Majority-Strategyproofness in Judgment Aggregation

Botan, S.; Endriss, U.

Publication date
2020

Document Version
Author accepted manuscript

Published in
AAMAS'20

Citation for published version (APA):
Botan, S., & Endriss, U. (2020). Majority-Strategyproofness in Judgment Aggregation. In *AAMAS'20: proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems : May 9-13, 2020, Auckland, New Zealand* (pp. 186-194). International Foundation for Autonomous Agents and Multiagent Systems. http://www.ifaamas.org/Proceedings/aamas2020/pdfs/p186.pdf

Majority-Strategyproofness in Judgment Aggregation

Sirin Botan
Institute for Logic, Language and Computation
University of Amsterdam

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

ABSTRACT

By a combination of well-known results in judgment aggregation, it is essentially impossible to design an aggregation rule that simultaneously satisfies two crucial requirements: to always return an outcome that is logically consistent, and to be immune to strategic manipulation. To address this dilemma, we put forward a novel notion of strategyproofness, which requires immunity to strategic manipulation only in certain well-defined situations—namely when either the truthful profile of individual judgments *or* the profile a would-be manipulator is trying to reach are majority-consistent. We argue that this constitutes an attractive compromise for aggregation rules one may want to use in practice, and we prove that several important rules are strategyproof in this sense. This includes, in particular, all rules belonging to the family of additive majority rules, such as the Kemeny rule and the Slater rule.

KEYWORDS

Judgment Aggregation; Social Choice Theory; Strategic Manipulation

ACM Reference Format:
Sirin Botan and Ulle Endriss. 2020. Majority-Strategyproofness in Judgment Aggregation. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 9 pages.

1 INTRODUCTION

Judgment aggregation is a powerful framework for analysing multi-agent decision making scenarios [21, 29]. In judgment aggregation we model the views held by individual agents as sets of propositional formulas, and try to design rules for aggregating such judgments into a single collective judgment that adequately represents the views held by the group. It generalises preference aggregation as traditionally studied in social choice theory [8] and is closely related to belief merging as long studied in AI [17].

A well-known difficulty in judgment aggregation, closely related to classical impossibility theorems in other areas of social choice theory [1, 20, 38], is the fact that it is essentially impossible to design an aggregation rule that is immune to manipulation by strategic agents while also ensuring that the rule will always return an outcome that is logically consistent [10, 12, 29, 34]. In this paper we propose—and study in detail—a weakening of the standard notion of strategyproofness aimed at circumventing this difficulty. This allows us to identify several judgment aggregation rules that offer a good compromise between the conflicting requirements of strategyproofness and guaranteed consistency. Before discussing this idea, let us first illustrate the problem.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

Example 1.1. Suppose three agents need to arrive at a collective decision regarding the four propositions p, q, $p \wedge q$, and $p \leftrightarrow q$. Let us consider two aggregation rules they might use. First, they could use the *premise-based rule*, which amounts to taking majority decisions on the premises p and q and then inferring the truth values for the conclusions $p \wedge q$ and $p \leftrightarrow q$. Second, they could use the *majority rule* and decide on all four propositions by majority.

	p	q	$p \wedge q$	$p \leftrightarrow q$
Agent 1	Yes	Yes	Yes	Yes
Agent 2	Yes	No	No	No
Agent 3	No	Yes	No	No
Premise-based	Yes	Yes	Yes	Yes
Majority	Yes	Yes	No	No

Under the premise-based rule, the outcome agrees with agent 2 on only a single proposition (p). Suppose agent 2 wants to maximise the number of propositions on which the outcome agrees with her own true judgment. Then she could manipulate and pretend that she disagrees with p (and thus agrees with $p \leftrightarrow q$), in which case the premise-based rule would agree with her true original judgment on *two* propositions ($p \wedge q$ and $p \leftrightarrow q$).

The majority rule does not suffer from this deficiency: if you care about the number of propositions agreeing with your own judgment, then it is always in your best interest to report your true judgment [10]. But the majority rule suffers from another—arguably even more debilitating—shortcoming: sometimes, as demonstrated by our example, the outcome returned by the rule will be *inconsistent*. Indeed, the majority rule proposes to accept both p and q but to reject their logical consequences $p \wedge q$ and $p \leftrightarrow q$. This is an instance of the infamous *discursive dilemma* [29]. △

Our proposal is to consider a carefully weakened notion of strategyproofness, parametrised by some domain $\mathcal{D}$ of profiles of individual judgments. Under this novel notion of strategyproofness we require immunity to manipulation only in two situations: when the truthful profile belongs to $\mathcal{D}$ or when the profile the manipulating agent might deviate to belongs to $\mathcal{D}$. While this notion is related to the idea of imposing a restriction on the domain on which the aggregation rule is defined [28], we do not actually impose any such restriction in our work. We specifically focus on the domain $\mathcal{M}$ of profiles that guarantee consistent outcomes under the majority rule. A rule that is $\mathcal{M}$-strategyproof will be immune to manipulation in all those cases in which the (strategyproof) majority rule would return a consistent outcome (and thus would be useable at all), while also returning consistent outcomes for all other profiles.

Related work. The study of strategic manipulation in judgment aggregation was initiated by Dietrich and List [10], who showed that only rules belonging to a very narrowly defined family are immune to manipulation. These rules, however, are inadequate for many applications, because they cannot guarantee the consistency of outcomes [12, 29, 34]. Prior work aimed at addressing this dilemma has focused on three approaches: (*i*) identifying *restricted domains* of profiles of individual judgments for which better performance of certain rules can be guaranteed [28]; (*ii*) analysing the computational complexity of the manipulation problem as a means of providing *complexity barriers* against unwanted behaviour [3, 15, 22]; and (*iii*) studying the extent to which limiting access to relevant information may serve as an effective *informational barrier* against manipulation [41]. The route we take in this paper—which is to offer a more fine-grained analysis of the concept of strategyproofness itself—is complementary to these approaches.

Related ideas on nonstandard variants of strategyproofness have been explored outside of judgment aggregation, albeit to a limited extent. Sato [36] considers the strategyproofness of voting rules with respect to manipulations restricted to adjacent preferences. Rules that are *adjacency-strategyproof* eliminate the possibility of manipulations that do not require agents to make large changes to their reported preference. The most closely related body of research to our own work concerns the manipulation of *social welfare functions* that map profiles of preference orders to collective preference orders. Bossert and Storcken [6] were the first to study this problem and suggested to model the preferences of agents over alternative preference orders in terms of the *Kendall-tau distance* between orders. While guaranteeing strategyproofness in this model is generally impossible [2, 6], Bossert and Sprumont [5] obtain positive results for a weak form of strategyproofness that considers only manipulations which bring about an outcome that is *between* an agent's true preference order and the current outcome. They find that several important rules, among them the Kemeny and the Slater rule, are strategyproof in this sense. Sato [37] presents several refinements of these results.

Contribution. We introduce the novel notion of *domain-strategyproofness*, together with its instantiation to *majority-strategyproofness*, and we show that all aggregation rules belonging to the family of *additive majority rules* [32] enjoy the latter property. This includes, in particular the well-known Kemeny, Slater, and Leximax rules. We also identify two further rules, the Maximal Condorcet rule and the Ranked Agenda rule, that satisfy the same property under somewhat more restrictive assumptions on the preferences of agents. To demonstrate that majority-strategyproofness is by no means a universal property of judgment aggregation rules—and not even of majority-preserving rules—we show that one further well-known rule, the Dodgson rule, fails to satisfy this property.

Paper outline. The remainder of the paper is organised as follows. After recalling relevant preliminaries from the judgment aggregation literature in Section 2, we define and discuss various notions of strategyproofness in Section 3. We then present our technical results for additive majority rules and further majority-preserving rules in Sections 4 and 5, respectively. We conclude in Section 6.

2 PRELIMINARIES

In this section we introduce the model we will be using throughout the paper. This is the standard set-based model of judgment aggregation going back to the seminal work of List and Pettit [29].

