutions was affected by the value of λ, and experiments for which this was not the case.[20] Other than that, the different experiments exhibited largely the same patterns: The Kemeny-Nash rule produces solutions that convincingly combine efficiency and fairness, and although increasing the λ parameter had a small (positive) effect on the average agreement ($\mathrm{Agr}_{\mathrm{AVG}}$), this effect was outweighed by the effect on the lowest agreement ($\mathrm{Agr}_{\mathrm{LOW}}$) and maximal agreement difference ($\mathrm{Agr}_{\mathrm{MD}}$). Because of this, in our presentation of the results, we focus on averages, rather than single experiments. Having said this, we end this section with the results of three single experiments that deviate from the general pattern.

To recapitulate, see Section 5.3 for more details, we have three different metrics: average agreement metric ($\mathrm{Agr}_{\mathrm{AVG}}$), lowest agreement metric ($\mathrm{Agr}_{\mathrm{LOW}}$) and maximal agreement difference metric ($\mathrm{Agr}_{\mathrm{MD}}$). The former metric is a measure for the efficiency of a rule, while the latter two pertain to the fairness of a rule. The metrics are defined in a relative way; they are optimised by the Kemeny, MaxHam and MaxEq rule (respectively to aforementioned order). This is done in the following way: take $\mathrm{Agr}_{\mathrm{AVG}}$ (which is optimised by the Kemeny rule) and suppose that for some profile $\boldsymbol{J}$ the average agreement of the Kemeny and Kemeny-Nash solution are, respectively, 4 and 5.[21] Then, for the Kemeny-Nash rule, the corresponding metric score is $\mathrm{Agr}_{\mathrm{AVG}} = \frac{4}{5}$. For the Kemeny rule this score is 1 by definition. For readability, we may refer to (e.g.) the average agreement metric simply as the average agreement.

Figure 5.1 shows the average results of the qualitative analysis, where the average is taken over all 30 experiments. The three different colours designate the different metrics, while different line styles are used to discriminate the rules. In particular, for the different metrics: $\mathrm{Agr}_{\mathrm{AVG}}$ (blue), $\mathrm{Agr}_{\mathrm{LOW}}$ (black) and $\mathrm{Agr}_{\mathrm{MD}}$(red). Rule-wise: Kemeny-Nash (solid), Kemeny (dotted), MaxHam (dashed) and MaxEq (dashed-dotted). Now—this is the fundamental result of this chapter—Figure 5.1 shows that Kemeny-Nash rule outperforms the Kemeny rule when it comes to fairness. Especially when $\lambda = 0$, the $\mathrm{Agr}_{\mathrm{LOW}}$ and the $\mathrm{Agr}_{\mathrm{MD}}$ for the Kemeny-Nash rule are

[20] As a boundary, we specify that the quality is affected if the symmetric difference (metric) of the solutions for $\lambda = 0.55$ and the solutions for $\lambda = 0$ is at least 0.01. (With one exception—PL6($n = 50$) (symmetric difference metric was 0.0065)—this corresponds to the requirement that the relative change (in at least one of three) is at least 0.05.

[21] For simplicity we assume that $|F_{\mathrm{kem}}(\boldsymbol{J})| = |F_{\mathrm{kem}}(\boldsymbol{J})| = 1$; in other cases the metric is an average of the different solutions.

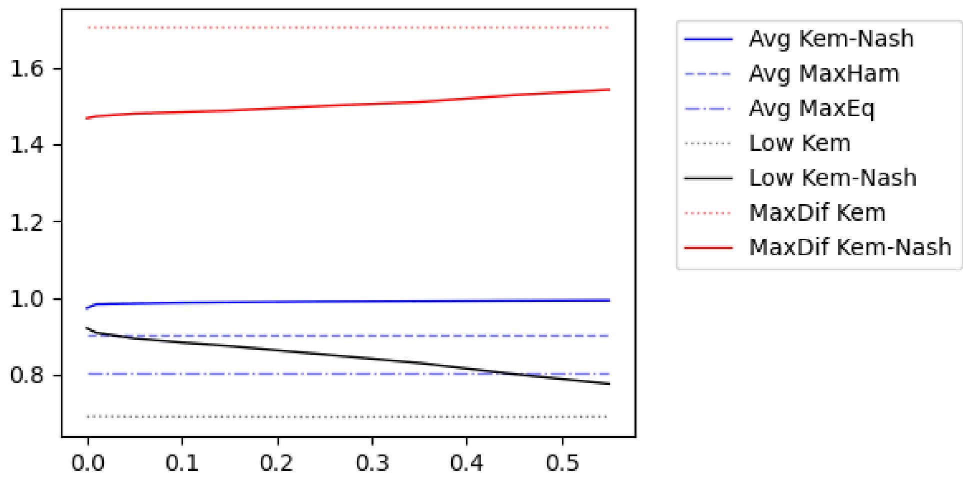
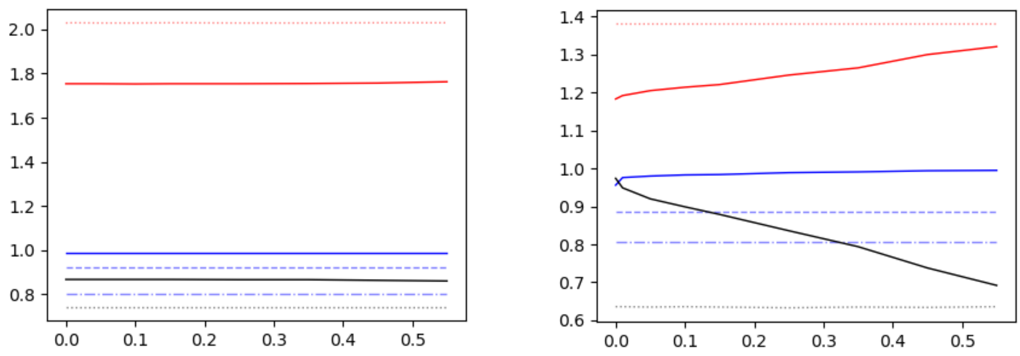