2.1 Notation and Terminology

Let $N = \{1, \ldots, n\}$ be a finite set of *agents*. We will assume that n is odd to avoid having to make tie-breaking decisions when computing majorities. Each agent submits their judgments on a (nonempty) set of formulas of propositional logic $\Phi = \Phi^+ \cup \Phi^-$, called the *agenda*, where Φ^+ is a set of nonnegated formulas, and $\Phi^- = \{\neg\varphi \mid \varphi \in \Phi^+\}$. A *judgment* J is a subset of Φ. We use $\mathcal{J}(\Phi) \subseteq 2^\Phi$ to denote the set of all judgments that are (logically) *consistent* as well as *complete*—in the sense of including one of φ and $\neg\varphi$ for every $\varphi \in \Phi^+$. Observe that any consistent judgment will also be *complement-free*, meaning that it cannot include both φ and $\neg\varphi$ for any $\varphi \in \Phi^+$. Any element of $\mathcal{J}(\Phi)$ is a permissible judgment J_i for an agent $i \in N$. We write $J =_\varphi J'$ to mean that judgments J and J' agree on formula φ.

The *Hamming distance* between two judgments J and J' in $\mathcal{J}(\Phi)$ is defined as $H(J, J') := |J \setminus J'| = |J' \setminus J|$. Thus, $H(J, J')$ is the number of elements in Φ^+ on which J and J' disagree. We say that judgment J' is *between* J and J'', if $J \cap J'' \subseteq J' \subseteq J \cup J''$. Observe that $J \cap J'' \subseteq J'$ if and only if $J' \subseteq J \cup J''$ in case all three judgments are both complete and complement-free.

A *profile* $\boldsymbol{J} = (J_1, \ldots J_n) \in \mathcal{J}(\Phi)^n$ is a vector of individual judgments, one for each agent in N. For any such profile $\boldsymbol{J}$ and any $\varphi \in \Phi$, the set $N_\varphi^{\boldsymbol{J}} := \{i \in N \mid \varphi \in J_i\}$ is the set of *supporters* of proposition φ, with $n_\varphi^{\boldsymbol{J}} := |N_\varphi^{\boldsymbol{J}}|$. The *majority judgment* associated with a given profile $\boldsymbol{J}$ is defined as $m(\boldsymbol{J}) := \{\varphi \in \Phi \mid n_\varphi^{\boldsymbol{J}} > \frac{n}{2}\}$. We say that profiles $\boldsymbol{J}$ and $\boldsymbol{J}'$ are *i-variants*, and we write $\boldsymbol{J} =_{-i} \boldsymbol{J}'$, if $J_k = J'_k$ for all agents $k \neq i$ (and possibly $J_i \neq J'_i$ for agent i).

Intuitively, an *aggregation rule* is a function that maps any given profile to a single judgment representing the collective judgment of the group. In this paper, we restrict attention to aggregation rules that, for any given profile of complete and consistent judgments, will return a collective judgment that also is complete and consistent. As we saw in the introduction, the majority rule—which returns $m(\boldsymbol{J})$ for any given profile $\boldsymbol{J}$—does *not* meet this requirement. In practice, even for an odd number of agents most natural rules are *irresolute*—meaning that they allow for the possibility of ties between several collective judgments and thus require a tie-breaking mechanism to settle on a single outcome. So, formally, an aggregation rule is a function $F : \mathcal{J}(\Phi)^n \to 2^{\mathcal{J}(\Phi)} \setminus \{\emptyset\}$.

In this paper, we focus on *majority-preserving* rules. A rule F is majority-preserving if $F(\boldsymbol{J}) = \{m(\boldsymbol{J})\}$ for all profiles $\boldsymbol{J}$ such that $m(\boldsymbol{J})$ is consistent. Majority-preserving rules constitute the bulk of well-studied rules in judgment aggregation [26].

2.2 Induced Preferences

Since agents hold and submit judgments rather than rankings over possible outcomes, we cannot directly reason about their preferences and incentives. Still, following Dietrich and List [10], we will assume that an agent's preferences over outcomes are related to their truthfully held judgments and that we can glean at least some information about their preferences by extrapolating from those judgments. Specifically, we assume that an agent's most preferred outcome is their own truthful judgment. In many cases it makes sense to also assume that agents like outcomes less the further away they are from their true judgment, according to some notion of distance. We write $J \succeq_i J'$ ($J \succ_i J'$), to mean that agent i weakly (strictly) prefers judgment J to judgment J'.

An agent i with true judgment $J_i \in \mathcal{J}(\Phi)$ is said to have *closeness-respecting preferences* if $J \cap J_i \supseteq J' \cap J_i$ implies $J \succeq_i J'$ for all $J, J' \in \mathcal{J}(\Phi)$. We focus on a special case of closeness-respecting preferences based on the Hamming distance: agent i has *Hamming preferences* in case $J \succeq_i J'$ if and only if $H(J, J_i) \leq H(J', J_i)$. In this paper, we will only consider agents with Hamming preferences over judgments, unless otherwise stated.

Assuming that agents have Hamming preferences amounts to assuming that they care equally about every proposition in the agenda. This is a strong assumption that will not be justified in all circumstances, but in the absence of domain-specific information about preferences it is arguably the most natural way to proceed. Hamming preferences have indeed been the dominant choice in the literature on strategic behaviour in judgment aggregation to date [3]. They have also been used to analyse strategic manipulation of social welfare functions [2, 6].

Because the rules we examine are irresolute—meaning they do not always return a single collective judgment—we need to extend agent preferences over judgments to preferences over *sets of judgments* to reason about manipulation of these rules. As at least one of the sets in our comparisons will always turn out to be a singleton, we do not need to explicitly specify the agents' preferences beyond these cases. Let $\succeq_i$ be the (weak) preference order of agent i over judgments in $\mathcal{J}(\Phi)$. Then $\mathring{\succeq}_i$ (with strict part $\mathring{\succ}_i$) is the corresponding preference extension over sets of judgments. For all preference extensions $\mathring{\succeq}_i$ we assume that $a \succeq_i b$ implies $\{a\} \mathring{\succeq}_i \{b\}$.

For any A and B in $2^{\mathcal{J}(\Phi)} \setminus \{\emptyset\}$, where $B = \{b\}$ is a singleton, and $\succeq_i$ is a preference order over judgments, we define the following three classes of preference extensions:

- $\mathring{\succeq}_i$ is a *cautious* extension if the following holds:
  - ▷ $A \mathring{\succeq}_i \{b\}$ if for all $a \in A$ we have $a \succeq_i b$.
  - ▷ $\{b\} \mathring{\succeq}_i A$ if for all $a \in A$ we have $b \succeq_i a$.
- $\mathring{\succeq}_i$ is an *optimistic* extension if the following holds:
  - ▷ $A \mathring{\succeq}_i \{b\}$ if there exists some $a \in A$ such that $a \succeq_i b$
  - ▷ $\{b\} \mathring{\succeq}_i A$ if for all $a \in A$ we have $b \succeq_i a$.
- $\mathring{\succeq}_i$ is a *pessimistic* extension if the following holds:
  - ▷ $A \mathring{\succeq}_i \{b\}$ if for all $a \in A$ we have $a \succeq_i b$
  - ▷ $\{b\} \mathring{\succeq}_i A$ if there exists some $a \in A$ such that $b \succeq_i a$.

The preference extensions attributed to Kelly [23], Gärdenfors [19], and Fishburn [18] are all cautious extensions in this sense.

We say an agent i *is* cautious if $\mathring{\succeq}_i$ is a cautious preference extension. Similarly, an agent can be optimistic or pessimistic—we call this the *type* of each agent. Furthermore, we say an agent is *minimally cautious* if $\mathring{\succeq}_i$ is a cautious preference extension and neither $A \mathring{\succeq}_i \{b\}$ nor $\{b\} \mathring{\succeq}_i A$ hold for any A and b not covered by the conditions defining cautious extensions. Minimally optimistic and minimally pessimistic agents are defined analogously.

3 NOTIONS OF STRATEGYPROOFNESS

Recall that our objective in this paper is to identify attractive judgment aggregation rules that display a reasonable degree of immunity to manipulation by strategic agents. We do so against a backdrop of myriad well-known impossibility results across different areas of social choice theory [10, 20, 31, 38]: designing strategyproof aggregation rules is difficult and often impossible.

In this section we first recall the standard definition of strategyproofness and review a well-known result showing that designing rules of practical interest that are strategyproof in this sense is essentially impossible. We then introduce a new, less demanding, alternative, namely $\mathcal{D}$-strategyproofness for some given domain $\mathcal{D}$ of profiles, and argue that $\mathcal{M}$-strategyproofness, for the domain $\mathcal{M}$ of majority consistent profiles, is of particular interest.