Figure 5.1: Results of the qualitative analysis, averaging over all experiments. On the x-axis we have the value of λ; the y-axis shows the metric scores. The solid lines indicate the metrics for the Kemeny-Nash rule: average agreement Agr_{AVG} (blue), lowest agreement Agr_{LOW} (black) and maximal agreement difference Agr_{MD} (red). The dotted lines are for the Kemeny rule: Agr_{LOW} (black) and Agr_{MD} (red). Finally, the Agr_{AVG} for the MaxHam and MaxEq rules are the dashed and dotted-dashed blue lines. For all metrics it holds that 1 is the optimal (and attainable) score; the further the score is from this value, the lower the quality of the solution. (See Section 5.3 for formal definitions of the metrics.) Notably, on average, the Kemeny-Nash rule yields solutions that are comparable to those of the Kemeny rule in terms of efficiency, but they are significantly better in terms of fairness. Increasing the λ parameter resulted in an increase of efficiency; but the increase is small compared to the corresponding decrease in fairness.

closer to the optimal values than they are for the Kemeny rule (revealed by the dotted lines being consistently further away from 1 than their corresponding solid line). We can further see that increasing λ does have a positive effect on Agr_{AVG}. However, this effect is small compared to the negative impact on both Agr_{LOW} and Agr_{MD} (the slope of the solid blue line is smaller than the slope of both the red and the black solid line).

As mentioned above, in some experiments the results were not (significantly) effected by increasing λ. In a similar fashion as above, Figure 5.2 shows the average results of the qualitative analysis; here the averages of the experiments that were not effected by variation of the λ parameter (5.2a) are separated from those that were effected (5.2b). Figure 5.2b shows, even more clearly than before, that the experiments that were effected by the variation of the λ parameter, were effected in a negative way. That is, the decrease of Agr_{LOW} and the increase in Agr_{MD} are crucially larger than the increase of Agr_{AVG}. As λ increases, we see that the quality of the Kemeny-Nash solutions approaches the quality of the Kemeny solutions; the



(a) Experiments without λ dependence. (b) Experiments with λ dependence.

Figure 5.2: Results of the qualitative analysis, averaging over the experiments that were not affected by variation of the λ parameter (left) and averaging over the experiments that were affected by the variation of λ (right). For both figures, we have the value of λ on the x-axis and the metric score on the y-axis. The solid lines indicate the metrics for the Kemeny-Nash rule: average agreement Agr_{AVG} (blue), lowest agreement Agr_{LOW} (black) and maximal agreement difference Agr_{MD} (red). The dotted lines are for the Kemeny rule: Agr_{LOW} (black) and Agr_{MD} (red). Finally, the Agr_{AVG} for the MaxHam and MaxEq rules are the dashed and dotted-dashed lines (blue). For all metrics it holds that 1 is the optimal (and attainable) score; the further the score is from this value, the lower the quality of the solution. (See Section 5.3 for formal definitions of the metrics. It is notable that (for $\lambda = 0$) the fairness (of the produced collective judgements) was significantly lower for the experiments that were not affected by the λ parameter. From the data of the experiments that were affected by λ; generally speaking, we can say that if the collective judgements are affected by λ, then this effect is negative. That is, the drop in the fairness of the collective judgements clearly outweighs the small increase in efficiency.

fairness advantages of the Kemeny-Nash rule are exchanged for a slight increase in efficiency. Finally, comparing Figure 5.2a and 5.2b: We see that, for $\lambda = 0$, the fairness of the solutions for the Kemeny-Nash and Kemeny rule, are remarkably lower for the experiments that were not effected by the increase of λ. In contrast, for the MaxHam and MaxEq rule, the efficiency of the solutions was comparable across the two kind of experiments.

To conclude this section, we show the results for the experiments PA3($n = 50$), PL3($n = 50$) and PL4($n = 50$) in Figure 5.3. These experiments clearly deviate from the average patterns. In the three experiments the quality of the Kemeny-Nash solutions is equal to the quality of MaxHam rule. In addition, for PA3($n = 50$) and PL3($n = 50$), the quality of the Kemeny-Nash solutions also equals the quality of the MaxEq solutions. Moreover, for PA3($n = 50$) and PL3($n = 50$), a large increase of Agr_{AVG} is accompanied by a negligible change in both Agr_{MD} and Agr_{LOW} . For PL4($n = 50$), the change in Agr_{AVG} is larger than the increase of Agr_{MD} (but smaller than the decrease in $\mathrm{Agr}_{\mathrm{LOW}}$).

5.5.2 Quantitative Analysis

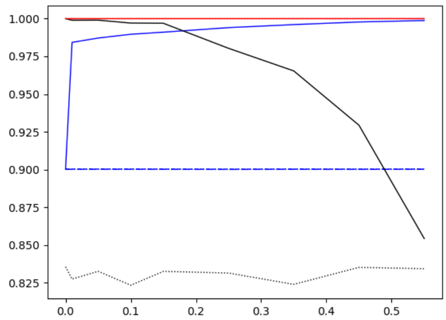
To repeat, in the quantitative analysis we measured the symmetric difference metric of the Kemeny and Kemeny-Nash rule.[22] Further we computed the symmetric differences of the Kemeny-Nash rule and the egalitarian rules; i.e., symmetric difference of Kemeny-Nash and MaxHam rule and that of the Kemeny-Nash and MaxEq rule. Similarly, we measured the symmetric difference of the Kemeny and MaxHam rule and that of the Kemeny and MaxEq rule. On average, the results show that the symmetric difference between the Kemeny and Kemeny-Nash rule is smaller than the other symmetric differences. Moreover, the symmetric difference between the Kemeny-Nash and MaxHam rule is smaller than that of the Kemeny and MaxHam rule; similarly, Kemeny-Nash and MaxEq rule is smaller than Kemeny and MaxEq rule. Figure 5.4 shows the average results. On the x-axis we have the value of the λ parameter; the y-axis shows the value of the symmetric difference metric, $d_{\triangle}$. The purple line depicts the symmetric difference of the Kemeny and Kemeny-Nash rule. The brown lines show the symmetric differences between Kemeny-Nash and MaxHam (solid) and between Kemeny-Nash and MaxEq (dashed). Similarly, the grey lines represent the symmetric differences between the Kemeny and MaxHam rules (solid) and between the Kemeny and MaxEq rule (dashed).

Not all experiments followed the average patterns. In the PL4($n = 50$) experiment we had that the symmetric difference between Kemeny and MaxHam was smaller than the symmetric difference between Kemeny-Nash and MaxHam, for $\lambda \in \{0.45, 0.55\}$. Similarly, for $\lambda \in \{0.45, 0.55\}$ in PL4($n = 50$), we had that the symmetric difference between Kemeny and MaxEq was smaller than the symmetric difference of the Kemeny-Nash and MaxEq rule. In five experiments—PA3($n = 15$), PA3($n = 50$), PA4($n = 50$), PL3($n = 50$) and PL4($n = 50$)—we had that for $\lambda = 0$, the symmetric difference between the Kemeny and Kemeny-Nash rule was greater than the symmetric difference between the Keney-Nash and MaxHam rule. Moreover—for PA3($n = 50$), PL3($n = 50$) and PL4($n = 50$), with $\lambda = 0$—we had that the symmetric difference of the Kemeny and Kemeny-Nash rule was also greater than that of the Kemeny-Nash and MaxEq rule. Figure 5.5 shows the results of the quantitative analysis for the experiments that did not follow the average trends.

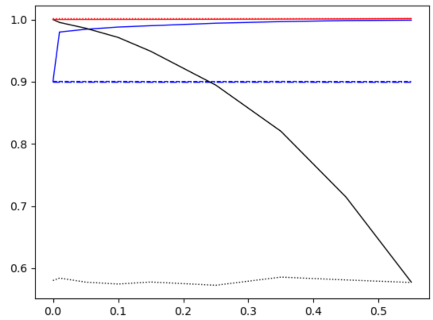
5.5.3 Zero-Effect Analysis

In all ten experiments with $n = 5$ judges, the Variance-Increasing Zero-Effect (VIZE) did not occur. Figure 5.6 shows the average VIZE, averaged over the experiments with $n = 15$ and $n = 50$ judges. On the x-axis we have the value of λ. It is noteworthy that, in general, the VIZE decreases as λ increases. The step from $\lambda = 0.45$ to $\lambda = 0.55$ is an exception; here the VIZE (slightly) increases. For 11/20 single experiments the VIZE increased from $\lambda = 0.45$ to $\lambda = 0.55$, for the other

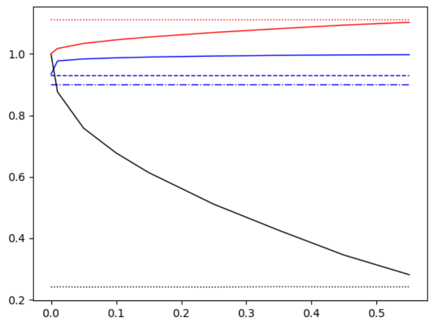
[22] When clear from context, we may simply refer to the metric as the symmetric difference.



(a) PA3($n = 50$).



(b) PL3($n = 50$).



(c) PL4($n = 50$).

Figure 5.3: Results of the qualitative analysis for PA3($n = 50$) (upper left), PL3($n = 50$) (upper right) and PL4($n = 50$) (below). The experiments are outliers; in these experiments the quality of the Kemeny-Nash solutions are equal to the quality of the MaxHam and MaxEq solutions (for PA3($n = 50$) and PL3($n = 50$)). For PL4($n = 50$) the quality of the Kemeny-Nash solutions is equal to the quality of the solutions of the MaxHam rule. Moreover, the experiments are outliers because of the effect that increasing λ from 0.00 to 0.01 has on the quality of the solutions. For PA3($n = 50$) and PL3($n = 50$), the increase in Agr_{AVG} is large, while the increase in Agr_{MD} and the decrease in Agr_{LOW} are negligible. For PL4($n = 50$), also when λ increases from 0.00 to 0.01, the increase in Agr_{AVG} is significantly larger than the increase in Agr_{MD} (but still crucially smaller than the decrease in Agr_{LOW}).

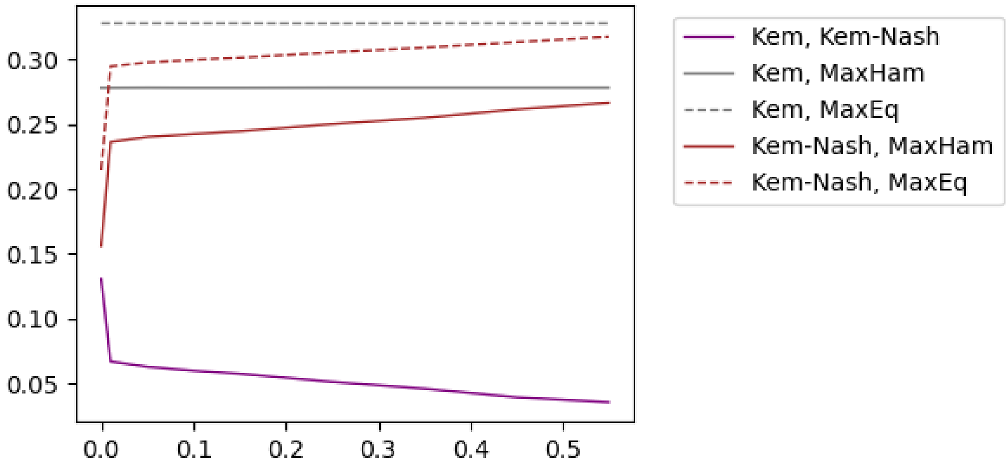


Figure 5.4: Results of the quantative analysis, averaging over all experiments. On the x-axis we have the value of λ; the y-axis shows the symmetric difference metric. The purple line represents the symmetric difference for the Kemeny and Kemeny-Nash rules. The brown lines show the symmetric difference between the Kemeny-Nash and egalitarian rules: symmetric difference between Kemeny-Nash and MaxHam rule (solid) and between the Kemeny-Nash and MaxEq rule (dashed). The grey lines indicate the symmetric difference between the Kemeny and MaxHam rule (solid), and between the Kemeny and MaxEq rule (dashed).

experiments the VIZE decreased on this interval. For all experiments the increase of the VIZE was non-positive on the range of $\lambda = 0$ to $\lambda = 0.45$.