3.1 Standard Strategyproofness

Let $\boldsymbol{J}$ be a profile such that J_i is agent i's truthful judgment, inducing her preference order $\succeq_i$ over judgments. Let $\mathring{\succeq}_i$ be i's preference order on sets of judgments. Then an irresolute aggregation rule F is *manipulable* by agent i in profile $\boldsymbol{J}$, if there exists a profile $\boldsymbol{J}' =_{-i} \boldsymbol{J}$ such that $F(\boldsymbol{J}') \mathring{\succ}_i F(\boldsymbol{J})$. An aggregation rule is *strategyproof* for a given type of agent if it is not manipulable by any agent of that type in any profile $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$.

The central result on strategyproofness in judgment aggregation is due to Dietrich and List [10]. It applies to *resolute* rules F (with $|F(\boldsymbol{J})| = 1$ for all profiles $\boldsymbol{J}$), meaning that the preference extension chosen plays no active role in the definition of strategyproofness.

Theorem 3.1 (Dietrich and List, 2007). *A resolute judgment aggregation rule F is strategyproof for all closeness-respecting preferences if and only if F is independent and monotonic.*

The axiom of independence requires that deciding whether F will accept φ is possible by only considering how the individual agents judge φ, while monotonicity requires that additional support for an accepted proposition φ never gets φ rejected.

Formally, F is independent and monotonic if and only if $N_\varphi^{\boldsymbol{J}} \subseteq N_\varphi^{\boldsymbol{J}'}$ implies $\varphi \in J \implies \varphi \in J'$ for $F(\boldsymbol{J}) = \{J\}$ and $F(\boldsymbol{J}') = \{J'\}$ [7]. Both axioms feature prominently in impossibility theorems, which essentially show that any rule that satisfies them is bound to return inconsistent outcomes for some profiles [12, 29, 34]. Indeed, among the standard aggregation rules, the only ones that satisfy both independence and monotonicity are the so-called *quota rules* [9], of which the majority rule is an example (quota rules accept a given proposition whenever a certain number of agents do). Although this class of rules can guarantee strategyproofness for a large family of preferences, they do not always return a consistent outcome and thus, arguably, are of little practical interest. This is why Theorem 3.1 must be interpreted as a negative result. Indeed, it suggests that there are no attractive rules that are strategyproof.

A first natural approach to overcoming this negative result is to restrict attention to strategyproofness for Hamming preferences only, rather than strategyproofness for *all* closeness-respecting preferences. But we will see in Section 4.2 that for the most well-known majority-preserving rules this also is not attainable.

3.2 Domain-Strategyproofness

Our approach is to introduce a weaker notion of strategyproofness, which we call *domain-strategyproofness*.

Consider an aggregation rule $F : \mathcal{J}(\Phi)^n \to 2^{\mathcal{J}(\Phi)} \setminus \{\emptyset\}$ and let $\mathcal{D} \subseteq \mathcal{J}(\Phi)^n$ be a subset of the set of admissible profiles. Let $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$ be a profile, with J_i being agent i's truthful judgment. Let $\succeq_i$ be agent i's preference order over judgments, and $\mathring{\succeq}_i$ her preference order over sets of judgments. We say that F is *$\mathcal{D}$-manipulable* by agent i in $\boldsymbol{J}$ if there exists another profile $\boldsymbol{J}' =_{-i} \boldsymbol{J}$ such that $F(\boldsymbol{J}') \mathrel{\mathring{\succ}_i} F(\boldsymbol{J})$ and at least one of $\boldsymbol{J}$ and $\boldsymbol{J}'$ belong to $\mathcal{D}$. If only $\boldsymbol{J}'$ belongs to $\mathcal{D}$, we say agent i can manipulate *to* $\mathcal{D}$. If only $\boldsymbol{J}$ belongs to $\mathcal{D}$, we say agent i can manipulate *from* $\mathcal{D}$.

Definition 3.2. *A rule is called $\mathcal{D}$-*strategyproof *for agents of a given type if it is not $\mathcal{D}$-manipulable by any agent $i \in N$ of that type in any profile $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$.*

This new notion of $\mathcal{D}$-strategyproofness is particularly useful when trying to improve upon aggregation rules that are known to be (fully) strategyproof but that can guarantee consistent outcomes only on a restricted domain $\mathcal{D}$ (as is the case for the majority rule). In such a case, a rule that is guaranteed to always return consistent outcomes and that is $\mathcal{D}$-strategyproof is an attractive alternative. Indeed, if a rule is strategyproof for $\mathcal{D}$, this tells us two things. First, if the truthful profile is in $\mathcal{D}$, then no agent has an incentive to manipulate. Second, if the profile that results after all judgments have been submitted is in $\mathcal{D}$, then we can be certain that the profile reported cannot have been the result of strategic manipulation. How does our notion of domain-strategyproofness relate to the use of *domain restrictions* in the judgment aggregation literature [11]? Domain restrictions have been a frequent source of positive results in social choice, starting with the seminal work of Black [4] and Sen [39]. They amount to restricting the input of an aggregation rule to a set of well-behaved profiles. Domain-strategyproofness similarly exploits the well-behavedness of a domain, but does so without restricting the actual input domain of the aggregation rule.

3.3 Majority-Strategyproofness

Let $\mathcal{M}(\Phi, n) \subseteq \mathcal{J}(\Phi)^n$ be the domain of all profiles for a given agenda and a given number of agents for which the majority outcome is consistent: $\mathcal{M}(\Phi, n) := \{\boldsymbol{J} \mid m(\boldsymbol{J}) \not\models \bot\}$. If Φ and n are clear from context, we simply write $\mathcal{M}$. The main notion of strategyproofness we will investigate in this paper is $\mathcal{M}$-strategyproofness.[1]

$\mathcal{M}$-strategyproofness, or *majority-strategyproofness*, of a majority-preserving rule guarantees that the majority outcome *will* in fact be preserved, even under the assumption that agents will manipulate if they have an incentive to do so. Such a rule would also guarantee that the number of manipulable profiles does not exceed the number of inconsistent outcomes given when using the (strategyproof) majority rule, as any manipulation will be between profiles where the majority rule would result in an inconsistent outcome. Thus, there is a sense in which $\mathcal{M}$-strategyproof rules will minimise the regret of the mechanism designer; if we—as the mechanism designer—care to a great extent about consistency and non-manipulability, it will never be preferable to use the majority rule over an $\mathcal{M}$-strategyproof majority-preserving rule that can guarantee consistency.

4 ADDITIVE MAJORITY RULES

In this section we prove that every judgment aggregation rule that belongs to the large family of *additive majority rules* is majority-strategyproof. This family includes some of the most important aggregation rules discussed in the literature, notably the Kemeny and the Slater rule. We first define and review this family of rules in some detail. We then show that its most prominent exponents are *not* fully strategyproof, before proving that nevertheless all rules in the family are majority-strategyproof.

4.1 Definition and Representative Rules

A judgment aggregation rule F is an *additive majority rule* (AMR)[2] if there exists a non-decreasing *gain function* $g : [0, n] \to \mathbb{R}$ with $g(k) < g(k')$ for any $k < \frac{n}{2}$ and $k' \geq \frac{n}{2}$ such that, for any profile $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$, the following condition is satisfied:

$$F(\boldsymbol{J}) \quad = \quad \underset{J \in \mathcal{J}(\Phi)}{\operatorname{argmax}} \sum_{\varphi \in J} g(n^{\boldsymbol{J}}_{\varphi})$$

Additive majority rules are based on the weighted majoritarian set, meaning that for each formula φ in the agenda the rule only looks at *how many* agents have φ in their judgment. Rules within this family differ only in how much they prioritise large majorities over small ones. Nehring and Pivato [32] call this the *elasticity* of the gain function. On one end of this spectrum lie rules for which the size of the majority does not play a large (or even any) role; on the other end, we find rules that prioritise large majorities over small ones. Observe that the requirement of $g(k) < g(k')$ for $k < \frac{n}{2}$ and $k' \geq \frac{n}{2}$ ensures that every AMR is majority-preserving.

The additive majority rules include three of the most studied majority-preserving rules in judgment aggregation. The first is the *Kemeny rule* F_{Kem}, defined by the simplest of gain functions:

$$g(x) \quad = \quad x$$

Thus, the Kemeny rule returns those consistent judgments that maximise a score computed as the number of times an individual agent agrees with the choice made for an individual proposition. Equivalently, we may think of the Kemeny rule as returning those judgments that minimise the average Hamming distance to the judgments in the profile. This rule generalises the well-known Kemeny rule for preference aggregation [24] and is also known under a number of other names, notably *distance-based rule* [35], *median rule* [33], and *prototype rule* [30].