In an attempt to find an analytical expression for the VIZE, we make the following proposal. Given a multiset of profiles $P \subseteq \mathcal{J}(\Phi, \Gamma)$ with n judgements, for the Kemeny-Nash rule F_{kn} we propose the following approximation for the VIZE (that depens on P only through the number of judgement n):

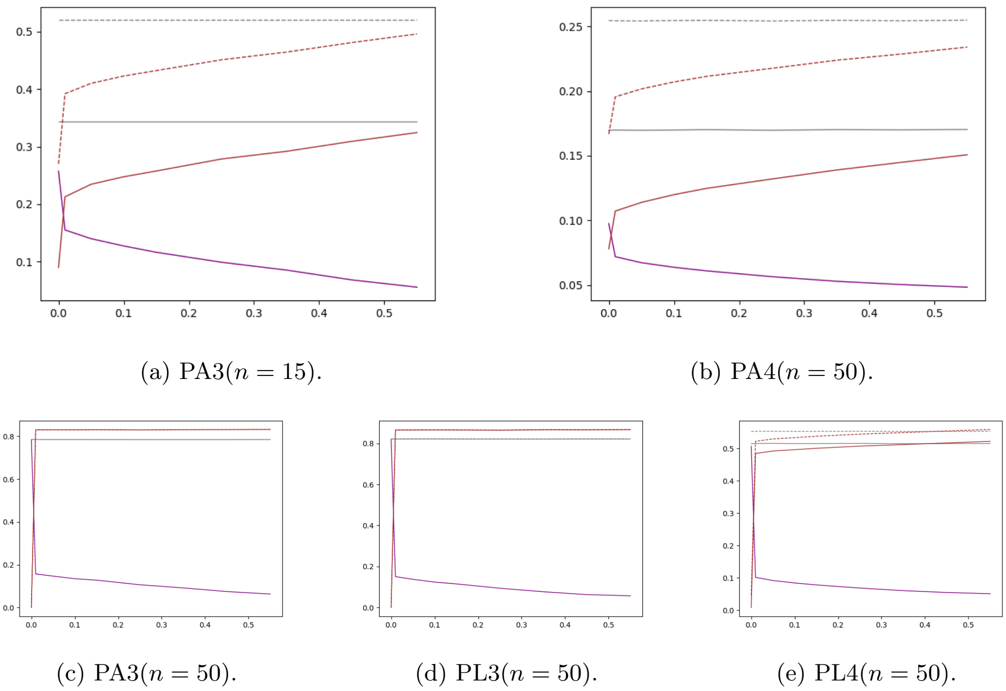
$$\text{VIZE}(F_{\text{kn}}, P) \approx C \cdot \exp(\sqrt{n}) \cdot \frac{|\{J \mid J, \overline{J} \in \mathcal{J}(\Phi, \Gamma)\}|}{|\mathcal{J}(\Phi, \Gamma)|} \cdot \frac{1}{|\mathcal{J}(\Phi, \Gamma)|}, \quad \text{with } C \in \mathbb{R} \tag{5.7}$$

We propose that the VIZE increases as:

(i) The number of judgements in the profile increase (the more judgements in the profile, the more consistent judgements are 'occupied' [i.e., appear in the profile], the higher the probability that two antipodal judgements are occupied);

(ii) The number of consistent antipodal judgements, divided by the total number of consistent judgement, increases (the greater this number, the higher the probability that any two judgements in the profile are antipodal);

(iii) The number of consistent judgements decreases (if there are only a few consistent judgements, the probability that a judgement is occupied is higher, so the probability that an antipodal pair is occupied is also higher).



(a) PA3($n = 15$).

(b) PA4($n = 50$).

(c) PA3($n = 50$).

(d) PL3($n = 50$).

(e) PL4($n = 50$).

Figure 5.5: Experiments that did not follow the average symmetric difference trends. On the x-axis we have the value of λ, and on the y-axis the symmetric difference metric. The purple line depicts the symmetric difference between Kemeny and Kemeny-Nash, the solid brown line between Kemeny-Nash and MaxHam and the dashed brown line between Kemeny-Nash and MaxEq. The solid and dashed grey lines show the symmetric difference between Kemeny and MaxHam and that of Kemeny and MaxEq (respectively). In all (presented) experiments the symmetric difference of Kemeny-Nash and MaxHam was smaller than the symmetric difference of Kemeny and Kemeny-Nash (for $\lambda = 0$). For the lower experiments we also have that the symmetric difference metric of Kemeny-Nash and MaxEq is smaller than the metric of Kemeny and Kemeny-Nash (also for $\lambda = 0$).

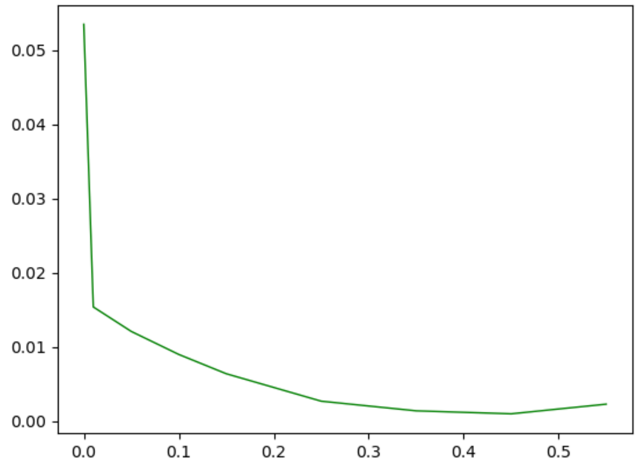


Figure 5.6: Results of the Variance-Increasing Zero-Effect (VIZE) analysis, averaging over all experiments with $n = 15$ and $n = 50$ judges. On the x-axis we have the value of the λ; the y-axis shows the value of VIZE. It is noteworthy that, as λ increases, the VIZE decreases; except from $\lambda = 0.45$ to $\lambda = 0.55$, here the value of the VIZE increases.