The *Slater rule* F_{Sla} is defined by the following gain function:

$$g(x) \quad = \quad \begin{cases} 0 & \text{if } 0 \leq x < \frac{n}{2} \\ 1 & \text{if } \frac{n}{2} \leq x \leq n \end{cases}$$

Thus, F_{Sla} rule considers all formulas accepted by a majority of agents as equal, and tries to respect as many of these majorities as possible without violating consistency. In particular, it will not distinguish between a unanimously accepted formula and one accepted by just $\lceil \frac{n}{2} \rceil$ agents. F_{Sla} generalises the Slater rule from preference aggregation [40] and is also known under several other names, such as *endpoint rule* [30] and *maximum-cardinality subagenda rule* [25].

A third AMR of some prominence in the literature is the *Leximax rule* F_{Lex} [16, 32]. It gives maximal preference to stronger majorities, meaning that it orders the formulas in the agenda in terms of the number of agents supporting them and then tries to accept as many formulas supported by a given number of agents as possible before

[1] Note that a rule being majority-preserving does not guarantee $\mathcal{M}$-strategyproofness. For example, a rule that outputs the majority judgment if consistent and otherwise outputs a fixed judgment clearly is majority-preserving but not $\mathcal{M}$-strategyproof.

[2] The family of additive majority rules was first identified by Nehring and Pivato [32]. Here we have slightly adapted their original definition to our needs: on the one hand, we only consider rules that weight all formulas equally, and on the other, we consider a slightly larger family of gain functions g.

moving on to formulas with fewer supporters. It is a refinement of another popular rule, the Ranked Agenda rule (see Section 5.2). The Leximax rule is the AMR with the following gain function:

$$g(x) \quad = \quad |\Phi|^x$$

Leximax lands on the opposite side of the spectrum compared to Slater; while Slater does not distinguish at all between small majorities and large ones, F_{Lex} will never prioritise any number of small majorities over a single large one. For example, it will choose a single formula accepted by n agents, over $|\Phi| - 1$ formulas each accepted by $n - 1$ agents.

The class of additive majority rules includes many more rules of practical interest. Let us highlight two further examples, characterised by the following gain functions:

$$g(x) = \sum_{k=1}^{x} \frac{1}{k} \qquad\qquad g(x) = x \sum_{k=0}^{x} \epsilon^k \text{ for } \quad \epsilon \ll 0$$

The first rule falls somewhere between Slater and Kemeny in terms of elasticity; like Kemeny, it distinguishes between small and large majorities, but the "marginal returns" gained from additional support diminish as majorities grow larger. The second rule is very close to the Kemeny rule, but will prioritise large majorities slightly more. The rule can be seen as a way to break ties between Kemeny outcomes; it gives extra importance to larger majorities only insofar as this can be helpful in differentiating between outcomes that othewise would be considered equally appealing.

4.2 Failure of Full Strategyproofness

Theorem 3.1 excludes the possibility of Kemeny, Slater, or Leximax being strategyproof for all closeness-respecting preferences. It leaves open, however, the possibility that they are strategyproof for Hamming preferences. Indeed Kemeny and Slater, whose standard distance-based definitions are closely tied to the Hamming distance, seem to be promising candidates for rules that are strategyproof in this sense. We are now going to see that this is not the case, and that all three rules are manipulable on the full domain for a sufficiently large agenda.

Athanasoglou [2] shows for social welfare functions that both Kemeny and Slater are manipulable for all preference extensions, when the number of alternatives exceeds three. As any preference profile can be embedded into judgment aggregation [14], and as the outcomes of the Kemeny and Slater judgment aggregation rules will agree with their social welfare function counterparts in the preference aggregation domain, we obtain the following result.

Proposition 4.1 (Athanasoglou 2016). *The Kemeny rule and the Slater rule are manipulable for all preference extensions.*

We now show that the same holds for the Leximax rule.

Proposition 4.2. *The Leximax rule is manipulable for all preference extensions.*

Proof. Let $\boldsymbol{J}$ be the profile below, taken from recent work by Lang et al. [26], with $\Phi^+ = \{p \wedge r, p \wedge s, q, p \wedge q, t\}$ and 16 agents, including one distinguished agent i:

	$p \wedge r$	$p \wedge s$	q	$p \wedge q$	t
6 agents	Yes	Yes	Yes	Yes	Yes
7 agents	No	No	Yes	No	No
2 agents	Yes	Yes	No	No	Yes
J_i	Yes	Yes	No	No	Yes
Maj	Yes	Yes	Yes	No	Yes

We first note the support for the formulas in the agenda Φ:

$$n_q^{\boldsymbol{J}} = 13 \qquad n_{\neg(p \wedge q)}^{\boldsymbol{J}} = 10 \qquad n_{p\wedge r}^{\boldsymbol{J}} = n_{p \wedge s}^{\boldsymbol{J}} = n_t^{\boldsymbol{J}} = 9$$

It is clear then that $F_{\text{Lex}}(\boldsymbol{J}) = J = \{\neg(p \wedge r), \neg(p \wedge s), q, \neg(p \wedge q), t\}$. Let $\boldsymbol{J'}$ be an i-variant of $\boldsymbol{J}$ where $J_i' = \{p \wedge r, p \wedge s, q, p \wedge q, t\}$. Then:

$$n_{\neg(p\wedge q)}^{\boldsymbol{J'}} = n_{p\wedge r}^{\boldsymbol{J'}} = n_{p\wedge s}^{\boldsymbol{J'}} = n_t^{\boldsymbol{J'}} = 9$$

Rejecting $p \wedge q$ will therefore no longer maximise gain, and simple calculation tells us $F_{\text{Lex}}(\boldsymbol{J'}) = J_i'$. As agent i has Hamming preferences, we know $J_i' \succ_i J$, which implies $F_{\text{Lex}}(\boldsymbol{J'}) \mathrel{\mathring{\succ}_i} F_{\text{Lex}}(\boldsymbol{J})$. □

Thus, strategyproofness on the full domain is is too demanding a property. It is unattainable for the salient additive majority rules, even when we restrict attention to Hamming preferences and are free to choose any preference extension.

4.3 Guaranteed Majority-Strategyproofness

While we cannot guarantee strategyproofness on the full domain, it turns out that $\mathcal{M}$-strategyproofness is attainable for Hamming preferences and a large class of preference extensions.

Before presenting our main result, we prove three technical lemmas. The first establishes a relation between majority outcomes in two profiles that are i-variants, and the second links the notion of betweenness to the Hamming distance.

Lemma 4.3. *For profiles $\boldsymbol{J} =_{-i} \boldsymbol{J'}$, $m(\boldsymbol{J})$ is between J_i and $m(\boldsymbol{J'})$.*

Proof. As all judgments involved are complete and complement-free, we simply need to show $m(\boldsymbol{J}) \subseteq J_i \cup m(\boldsymbol{J'})$. Take any $\varphi \in m(\boldsymbol{J})$. Suppose $\varphi \notin J_i$. If $J_i' =_\varphi J_i$, then $N_\varphi^{\boldsymbol{J'}} = N_\varphi^{\boldsymbol{J}}$, so $\varphi \in m(\boldsymbol{J'})$. But if $J_i' \neq_\varphi J_i$, then $\varphi \in J_i'$ and $n_\varphi^{\boldsymbol{J'}} > n_\varphi^{\boldsymbol{J}}$, so again $\varphi \in m(\boldsymbol{J'})$. □

The following is implicit in the work of Duddy and Piggins [13], who prove the equivalent statement for preference orders. We give a proof for the sake of completeness.

Lemma 4.4. *If for complete and complement-free judgment sets J, J', J'', it is the case that J' is between J and J'', then we have that $H(J, J'') = H(J, J') + H(J', J'')$.*

Proof. By definition of betweenness, $J' \subseteq J \cup J''$. To see that

$$H(J', J) + H(J', J'') = |(J'' \setminus J \cup J \setminus J'') \cap J'|$$

note that for any $\varphi \in J'$, there are three cases we need to consider: either $\varphi \in J \setminus J'$; or $\varphi \in J' \setminus J$; or $\varphi \in J \cap J'$. If $\varphi \in J \cap J'$, this means that considering φ does not add to the Hamming distance from J' to J nor to the Hamming distance from J to J''. Thus we only need to consider the first two of three possible cases in order to find the sum of the two Hamming distances. In other words, we can simply count the number of times J and J'' disagree on formulas in J'.

Since $H(J', J) + H(J', J'')$ is the Hamming distance between J and J'' restricted only to the formulas present in J', this distance cannot exceed $H(J, J'')$, meaning it must be the case that $H(J, J') + H(J', J'') \leq H(J, J'')$. This together with the triangle inequality $H(J, J'') \leq H(J, J') + H(J', J'')$ proves the claim. $\square$

Our final lemma establishes a relationship between majority outcomes and the outcomes of an AMR, in terms of the Hamming distance. By definition, the Slater rule satisfies the property in Lemma 4.5. We show that the same is true for any AMR when restricting our scope to i-variants. This will be useful for proving $\mathcal{M}$-strategyproofness for the class as a whole.

Lemma 4.5. *Let F be an additive majority rule and let $\boldsymbol{J}$ and $\boldsymbol{J'}$ be two profiles such that $\boldsymbol{J} =_{-i} \boldsymbol{J'}$ for some agent i, and such that $m(\boldsymbol{J'})$ is consistent. Then $H(m(\boldsymbol{J}), m(\boldsymbol{J'})) \geq H(m(\boldsymbol{J}), J^*)$ for all $J^* \in F(\boldsymbol{J})$.*

Proof. Let g be the non-decreasing gain function defining F and fix an arbitrary judgment set $J^* \in F(\boldsymbol{J})$. Let $k = H(m(\boldsymbol{J}), m(\boldsymbol{J'}))$ and $k' = H(m(\boldsymbol{J}), J^*)$. So we need to show that $k \geq k'$.

We first derive a constraint on k. Observe that agent i can change the majority outcome for a formula φ under profile $\boldsymbol{J}$ only in case $n_\varphi^{\boldsymbol{J}}$ is equal to either $\lfloor \frac{n}{2} \rfloor$ or $\lceil \frac{n}{2} \rceil$. With this in mind, we can write the total gain for formulas $\varphi \in m(\boldsymbol{J'})$ under profile $\boldsymbol{J}$ as follows:

$$\begin{aligned}
&\sum_{\varphi \in m(\boldsymbol{J'})} g(n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) + \sum_{\varphi \in m(\boldsymbol{J'}) \setminus m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) - \sum_{\varphi \in m(\boldsymbol{J}) \setminus m(\boldsymbol{J'})} g(n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) + k \cdot g(\lfloor \tfrac{n}{2} \rfloor) - k \cdot g(\lceil \tfrac{n}{2} \rceil)
\end{aligned}$$

Next, we derive a similar constraint on k'. Let us compute the total gain for formulas $\varphi \in J^*$ under the same profile $\boldsymbol{J}$:

$$\begin{aligned}
&\sum_{\varphi \in J^*} g(n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) + \sum_{\varphi \in J^* \setminus m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) - \sum_{\varphi \in m(\boldsymbol{J}) \setminus J^*} g(n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) + \sum_{\varphi \in J^* \setminus m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) - \sum_{\varphi \in J^* \setminus m(\boldsymbol{J})} g(n - n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) + \sum_{\varphi \in J^* \setminus m(\boldsymbol{J})} \left[g(n_\varphi^{\boldsymbol{J}}) - g(n - n_\varphi^{\boldsymbol{J}}) \right]
\end{aligned}$$

As g is a non-decreasing function, $g(n_\varphi^{\boldsymbol{J}}) - g(n - n_\varphi^{\boldsymbol{J}})$ is non-decreasing in $n_\varphi^{\boldsymbol{J}}$. Hence, given that the maximal value that $n_\varphi^{\boldsymbol{J}}$ can take for any $\varphi \notin m(\boldsymbol{J})$—and thus for any $\varphi \in J^* \setminus m(\boldsymbol{J})$—is $\lfloor \frac{n}{2} \rfloor$, the last sum in the equation above is at most equal to $k' \cdot \left[g(\lfloor \frac{n}{2} \rfloor) - g(n - \lfloor \frac{n}{2} \rfloor) \right] = k' \cdot \left[g(\lfloor \frac{n}{2} \rfloor) - g(\lceil \frac{n}{2} \rceil) \right]$. So we obtain:

$$\sum_{\varphi \in J^*} g(n_\varphi^{\boldsymbol{J}}) \leq \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) + k' \cdot \left[g(\lfloor \tfrac{n}{2} \rfloor) - g(\lceil \tfrac{n}{2} \rceil) \right]$$

Finally, let us combine the constraints on k and k' that we have derived. Recall that, by assumption, $m(\boldsymbol{J'})$ is a consistent judgment set. So it is available as a potential outcome under profile $\boldsymbol{J}$. Thus, the score of J^*, one of the *actual* outcomes under $\boldsymbol{J}$, must be at least as high as that of $m(\boldsymbol{J'})$:

$$\sum_{\varphi \in J^*} g(n_\varphi^{\boldsymbol{J}}) \geq \sum_{\varphi \in m(\boldsymbol{J'})} g(n_\varphi^{\boldsymbol{J}})$$

Putting everything together, and keeping in mind that $g(\lfloor \frac{n}{2} \rfloor) - g(\lceil \frac{n}{2} \rceil) < 0$, we obtain $k \geq k'$ as claimed. $\square$

We can now combine the three lemmas to get our main result.

Theorem 4.6. *Additive majority rules are $\mathcal{M}$-strategyproof for cautious, optimistic, and pessimistic agents.*

Proof. Let F be the AMR defined by the non-decreasing gain function g, and let $\boldsymbol{J}$ and $\boldsymbol{J'}$ be two profiles such that $\boldsymbol{J} =_{-i} \boldsymbol{J'}$ for some agent i, and J_i is agent i's truthful opinion. We need to show that, if $m(\boldsymbol{J})$ or $m(\boldsymbol{J'})$ is consistent, then it must be the case that $F(\boldsymbol{J}) \mathrel{\mathring{\succeq}_i} F(\boldsymbol{J'})$ whenever agent i is cautious, pessimistic, or optimistic. From Lemmas 4.3 and 4.4 together, we obtain:

$$H(J_i, m(\boldsymbol{J'})) = H(J_i, m(\boldsymbol{J})) + H(m(\boldsymbol{J}), m(\boldsymbol{J'})) \tag{i}$$

Note that if both $m(\boldsymbol{J})$ and $m(\boldsymbol{J'})$ are consistent, then as F is majority-preserving, $F(\boldsymbol{J}) = \{m(\boldsymbol{J})\}$ and $F(\boldsymbol{J'}) = \{m(\boldsymbol{J'})\}$. Any possible manipulation between these profiles would therefore imply a possible manipulation of the majority rule. However, Theorem 3.1 tells us no manipulation of the majority rule is possible. Thus, we need only consider the following two cases.

Case 1: For inconsistent $m(\boldsymbol{J})$ and consistent $m(\boldsymbol{J'})$, Lemma 4.5 says that for any outcome $J^* \in F(\boldsymbol{J})$, it is the case that $H(m(\boldsymbol{J}), J^*) \leq H(m(\boldsymbol{J}), m(\boldsymbol{J'}))$. We need to show that $H(J_i, J^*) \leq H(J_i, m(\boldsymbol{J'}))$.

Take an arbitrary judgment set $J^* \in F(\boldsymbol{J})$. Combining the triangle inequality with Lemma 4.5 and (i), we get (ii):

$$\begin{aligned}
H(J_i, J^*) &\leq H(J_i, m(\boldsymbol{J})) + H(m(\boldsymbol{J}), J^*) \\
&\leq H(J_i, m(\boldsymbol{J})) + H(m(\boldsymbol{J}), m(\boldsymbol{J'})) \\
&= H(J_i, m(\boldsymbol{J'}))
\end{aligned} \tag{ii}$$

In other words, for any $J^* \in F(\boldsymbol{J})$ and the unique $J' = m(\boldsymbol{J'}) \in F(\boldsymbol{J'})$, we have that $J^* \succeq_i J'$. So if agent i is cautious, optimistic, or pessimistic, then $F(\boldsymbol{J}) \mathrel{\mathring{\succeq}_i} F(\boldsymbol{J'})$ as required.

Case 2: For consistent $m(\boldsymbol{J})$ and inconsistent $m(\boldsymbol{J'})$, we know by Lemma 4.5 that $H(m(\boldsymbol{J'}), J^*) \leq H(m(\boldsymbol{J}), m(\boldsymbol{J'}))$ for any $J^* \in F(\boldsymbol{J'})$. We now need to show that $H(J_i, J^*) \geq H(J_i, m(\boldsymbol{J}))$.