We believe that these relations make at least some sense, but the approximation should be really seen as a first guess. The particular relations we used (e.g., VIZE$\propto$ $\exp(\sqrt{n})$) are rather arbitrary, chosen to fit the data. In Table 5.3 we present the measured value of the VIZE(F_{kn}, P) (left) and, assuming our presented approximation, the induced proportionality constant C (right). For experiments that were not (significantly) effected by introducing the λ parameter, we print the values in bold. We included average values for the experiments that were not (significantly) effected by the value of λ, and for the experiments that were effected, separately.

We hypothesised (in Section2.4) that increasing λ leads to a reduction of the VIZE. If this is true, it would be reasonable to suggest that if the VIZE is already small before the introduction of λ, the results of the experiment are not (or less) significantly effected by λ. On average, we can see that both assumptions are (to some extent) supported by the data.

Considering the proposed analytical expression for the VIZE, we see that the different experiments imply very different proportionality constants; the proposed analytical expression is actually not really substantiated by the data.

5.5.4 Number of Solutions

As we mentioned above, the (average) number of solutions that a rule returns for a single profile is a relevant figure. If this number is larger than 1, the individuals

VIZE(F_{kn}, P)			Proportionality Constant C		
$n = 5$	$n = 15$	$n = 50$	$n = 5$	$n = 15$	$n = 50$
0.0000	.0135	.0363	1.2	6.0	147.2
0.0000	.0362	.2735	1.6	8.0	196.2
0.0000	.0104	.0069	**0.6**	2.8	69.7
0.0000	**.0024**	**.0015**	**0.1**	**0.4**	**9.2**
0.0000	**.0082**	.0729	**0.4**	**2.0**	49.1
0.0000	**.0017**	**.0014**	**0.0**	**0.1**	**2.3**
0.0000	.0476	.3052	1.2	6.0	147.2
0.0000	.0212	.2036	**0.6**	3.0	73.6
0.0000	**.0037**	.0149	**0.1**	**0.8**	18.4
0.0000	**.0029**	**.0052**	**0.1**	**0.4**	**9.2**
Avg. no λ effect			Avg. no λ effect		
.0000	**.0038**	**.0027**	**0.3**	**0.7**	**6.9**
Avg. λ effect			Avg. λ effect		
.0000	.0258	.1305	1.3	5.2	100.2

Table 5.3: VIZE (for $\lambda = 0$) and proportionality factor compared. Numbers for scenarios that were not affected by λ are printed in bold. Evidently, on average, these experiments have a lower VIZE and a lower proportionality factor.

are faced with the (non-trivial) problem of selecting a single outcome from the set of 'best' outcomes.

Figure 5.7 shows the average number of returned solutions for the different rules: Kemeny-Nash (yellow), Kemeny (blue), MaxHam (black) and MaxEq (red). It is noteworthy that moving from $\lambda = 0$ to $\lambda = 0.01$, the number of returned solutions for the Kemeny-Nash rule exhibits a significant drop, from 1.77 to 1.08. For $\lambda > 0$ the number of solutions of the Kemeny-Nash rule is lower than that of the Kemeny, MaxHam and MaxEq rule. It is also remarkable that the number of returned solutions for the Kemeny and Kemeny-Nash rule is crucially lower than that of the MaxHam and MaxEq rule.

5.5.5 Time Analysis

We tested the time performance for our different methods—BF, ASP with optimisation (ASP OPT) and ASP with saturation (ASP SAT)—with different rules. For the ASP SAT method roughly half of the scenarios did not terminate for the Kemeny-Nash rule ($\lambda = 0$) and $n = 5$ judges. For this reason the method is deemed useless for the purpose of this work, and the time results are not included.

For the BF solver and ASP OPT solver, the average time performance (over the ten partial profiles) are shown in Table 5.4. We show the time required to generate the scenario object (second column) and the times needed for a BF solver and ASP OPT solver iteration (third and fifth column, respectively). In our implementation

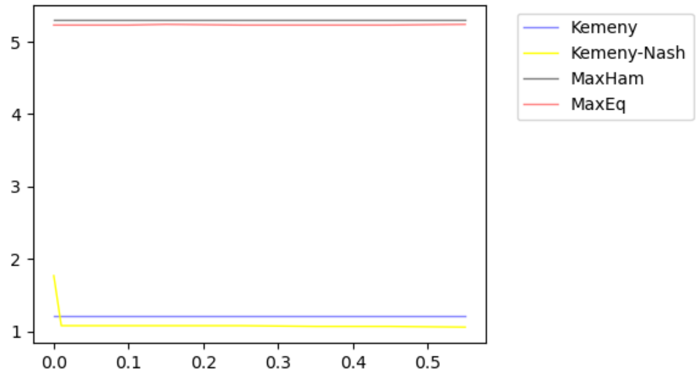


Figure 5.7: Average number of solutions, returned for a single profile, for the different rules. The x-axis shows the value of λ. The lines indicate the number of solutions (per profile) for the Kemeny-Nash rule (yellow), the Kemeny rule (blue), the MaxHam rule (black), and the MaxEq rule (red). It is notable that the number of solutions for the Kemeny-Nash rule shows a quick drop (1.77 to 1.08) when $\lambda = 0$ is increased to $\lambda = 0.01$. When λ is further increased, the average number of collective judgements is not further effected. For $\lambda \geq 0.01$ the number of solutions for the Kemeny-Nash rule is lower than for any of the other rules.

both methods require the creation of the scenario object. However, for the ASP (OPT) solver this step could be circumvented; for that reason we also include the sum of scenario creation and BF solver iteration (fourth column). To compare the time performance of the different methods we should compare the fourth and fifth column. It can be seen that for $n = 5$ judges the time performance for the BF solver and ASP OPT solver are comparable, but the ASP solver is faster. For the Kemeny rule with $n = 15$ judges, and $n = 50$ judges, the BF solver was significantly faster. For the Kemeny-Nash rule, with $\lambda = 0$, the ASP method failed to terminate (in any of the partial scenarios) for $n = 50$ judges; for $n = 15$ judges the BF solver was significantly faster. Finally, for the Kemeny-Nash rule with $\lambda = 0.05$, the ASP solver did not terminate for $n = 15$ and $n = 50$ judges (in any of the ten partial scenarios).