Take an arbitrary judgment set $J^* \in F(\boldsymbol{J'})$. We again use the triangle inequality, Lemma 4.5, and (i) to get (iii):

$$\begin{aligned}
H(J_i, J^*) &\geq H(J_i, m(\boldsymbol{J'})) - H(m(\boldsymbol{J'}), J^*) \\
&\geq H(J_i, m(\boldsymbol{J'})) - H(m(\boldsymbol{J}), m(\boldsymbol{J'})) \\
&= H(J_i, m(\boldsymbol{J}))
\end{aligned} \tag{iii}$$

In other words, for any $J^* \in F(\boldsymbol{J'})$ and the unique $J = m(\boldsymbol{J}) \in F(\boldsymbol{J})$, we have that $J \succeq_i J^*$. Again, if agent i has cautious, optimistic, or pessimistic preferences, then we get $F(\boldsymbol{J}) \mathrel{\mathring{\succeq}_i} F(\boldsymbol{J'})$. $\square$

Inspection of our proof shows that $\mathcal{M}$-strategyproofness is guaranteed for every AMR under any preference extension for which, first, $a \succ_i b$ for all $a \in A$ implies $A \mathrel{\mathring{\succ}_i} \{b\}$ and, second, $a \succ_i b$ for all $b \in B$ implies $\{a\} \mathrel{\mathring{\succ}_i} B$. The cautious, optimistic, and pessimistic preference extensions mentioned in the statement of the theorem are particularly natural exponents of this class of extensions.

Corollary 4.7. *The Kemeny, Slater, and Leximax rules are $\mathcal{M}$-strategyproof for cautious, optimistic, and pessimistic agents.*

Let us briefly review how our results relate to known domain restrictions that guartantee a consistent majority outcome, the most prominent example of which is *unidimensional alignment* [28]. Let $\mathcal{U}(\Phi, n)$ be the domain of unidimensionally aligned profiles for Φ and n. As $\mathcal{U}(\Phi, n) \subseteq \mathcal{M}(\Phi, n)$ [28], we immediately obtain:

COROLLARY 4.8. *Additive majority rules are $\mathcal{U}$-strategyproof for cautious, optimistic, and pessimistic agents.*

Clearly, this holds for any domain restriction in judgment aggregation that guarantees a consistent majority.

The majority-strategyproofness of additive majority rules presents a strong argument for their use *in lieu* of the majority rule. They offer an alternative that guarantees consistency, and ensures that the majority will be preserved in all cases. Importantly they also offer a post-aggregation "check" for majority consistent outcomes, meaning it is possible to recognise cases where no manipulation can have occurred, thereby ensuring we can trust the outcome.

5 FURTHER AGGREGATION RULES

In this section we first examine two rules, the Maximal Condorcet rule and the Ranked Agenda rule, that are related to the additive majority rules in that they will always return a superset of the outcome of some AMR. It turns out that this particular relationship affords these rules a certain level of protection against manipulation. We also present an example of a majority-preserving rule, the Dodgson rule, that is highly susceptible to manipulation.

5.1 The Maximal Condorcet Rule

For a set of formulas $S \subseteq \Phi$, a set $S' \subseteq S$ is a *maximally consistent* subset of S if and only if (*i*) S' is consistent and (*ii*) there is no consistent set S'' such that $S' \subset S'' \subseteq S$. Let $S^+ = \{J \mid J \in \mathcal{J}(\Phi) \text{ and } J \supseteq S\}$. Let $C(J)$ denote the set of all maximally consistent subsets of the judgment J. The *Maximal Condorcet rule* is defined as follows:

$$F_{\mathrm{MC}}(\boldsymbol{J}) \quad = \quad \{J^+ \mid J \in C(m(\boldsymbol{J}))\}$$

F_{MC} is also known as the (rule returning the) *Condorcet (admissible) set* [33] and the *maximal sub-agenda rule* [27].

Observe that F_{Sla} is a refinement of F_{MC} in that $F_{\mathrm{Sla}}(\boldsymbol{J}) \subseteq F_{\mathrm{MC}}(\boldsymbol{J})$ for all profiles $\boldsymbol{J}$. This is clear from the standard definition of F_{Sla} as the rule that selects the maximal consistent subset of the majority in terms of cardinality. The proximity of Maximal Condorcet to the additive majority rules means that it retains some level of immunity to manipulation. We are going to show that for minimally cautious agents, F_{MC} is $\mathcal{M}$-strategyproof. We first state some weaker strategyproofness results for minimally pessimistic and minimally optimistic agents, although we show F_{MC} is still manipulable to and from $\mathcal{M}$ for pessimistic and optimistic agents, respectively.

EXAMPLE 5.1 (MANIPULATION *TO* MAJORITY). Let $\boldsymbol{J}$ be the profile below, where J_1 is agent 1's truthful opinion, and $\boldsymbol{J}'$ is a 1-variant where $J_1' = \{p, \neg q, \neg(p \wedge q), p \wedge r\}$.

	p	q	$p \wedge q$	$p \wedge r$
J_1	Yes	Yes	Yes	Yes
J_2	Yes	No	No	Yes
J_3	No	Yes	No	No
Maj	Yes	Yes	No	Yes

We can see that $F_{\mathrm{MC}}(\boldsymbol{J}) = \{\{p, q, p \wedge q, p \wedge r\}, \{p, \neg q, \neg(p \wedge q), p \wedge r\}, \{\neg p, q, \neg(p \wedge q), \neg(p \wedge r)\}\}$, and since $m(\boldsymbol{J}')$ is consistent, $F_{\mathrm{MC}}(\boldsymbol{J}') = \{\{p, \neg q, \neg(p \wedge q), p \wedge r\}\}$. As $\{p, \neg q, \neg(p \wedge q), p \wedge r\} \succ_1 \{\neg p, q, \neg(p \wedge q), \neg(p \wedge r)\}$, agent 1 can successfully manipulate from from $\boldsymbol{J}$ to $\boldsymbol{J}'$—meaning *to* the majority—if she is pessimistic. △

Due to the aforementioned relationship between Maximal Condorcet and Slater, we get the following result.

PROPOSITION 5.1. *A minimally pessimistic agent cannot manipulate the Maximal Condorcet rule from majority.*

PROOF. Let $\boldsymbol{J}$ and $\boldsymbol{J}'$ be two profiles such that $\boldsymbol{J} =_{-i} \boldsymbol{J}'$ and $F_{\mathrm{MC}}(\boldsymbol{J}) = \{m(\boldsymbol{J})\}$. Suppose for contradiction that there is a minimally pessimistic agent i, with truthful opinion J_i, who can manipulate from $\boldsymbol{J}$ to $\boldsymbol{J}'$. Then $J' \succ_i m(\boldsymbol{J})$ for all $J' \in F_{\mathrm{MC}}(\boldsymbol{J}')$. As $F_{\mathrm{Sla}}(\boldsymbol{J}') \subseteq F_{\mathrm{MC}}(\boldsymbol{J}')$, this would constitute a successful manipulation of Slater by a pessimistic agent, which contradicts Corollary 4.7. □

EXAMPLE 5.2 (MANIPULATION *FROM* MAJORITY). Let $\boldsymbol{J}$ be the profile below, and suppose J_1 is agent 1's truthful opinion. Let $\boldsymbol{J}' =_{-1} \boldsymbol{J}$ be the profile which differs only in that agent 1 submits $J_1' = \{a, b, c, \neg d, (a \wedge \neg d) \rightarrow (b \wedge c)\}$.

	p	q	r	s	$(p \wedge \neg s) \rightarrow (q \wedge r)$
J_1	Yes	Yes	Yes	Yes	Yes
J_2	Yes	No	No	Yes	Yes
J_3	No	No	No	No	Yes
Maj	Yes	No	No	Yes	Yes

As $m(\boldsymbol{J})$ is consistent, $F_{\mathrm{MC}}(\boldsymbol{J}) = \{m(\boldsymbol{J})\}$. For $\boldsymbol{J}'$, the majority, $m(\boldsymbol{J}') = \{p, \neg q, \neg r, \neg s, (p \wedge \neg s) \rightarrow (q \wedge r)\}$, is not consistent. It is simple to confirm $\{p, \neg s, (p \wedge \neg s) \rightarrow (q \wedge r)\} \in C(m(\boldsymbol{J}'))$, and thus that $J^* = \{p, q, r, \neg s, (p \wedge \neg s) \rightarrow (q \wedge r)\} \in F_{\mathrm{MC}}(\boldsymbol{J}')$. We calculate the distances from J_1 to find that $J^* \succ_1 m(\boldsymbol{J})$. As there exists some strictly better outcome in $F_{\mathrm{MC}}(\boldsymbol{J}')$, agent 1 can manipulate Maximal Condorcet *from* majority if she is an optimistic agent. △

PROPOSITION 5.2. *A minimally optimistic agent cannot manipulate the Maximal Condorcet rule to majority.*

PROOF. Let $\boldsymbol{J}$ and $\boldsymbol{J}'$ be two profiles such that $\boldsymbol{J} =_{-i} \boldsymbol{J}'$ and $F_{\mathrm{MC}}(\boldsymbol{J}') = \{m(\boldsymbol{J}')\}$. Suppose for contradiction that there is a minimally optimistic agent i, with truthful opinion J_i, who can manipulate from $\boldsymbol{J}$ to $\boldsymbol{J}'$. Then $m(\boldsymbol{J}') \succ_i J^*$ for all $J^* \in F_{\mathrm{MC}}(\boldsymbol{J})$. As $F_{\mathrm{Sla}}(\boldsymbol{J}) \subseteq F_{\mathrm{MC}}(\boldsymbol{J})$, this would constitute a successful manipulation of Slater by an optimistic agent, which contradicts Corollary 4.7. □

PROPOSITION 5.3. *The Maximal Condorcet rule is $\mathcal{M}$-strategyproof for minimally cautious agents.*

PROOF. By definition, if a minimally optimistic (pessimistic) agent cannot manipulate a rule to (from) the majority, then a minimally cautious agent cannot either. This, together with Proposition 5.2, shows that minimally cautious agents cannot manipulate F_{MC} *to* (*from*) majority. This establishes $\mathcal{M}$-strategyproofness of Maximal Condorcet for minimally cautious agents. □

Thus, while a pessimistic or optimistic agent might manipulate the Maximal Condorcet rule, the rule benefits from its relationship with Slater in terms of $\mathcal{M}$-strategyproofness for cautious agents.

Note however that, while Slater is $\mathcal{M}$-strategyproof for *any* cautious agent, Maximal Condorcet provides the same protection only against those cautious agents who are minimally cautious.

5.2 The Ranked Agenda Rule

The *Ranked Agenda rule* F_{RA} is a generalisation of the Ranked Pairs voting rule [42]. We do not explicitly define the this rule here, but refer to Lang et al. [26] for a precise definition. It is similar to the Leximax rule in that it prioritises large majorities over small ones, but it does not break ties by "looking ahead" to maximise gain as Leximax does. While F_{RA} is not itself an AMR, we have $F_{\mathrm{Lex}}(\boldsymbol{J}) \subseteq F_{\mathrm{RA}}(\boldsymbol{J})$ for all profiles $\boldsymbol{J}$ [26]. Exploiting this connection to an AMR we have shown to be $\mathcal{M}$-strategyproof before we obtain the following results (using the same approach as in Section 5.1).

Proposition 5.4. *A minimally pessimistic agent cannot manipulate the Ranked Agenda rule from majority.*

Proposition 5.5. *A minimally optimistic agent cannot manipulate the Ranked Agenda rule to majority.*

Proposition 5.6. *The Ranked Agenda rule is $\mathcal{M}$-strategyproof for minimally cautious agents.*

5.3 The Dodgson Rule

We conclude our examination by straying even further afield from the additive majority rules. In order to define the next rule, we first define the Hamming distance *between two profiles* as $H_P(\boldsymbol{J}, \boldsymbol{J}') := \sum_{i \in N} H(J_i, J_i')$. The *Dodgson rule* (for odd n) is defined as follows:

$$F_{\mathrm{Dod}}(\boldsymbol{J}) \quad = \quad \{\, m(\boldsymbol{J}') \mid \operatorname*{argmin}_{\boldsymbol{J}' \in \mathcal{M}(\Phi, n)} H_P(\boldsymbol{J}, \boldsymbol{J}') \,\}$$

This rule is also known as the *minimal-profile-change rule* [26] and as the *"full" distance-based rule* [30]. F_{Dod} chooses those judgments that can be reached by making the smallest number of atomic changes to the profile, where an atomic change consists in changing the judgment of a single agent on a single formula. This is clearly a majority-preserving rule, but it is not an AMR. Indeed, it also lacks the strategyproofness properties of the previous majority-preserving rules examined in this paper.

Proposition 5.7. *Dodgson fails $\mathcal{M}$-strategyproofness for all preference extensions.*

Proof. Let Φ be an agenda with $|\Phi^+| = 10$. Consider the profile $\boldsymbol{J}$ below, with J_1 being agent 1's true judgment:

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7	φ_8	φ_9	φ_{10}
J_1	No	Yes	Yes	No	No	No	No	No	No	No
J_2	No	No	No	Yes	No	Yes	Yes	No	Yes	Yes
J_3	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No
Maj	No	No	No	No	No	Yes	Yes	No	Yes	No

Suppose that—besides J_1, J_2, and J_3 appearing $\boldsymbol{J}$—the only other judgments that are consistent are J_4, J_5, J_6, and J_7 shown below.[3]

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7	φ_8	φ_9	φ_{10}
J_4	No	No	No	No	No	Yes	Yes	No	Yes	No
J_5	Yes	Yes	No	Yes	No	No	No	No	No	No
J_6	No	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes
J_7	No	Yes	No	Yes	No	Yes	Yes	No	No	No

As the majority outcome is consistent, $F_{\mathrm{Dod}}(\boldsymbol{J}) = \{m(\boldsymbol{J})\} = \{J_4\}$.

Let $\boldsymbol{J}'$ be an i-variant of $\boldsymbol{J}$ with $J_1' = J_5$, making $m(\boldsymbol{J}')$ inconsistent. We see that the minimal number of atomic changes we can make to the profile $\boldsymbol{J}'$—while ensuring all input judgments are consistent—is 1, as $H(J_2, J_6) = 1$. For all other relevant pairwise comparisons of admissible judgments, the Hamming distance between them is 2 or greater. Indeed, replacing J_2 with J_6 will result in profile $\boldsymbol{J}^* = (J_5, J_6, J_3)$, with a consistent majority outcome. Thus $F_{\mathrm{Dod}}(\boldsymbol{J}') = \{m(\boldsymbol{J}^*)\} = \{J_7\}$. As $J_7 \succ_1 J_4$, it must be the case that $F(\boldsymbol{J}') \mathrel{\mathring{\succ}}_1 F(\boldsymbol{J})$, making this a successful manipulation from $\mathcal{M}$. □

For cautious (and pessimistic) agents, the following example shows manipulation is possible both to and from $\mathcal{M}$. Thus, for this type of agent, Dodgson will also fail to provide the post-aggregation guarantee that no manipulation has occurred.

Example 5.3. Let $\boldsymbol{J}$ be the profile below, where J_1 is agent 1's true judgment, and suppose she is cautious. Let Φ be an agenda such that J_1, J_2, and J_3 are the only consistent judgments.

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6
J_1	No	No	No	Yes	No	No
J_2	No	No	Yes	No	Yes	Yes
J_3	Yes	Yes	Yes	Yes	Yes	Yes
Maj	No	No	Yes	Yes	Yes	Yes

Note that the majority outcome is not consistent. It is easy to check that $F(\boldsymbol{J}) = \{J_2, J_3\}$. Now let $\boldsymbol{J}'$ be an i-variant of $\boldsymbol{J}$, where $J_1' = J_2$. Then $F_{\mathrm{Dod}}(\boldsymbol{J}') = m(\boldsymbol{J}') = J_2$, as agent 1 prefers J_2 over J_3. As she is a cautious agent, we have $F_{\mathrm{Dod}}(\boldsymbol{J}') \mathrel{\mathring{\succ}}_1 F_{\mathrm{Dod}}(\boldsymbol{J})$ which is a successful manipulation *to* $\mathcal{M}$. △

The case of the Dodgson rule thus presents a clear example showing that by no means all majority-preserving rules are associated with some level of $\mathcal{M}$-strategyproofness.