The results validate our choice to use the BF solver method for the final experiments.

5.6 Discussion

Here we discuss the results of the different parts of our analysis. Briefly, we can be happy about the results of the qualitative analysis, because they convincingly show that the Kemeny-Nash rule (indeed) provides solutions that are both fair (compared to the Kemeny rule) and efficient (compared to the egalitarian rules). This alone makes a further study of the Kemeny-Nash rule relevant. With our results we have

Rule (Judges)	Scen. (s)	BF (s)	Scen. + BF (s)	ASP OPT (s)
Kemeny rule				
5 judges	$6.68 \cdot 10^{-2}$	$1.62 \cdot 10^{-4}$	$6.69 \cdot 10^{-2}$	$4.08 \cdot 10^{-2}$
15 judges	$6.14 \cdot 10^{-2}$	$4.47 \cdot 10^{-4}$	$6.18 \cdot 10^{-2}$	$1.15 \cdot 10^{-1}$
50 judges	$5.99 \cdot 10^{-2}$	$1.10 \cdot 10^{-3}$	$6.10 \cdot 10^{-2}$	1.04
Kemeny-Nash rule ($\lambda = 0$)				
5 judges	$6.68 \cdot 10^{-2}$	$1.63 \cdot 10^{-4}$	$6.69 \cdot 10^{-2}$	$4.14 \cdot 10^{-2}$
15 judges	$6.14 \cdot 10^{-2}$	$4.48 \cdot 10^{-4}$	$6.18 \cdot 10^{-2}$	1.20
50 judges	$5.99 \cdot 10^{-2}$	$1.10 \cdot 10^{-3}$	$6.10 \cdot 10^{-2}$	n/a
Kemeny-Nash rule ($\lambda = 0.05$)				
5 judges	$6.68 \cdot 10^{-2}$	$2.16 \cdot 10^{-4}$	$6.70 \cdot 10^{-2}$	$4.35 \cdot 10^{-2}$
15 judges	$6.14 \cdot 10^{-2}$	$4.75 \cdot 10^{-4}$	$6.19 \cdot 10^{-2}$	n/a
50 judges	$5.99 \cdot 10^{-2}$	$1.20 \cdot 10^{-3}$	$6.11 \cdot 10^{-2}$	n/a

Table 5.4: Average (over ten partial scenarios) time performance of the BF solver method and the ASP (with optimisation) solver method, for the Kemeny and Kemeny-Nash rule (for $\lambda = 0$ and $\lambda = 0.05$). The second column shows the time required to create the scenario object. The third and fifth column show the (average) time of one solver iteration. Both solver methods (BF and ASP) require creation of the scenario object, however for the ASP solver it would be easy to circumvent this step. The results show that, even if we sum the time needed to create the scenario and the time needed for a BF solver iteration, the BF method is significantly faster for $n = 15$ and $n = 50$ judges. For $n = 5$ judges, the time performance is comparable, albeit the ASP method was a bit faster.

not been able to explain how the structure of a particular scenario relates to the (qualitative and quantitative) results for that scenario. In particular, while we were able to abstract general patterns in the data, we were not able to explain why particular scenarios did not follow the general trend(s). Moreover, we do not know how the structure of the scenario determines whether or not the scenario is affected by variation of the λ parameter.

Qualitative analysis. The results validate the hypothesis that the Kemeny-Nash rule mitigates the efficiency of the Kemeny rule with the fairness of the egalitarian rules (MaxHam and MaxEq). On average, the efficiency of the Kemeny-Nash rule was significantly higher than the efficiency of the egalitarian rules. Conversely, the fairness of the Kemeny-Nash solutions was crucially better than the fairness of the Kemeny solutions. In 3/30 experiments this average pattern was not followed. In PA3($n = 50$), PL3($n = 50$) and PL4($n = 50$), the quality of the Kemeny-Nash solutions coincided with the quality of the solutions of one (or both) of the egalitarian rules.

We hypothesised that introducing the λ parameter has a positive effect on the quality of the solutions (as it decreases the VIZE). This hypothesis is countered by the results. Overall we saw that (when λ increased) the increase in $\mathrm{Agr}_{\mathrm{AVG}}$ was small, compared to the decrease of $\mathrm{Agr}_{\mathrm{LOW}}$ and to the increase of $\mathrm{Agr}_{\mathrm{MD}}$. The experiments PA3($n = 50$) and PL3($n = 50$) are a clear exception to this rule, for these experiments changing λ from 0.00 to 0.01 is clearly improving the quality of the solutions.

Further, we saw that in 15/30 experiments the λ parameter had no (fundamental) effect on the quality of the solutions. Moreover, we saw that the quality of the solutions of this group of experiments was significantly lower for the Kemeny and Kemeny-Nash rule—compared to the experiments that did depend on λ, for $\lambda = 0$. We were unable to find out what features of an experiment determine the influence of the λ parameter, or determine the average quality of the solutions.

Quantitative analysis. On average, the results show that (for $\lambda = 0$) the symmetric differences between the Kemeny-Nash and Kemeny rule is smaller than the symmetric difference of both the Kemeny-Nash and MaxHam rule and the Kemeny-Nash and MaxEq rule. On average, the solutions of the Kemeny-Nash rule lie closer to those of the Kemeny rule than those of the MaxHam and MaxEq rules. This is in correspondence with the results of the qualitative analysis; for the Kemeny-Nash rule, $\mathrm{Agr}_{\mathrm{AVG}}$ was consistently closer to 1 than both $\mathrm{Agr}_{\mathrm{LOW}}$ and $\mathrm{Agr}_{\mathrm{MD}}$.

Moreover, we saw that the symmetric difference between the Kemeny-Nash rule and the MaxHam (MaxEq) rule was lower than the symmetric difference between the Kemeny and MaxHam (MaxEq) rule. This also corresponds to the results of the qualitative analysis: $\mathrm{Agr}_{\mathrm{LOW}}$ ($\mathrm{Agr}_{\mathrm{MD}}$) were generally closer to 1 for the Kemeny-Nash rule (compared to the Kemeny rule).

Variance-Increasing Zero-Effect analysis. The results have shown that the VIZE is not so helpful to understand the difference between the Kemeny-Nash and Kemeny judgement aggregation rules.

As we expected, as λ increases, the VIZE (generally) decreases; but this decrease was not accompanied by a better quality of the solutions. The results show that the VIZE increases in the step from $\lambda = 0.45$ to $\lambda = 0.55$; the reason for this is not known.

Running time: BF versus ASP. In our implementation the time performance of the ASP solver was either comparable to that of the BF solver, or it was significantly slower. This is contrary to what one would expect: De Haan and Slavkovik (2019) and Botan et al. (2021) argue that ASP encodings should be particularly suitable for solving judgement aggregation problems. We note that our ASP encoding is 'wrapped' in a python package, which is moreover integrated with a BF solver method; we suggest that this could be the reason that the full potential of the ASP framework is not utilised in our implementation.

Number of solutions. From our examination of the (average) number of collective outcomes that is returned for a single profile, for the different rules, we learned that this number is significantly higher (approximately four times) for the egalitarian rules than it is for the Kemeny-Nash and Kemeny rule. Further, we learned that the average number of solutions quickly drops (from 1.77 to 1.08) when $\lambda = 0$ is increased to $\lambda = 0.01$. Contrary to our results of the qualitative analysis, this is an argument in favour of setting λ to a small non-zero value.

Chapter 6

Conclusion

In this chapter we summarise our results and suggest directions for future research.

6.1 Thesis Summary

In this thesis we studied a modification to the Kemeny rule; we introduced the Kemeny-Nash judgement aggregation procedure. While the Kemeny rule can be interpreted as maximising utilitarian social welfare, by maximising the product of individual utilities, the Kemeny-Nash rule maximises Nash social welfare instead. In other areas of social choice theory, notably in the fair division literature, maximising Nash social welfare is known to provide solutions that are both fair and efficient.

Alongside the Kemeny-Nash rule, we introduced the parameterised Kemeny-Nash rule. In the parameterised variant—when we maximise the product of individual utilities—all zero-terms are replaced by a small positive value $0 < \lambda \ll 1$.

In the main body of this work (Chapters 3, 4, and 5) used different tools to examine the collective judgements that are returned by the Kemeny-Nash rule, and set them apart from the judgements that are returned by the Kemeny rule. In subsequent order, we studied axiomatic properties, the computational complexity of the outcome determination problem, and we did experiments. Below we summarise the main results for each of the chapters.

In Chapter 3 we studied the characterisation of the Kemeny rule by Nehring and Pivato (2022). While the Kemeny-Nash rule satisfied none of the three axioms ((`C`), (`R`), and (`ESME`)) that characterise the Kemeny rule, the parameterised Kemeny-Nash rule did satisfy two of them ((`C`) and (`R`)).

For the other axioms we studied, neither the Kemeny-Nash, nor its parameterised variant, satisfied (`MP`), this can be seen as a serious deficit of the rules. Further, we studied two equity principles ((`SHE`) and (`PD`)),

In Chapter 4 we showed that the outcome determination problem for both the Kemeny-Nash rule, as well as the parameterised Kemeny-Nash rule, is contained in Δ_2^p, and hard for Θ_2^p. Although we were not able to show completeness, if the problems are complete for some class C, the possibilities of C are narrowed down. For the Kemeny rule the outcome determination problem is known to be complete for Θ_2^p.

In Chapter 5 we did an experimental analysis. The analysis convincingly showed that the collective judgements that are computed by the Kemeny-Nash rule are significantly better in terms of fairness, and only slightly worse in terms of efficiency, than the collective judgements that are produced by the Kemeny rule. Our hypothesis that the instances of the VIZE—roughly, scenarios for which the spread of the individual utilities is higher under the Kemeny-Nash rule, than under the Kemeny rule—was refuted. The Kemeny-Nash rule outperformed the parameterised Kemeny-Nash rule.

6.2 Future Work

Even though we were unable to set the fairness of the Kemeny-Nash rule apart from the fairness of the Kemeny rule, our experimental analysis substantiates a further study of the Kemeny-Nash rule. To summarise important open questions of this work; it would be relevant to (i) find or introduce fairness properties that distinguish the Kemeny-Nash rule from the Kemeny rule, (ii) further analyse completeness of the outcome determination problem for the Kemeny-Nash rule, and (iii) determine in what circumstances the Kemeny-Nash rule be (particularly) suitable; we suggest that finding an analytical expression for the VIZE might be helpful in this respect.

Finally, although we did consider different judgement aggregation scenarios in our experimental approach, in no part of this work we treated the structure of the agenda in any principled manner. As we mentioned in the In other judgement aggregation problems—especially problems of an axiomatic nature—examining the structure of the agenda led to new insights. Thus, a suggestion for future research would be to study how the structure of the agenda influences the behaviour of the Kemeny-Nash rule.

Bibliography

Arora, Sanjeev and Boaz Barak (2009), *Computational Complexity: A Modern Approach*, 1st edition. Cambridge University Press.

Arrow, Kenneth J. (1951), *Social Choice and Individual Values*, 1st edition. New Haven, CT: Cowles Foundation.

Awad, Edmond, Richard Booth, Fernando Tohmé, and Iyad Rahwan (2017), "Judgement Aggregation In Multi-Agent Argumentation." *Journal of Logic and Computation*, 27, 227–259.

Baumeister, Dorothea, Gábor Erdélyi, Olivia J. Erdélyi, and Jörg Rothe (2015), "Complexity of Manipulation and Bribery in Judgment Aggregation for Uniform Premise-Based Quota Rules." *Mathematical Social Sciences*, 76, 19–30.

Baumeister, Dorothea, Jörg Rothe, and Ann-Kathrin Selker (2017), "Strategic Behavior in Judgment Aggregation." In *Trends in Computational Social Choice* (Ulle Endriss, ed.), chapter 8, AI Access.

Botan, Sirin, Ronald de Haan, Marija Slavkovik, and Zoi Terzopoulou (2021), "Egalitarian Judgment Aggregation." In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2021)*, 214–222, IFAAMAS.