6 CONCLUSION

We have introduced a novel weakening of strategyproofness, which we called domain-strategyproofness. We have argued that in the absence of full strategyproofness, domain-strategyproofness often offers a sufficiently strong barrier against manipulation. We have focused in particular on the majority-consistent domain, and examined majority-preserving aggregation rules, showing varying levels of strategyproofness for several prominent rules from the judgment aggregation literature. Our results make a strong case for the use of additive majority rules, a class of rules that includes both the Kemeny rule and the Slater rule.

As strategyproof rules are hard to come by in social choice in general, we have argued that domain-strategyproofness offers an attractive way out of this dilemma. While our results in judgment aggregation also hold for social welfare functions in preference aggregation, it still remains to be seen whether similar results can be obtained for Condorcet extensions in voting—an arena where finding attractive strategyproof rules is similarly challenging.

[3] We note that, by a result of Dokow and Holzman [12], it is possible to construct an agenda with these structural properties (and we can, conveniently, abstract away from the specifics of Φ).

REFERENCES

[1] Kenneth Arrow. 1950. A Difficulty in the Concept of Social Welfare. *Journal of Political Economy* 58, 4 (1950), 328–46.

[2] Stergios Athanasoglou. 2016. Strategyproof and Efficient Preference Aggregation with Kemeny-based Criteria. *Games and Economic Behavior* 95 (2016), 156–167.

[3] Dorothea Baumeister, Jörg Rothe, and Ann-Kathrin Selker. 2017. Strategic Behavior in Judgment Aggregation. In *Trends in Computational Social Choice*, Ulle Endriss (Ed.). AI Access, Chapter 8, 145–168.

[4] Duncan Black. 1948. On the Rationale of Group Decision-making. *Journal of Political Economy* 56, 1 (1948), 23–34.

[5] Walter Bossert and Yves Sprumont. 2014. Strategy-proof Preference Aggregation: Possibilities and Characterizations. *Games and Economic Behavior* 85 (2014), 109–126.

[6] Walter Bossert and Ton Storcken. 1992. Strategy-proofness of Social Welfare Functions: The Use of the Kemeny Distance between Preference Orderings. *Social Choice and Welfare* 9, 4 (1992), 345–360.

[7] Sirin Botan, Arianna Novaro, and Ulle Endriss. 2016. Group Manipulation in Judgment Aggregation. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2016)*. IFAAMAS.

[8] Franz Dietrich and Christian List. 2007. Arrow's Theorem in Judgment Aggregation. *Social Choice and Welfare* 29, 1 (2007), 19–33.

[9] Franz Dietrich and Christian List. 2007. Judgment Aggregation by Quota Rules: Majority Voting Generalized. *Journal of Theoretical Politics* 19, 4 (2007), 391–424.

[10] Franz Dietrich and Christian List. 2007. Strategy-proof Judgment Aggregation. *Economics & Philosophy* 23, 3 (2007), 269–300.

[11] Franz Dietrich and Christian List. 2010. Majority Voting on Restricted Domains. *Journal of Economic Theory* 145, 2 (2010), 512–543.

[12] Elad Dokow and Ron Holzman. 2010. Aggregation of Binary Evaluations. *Journal of Economic Theory* 145, 2 (2010), 495–511.

[13] Conal Duddy and Ashley Piggins. 2012. A Measure of Distance Between Judgment Sets. *Social Choice and Welfare* 39, 4 (2012), 855–867.

[14] Ulle Endriss. 2016. Judgment Aggregation. In *Handbook of Computational Social Choice*, F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia (Eds.). Cambridge University Press.

[15] Ulle Endriss, Umberto Grandi, and Daniele Porello. 2012. Complexity of Judgment Aggregation. *Journal of Artificial Intelligence Research* 45 (2012), 481–514.

[16] Patricia Everaere, Sébastien Konieczny, and Pierre Marquis. 2014. Counting Votes for Aggregating Judgments. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2014)*. IFAAMAS, 1177–1184.

[17] Patricia Everaere, Sébastien Konieczny, and Pierre Marquis. 2017. An Introduction to Belief Merging and its Links with Judgment Aggregation. In *Trends in Computational Social Choice*, Ulle Endriss (Ed.). AI Access, Chapter 7, 123–143.

[18] Peter C. Fishburn. 1972. Even-chance Lotteries in Social Choice Theory. *Theory and Decision* 3, 1 (1972), 18–40.

[19] Peter Gärdenfors. 1976. Manipulation of Social Choice Functions. *Journal of Economic Theory* 13, 2 (1976), 217–228.

[20] Allan Gibbard. 1973. Manipulation of Voting Schemes: A General Result. *Econometrica* 41, 4 (1973), 587.

[21] Davide Grossi and Gabriella Pigozzi. 2014. *Judgment Aggregation: A Primer*. Morgan & Claypool Publishers.

[22] Ronald de Haan. 2017. Complexity Results for Manipulation, Bribery and Control of the Kemeny Judgment Aggregation Procedure. In *Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems (AAMAS-2017)*. IFAAMAS.

[23] Jerry S. Kelly. 1977. Strategy-proofness and Social Choice Functions without Singlevaluedness. *Econometrica: Journal of the Econometric Society* (1977), 439–446.

[24] John G. Kemeny. 1959. Mathematics without Numbers. *Daedalus* 88, 4 (1959), 577–591.

[25] Jérôme Lang, 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 (TARK-2011)*. ACM.

[26] Jérôme Lang, Gabriella Pigozzi, Marija Slavkovik, Leendert van der Torre, and Srdjan Vesic. 2017. A Partial Taxonomy of Judgment Aggregation Rules and Their Properties. *Social Choice and Welfare* 48, 2 (2017), 327–356.

[27] Jérôme Lang and Marija Slavkovik. 2014. How Hard is it to Compute Majority-preserving Judgment Aggregation Rules? In *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI-2014)*. IOS Press, 501–506.

[28] Christian List. 2003. A Possibility Theorem on Aggregation over Multiple Interconnected Propositions. *Mathematical Social Sciences* 45 (2003), 1–13.

[29] Christian List and Philip Pettit. 2002. Aggregating Sets of Judgments: An Impossibility Result. *Economics and Philosophy* 18, 1 (2002), 89–110.

[30] Michael K. Miller and Daniel Osherson. 2009. Methods for Distance-based Judgment Aggregation. *Social Choice and Welfare* 32, 4 (2009), 575–601.

[31] Eitan Muller and Mark Allen Satterthwaite. 1977. The Equivalence of Strong Positive Association and Strategy-proofness. *Journal of Economic Theory* 14, 2 (1977), 412–418.

[32] Klaus Nehring and Marcus Pivato. 2019. Majority Rule in the Absence of a Majority. *Journal of Economic Theory* 183 (2019), 213–257.

[33] Klaus Nehring, Marcus Pivato, and Clemens Puppe. 2014. The Condorcet Set: Majority Voting over Interconnected Propositions. *Journal of Economic Theory* 151 (2014), 268–303.

[34] Klaus Nehring and Clemens Puppe. 2007. The Structure of Strategy-proof Social Choice – Part I: General Characterization and Possibility Results on Median Spaces. *Journal of Economic Theory* 135, 1 (2007), 269–305.

[35] Gabriella Pigozzi. 2006. Belief Merging and the Discursive Dilemma: An Argument-based Account to Paradoxes of Judgment Aggregation. *Synthese* 152, 2 (2006), 285–298.

[36] Shin Sato. 2013. A Sufficient Condition for the Equivalence of Strategy-proofness and Nonmanipulability by Preferences Adjacent to the Sincere one. *Journal of Economic Theory* 148, 1 (2013), 259–278.

[37] Shin Sato. 2015. Bounded Response and the Equivalence of Nonmanipulability and Independence of Irrelevant Alternatives. *Social Choice and Welfare* 44 (2015), 133–149.

[38] Mark Allen Satterthwaite. 1975. Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory* 10, 2 (1975), 187–217.

[39] Amartya K. Sen. 1966. A Possibility Theorem on Majority Decisions. *Econometrica: Journal of the Econometric Society* (1966), 491–499.

[40] Patrick Slater. 1961. Inconsistencies in a Schedule of Paired Comparisons. *Biometrika* 48, 3–4 (1961), 303–312.

[41] Zoi Terzopoulou and Ulle Endriss. 2019. Strategyproof Judgment Aggregation under Partial Information. *Social Choice and Welfare* 53, 3 (2019), 415–442.

[42] Thorwald Nicolaus Tideman. 1987. Independence of Clones as a Criterion for Voting Rules. *Social Choice and Welfare* 4, 3 (1987), 185–206.