Botan, Sirin, Ronald de Haan, Marija Slavkovik, and Zoi Terzopoulou (2023), "Egalitarian Judgment Aggregation." *Autonomous Agents and Multi-Agent Systems*, 37, 16.

Brandt, Felix, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (2016), "Introduction to Computational Social Choice." In *Handbook of Computational Social Choice* (F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, eds.), chapter 1, Cambridge University Press.

Caragiannis, Ioannis, David Kurokawa, Hervé Moulin, Ariel D. Procaccia, Nisarg Shah, and Junxing Wang (2019), "The Unreasonable Fairness of Maximum Nash Welfare." *ACM Transactions on Economics and Computation (TEAC)*, 7, 1–32.

Chen, Zhi-Zhong and Seinosuke Toda (1995), "The Complexity of Selecting Maximal Solutions." *Information and Computation*, 119, 231–239.

De Haan, Ronald and Marija Slavkovik (2017), "Complexity Results for Aggregating Judgments Using Scoring or Distance-Based Procedures." In *Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems (AAMAS-2017)*, 952–961.

De Haan, Ronald and Marija Slavkovik (2019), "Answer Set Programming for Judgment Aggregation." In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)*, AAAI Press.

Dietrich, Franz (2010), "The Possibility of Judgment Aggregation on Agendas with Subjunctive Implications." *Journal of Economic Theory*, 145, 603–638.

Dietrich, Franz and Christian List (2007), "Judgment Aggregation by Quota Rules: Majority Voting Generalized." *Journal of Theoretical Politics*, 19, 391–424.

Dietrich, Franz and Philippe Mongin (2010), "The Premiss-Based Approach to Judgment Aggregation." *Journal of Economic Theory*, 145, 562–582.

Dubey, Ram Sewak (2016), "On Construction of Social Welfare Orders Satisfying Hammond Equity and Weak Pareto Axioms." *Mathematical Social Sciences*, 84, 119–124.

Eiter, Thomas and Georg Gottlob (1995), "Identifying the Minimal Transversals of a Hypergraph and Related Problems." *SIAM Journal on Computing*, 24, 1278–1304.

Endriss, Ulle (2010), "Lecture Notes on Fair Division." ILLC, University of Amsterdam, URL `https://research.illc.uva.nl/COMSOC/estoril-2010/docs/endriss-fair-division.pdf`.

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

Endriss, Ulle, Ronald de Haan, Jérôme Lang, and Marija Slavkovik (2020), "The Complexity Landscape of Outcome Determination in Judgment Aggregation." *Journal of Artificial Intelligence Research*, 69, 687–731.

Grandi, Umberto (2012), *Binary Aggregation with Integrity Constraints*. Ph.D. thesis, University of Amsterdam, ILLC.

Grandi, Umberto and Ulle Endriss (2013), "Lifting Integrity Constraints in Binary Aggregation." *Artificial Intelligence*, 199–200, 45–66.

Halmos, Paul Richard (1974), *Measure Theory*, [new ed.] edition. Graduate Texts in Mathematics 18, Springer-Verlag.

Hammond, Peter Jackson (1976), "Equity, Arrow's Conditions, and Rawls' Difference Principle." *Econometrica: Journal of the Econometric Society*, 793–804.

Hara, Chiaki, Tomoichi Shinotsuka, Kotaro Suzumura, and Yongsheng Xu (2008), "Continuity and Egalitarianism in the Evaluation of Infinite Utility Streams." *Social Choice and Welfare*, 31, 179–191.

Jensen, Johan L. W. V. (1906), "Sur Les Fonctions Convexes Et Les Inégalités Entre Les Valeurs Moyennes." *Acta Mathematica*, 30, 175–193.

Kornhauser, Lewis A. and Lawrence G. Sager (1986), "Unpacking the Court." *The Yale Law Journal*, 96, 82–117.

Kornhauser, Lewis A. and Lawrence G. Sager (1993), "The One and the Many: Adjudication in Collegial Courts." *California Law Review*, 81, 1–59.

Krentel, Mark W. (1988), "The Complexity of Optimization Problems." *Journal of Computer and System Sciences*, 36, 490–509.

Lang, Jérôme, Gabriella Pigozzi, Marija Slavkovik, and Leendert van der Torre (2011), "Judgment Aggregation Rules Based on Minimization." In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, 238–246.

List, Christian and Philip Pettit (2002), "Aggregating Sets of Judgments: An Impossibility Result." *Economics and Philosophy*, 18, 89–110.

Miller, Michael K. and Daniel Osherson (2009), "Methods for Distance-Based Judgment aggregation." *Social Choice and Welfare*, 32, 575–601.

Mongin, Philippe (2008), "Factoring Out the Impossibility of Logical Aggregation." *Journal of Economic Theory*, 141, 100–113.

Moulin, Hervé (1988), *Axioms of Cooperative Decision Making*, volume 15 of *Econometric Society Monographs*. Cambridge University Press.

Nash, John F. (1950), "The Bargaining Problem." *Econometrica*, 18, 155–162.

Nehring, Klaus and Marcus Pivato (2022), "The Median Rule in Judgement Aggregation." *Economic Theory*, 73, 1051–1100.

Pettit, Philip (2001), "Deliberative Democracy and the Discursive Dilemma." *Philosophical Issues*, 11, 268–299.

Pigozzi, Gabriella (2006), "Belief Merging and the Discursive Dilemma: an Argument-Based Account to Paradoxes of Judgment Aggregation." *Synthese*, 152, 285–298.

Sen, Amartya (1974), "Rawls Versus Bentham: An Axiomatic Examination of the Pure Distribution Problem." *Theory and Decision*, 4, 301–309.

Sen, Amartya (1976), "Welfare Inequalities and Rawlsian Axiomatics." *Theory and Decision*, 7, 243–262.

Slavkovik, Marija (2014), “Not All Judgment Aggregation Should be Neutral.” In *ECSI*, 198–211. CHECK REFERENCE.

Varian, Hal R. (1974), “Equity, Envy, and Efficiency.” *Journal of Economic Theory*, 9, 63–91.

Wagner, Klaus W. (1990), “Bounded Query Classes.” *SIAM Journal on Computing*, 19, 833–846.