The impossibility of non-manipulable probability aggregation

Franz Dietrich
CNRS & Paris School of Economics

Christian List
LMU Munich

Discussion paper, 27 December 2023

Abstract

A probability aggregation rule assigns to each profile of probability functions across a group of individuals (representing their individual probability assignments to some propositions) a collective probability function (representing the group's probability assignment). The rule is "non-manipulable" if no group member can manipulate the collective probability for any proposition in the direction of his or her own probability by misrepresenting his or her probability function ("strategic voting"). We show that, except in trivial cases, no probability aggregation rule satisfying two mild conditions (non-dictatorship and consensus preservation) is non-manipulable.

1 Introduction

It is widely recognized that probability aggregation – aggregating a profile of probability functions across a group of individuals into a single probability function for the group as a whole – is immune to some of the well-known social-choice-theoretic pathologies that bedevil both the aggregation of preferences and the aggregation of "true/false" judgments. When preferences are aggregated across a group of voters, legislators etc., the majority preferences can be cyclical – with majorities preferring A to B, B to C, and yet C to A – even when all individual preferences are free from such cycles (e.g., one of three individuals might prefer A to B to C, a second B to C to A, and a third C to A to B). And when "true/false" judgments are aggregated, say, in an expert panel, multi-member court, or committee, the majority judgments can be logically inconsistent even if all individual judgments are consistent, as illustrated in Table 1(a) (List and Pettit 2004). By contrast, when probabilities are aggregated, we can arrive at coherent collective probability assignments by averaging the individual probability assignments (for reviews, see Genest and Zidek 1986 and Dietrich and List 2016). To illustrate, if we reinterpret the judgments in Table 1(a) as probability assignments (where "true" and "false" correspond to probabilities 1 and 0, respectively), we arrive at a probabilistically coherent average assignment, as shown in Table 1(b).

Table 1: A simple example

(a) Binary case			
	p	q	$p \wedge q$
Individual 1	True	True	True
Individual 2	True	False	False
Individual 3	False	True	False
Majority	True	True	False

(b) Probabilistic case			
	p	q	$p \wedge q$
Individual 1	1	1	1
Individual 2	1	0	0
Individual 3	0	1	0
Majority	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$

Similarly, in preference or judgment aggregation, some plausible conditions (e.g., "unanimous views must never be overruled", "aggregation must be done in a pairwise or propositionwise manner") can only be satisfied by "dictatorial" aggregation rules, where one individual always determines the collective outcome – a result known as Arrow's theorem (1951/1963) for preference aggregation, a version of which also holds for the aggregation of "true/false" judgments (Dietrich and List 2007a, Dokow and Holzman 2010). By contrast, when probabilities are aggregated, the analogous conditions characterize the class of linear averaging rules, which seem democratic and well-behaved (Aczél and Wagner 1980, McConway 1981). In short, some of the notorious aggregation-theoretic pathologies do not occur when the aggreganda are probabilities rather than than "true/false" judgments or preference orderings. Probability aggregation, perhaps because of its greater informational richness, admits possibilities where the aggregation of those other aggreganda runs into impossibilities (see Dietrich and List 2017a).

In this paper, we show that, unfortunately, this happy observation does not carry over to the quest for aggregation rules that are immune to strategic voting. In the context of preference aggregation, the celebrated Gibbard-Satterthwaite theorem shows that, when there are more than two options, virtually all non-dictatorial aggregation rules violate "strategy-proofness": voters sometimes have incentives to misrepresent their preferences, so as to achieve an outcome they prefer by the lights of their true preferences (Gibbard 1973, Satterthwaite 1975). In the context of "true/false" judgment aggregation, something similar holds: under mild conditions, all non-dictatorial aggregation rules are manipulable by strategic voting (Dietrich and List 2007b). We will prove that non-manipulable probability aggregation is also essentially impossible, unless probabilities are assigned only to a single proposition and its negation.

Although there is much work on probability aggregation and much work on strategic voting in electoral contexts, there is little work on strategic voting in the context of probability aggregation (for a notable exception, see Laraki and Varloot 2022). The existing work tends to proceed by applying the established framework of the Gibbard-Satterthwaite theorem to the case of probability aggregation. In the framework of Gibbard and Satterthwaite, individuals (voters) each have a preference ordering over a set of options (candidates), and those preference orderings are aggregated into a winning outcome. For instance, voters each rank candidates A, B, and C in an order of preference, and the aggregation rule selects one candidate as the winner. The Gibbard-Satterthwaite theorem states that "strategy-proofness" is essentially unattainable unless one fixed individual is made a "dictator". The application of this result to probability aggregation takes the set of options ("candidates") to be the set of all possible probability functions on a given algebra of propositions, and assumes that probability aggregation, in effect, takes the form of preference aggregation over those "options". The Gibbard-Satterthwaite theorem now implies that voters will sometimes be incentivized to misrepresent their preferences over the possible probability functions, just as in ordinary elections they are sometimes incentivized to misrepresent their preferences over the electoral candidates.[1]

In this paper, we will proceed differently, taking inspiration from the theory of judgment aggregation rather than preference aggregation. Instead of going via the detour of the original Gibbard-Satterthwaite theorem and introducing preference orderings over probability functions, we will define "non-manipulability" directly, as the requirement that there be no opportunities for individuals to manipulate the collective probability assignments by misrepresenting their own probability assignments, in a way that is inspired by a similar definition of non-manipulability in binary ("true/false") judgment aggregation (Dietrich and List 2007b). The resulting analysis is simpler than one that introduces preferences over probability functions, while still diagnosing how probability aggregation is vulnerable to strategic manipulation.

[1] In a variant of the Gibbard-Satterthwaite framework, to which the theorem still applies, voters have full preference orderings over all the options but vote only for a single option. Laraki and Varloot (2022) employ that variant, which allows them to assume that individuals have full preference orderings over all possible probability functions while voting only for a single (most preferred) probability function. Laraki and Varloot, drawing on Freeman et al. (2021), also investigate how strategy-proof probability aggregation might become possible when the domain of admissible preference orderings is suitably restricted, such as to "single-peaked" ones (cf. Moulin 1980).

2 Definitions

We consider a set $N = \{1, 2, ..., n\}$ of two or more individuals (experts, jurors, policy-makers, committee members, or simply epistemic peers, ...) who each assign probabilities to some propositions, and we are looking for a probability aggregation rule: a method of aggregating these n individual probability assignments into a corresponding collective probability assignment.

Formally, a *proposition* (often also called *event*) is a subset of an underlying non-empty set Ω of possible worlds. Thus a proposition is identified with the set of those worlds in which the proposition is true. The empty set is called the *contradictory* proposition. The set Ω is called the *tautological* proposition. Any other proposition is called *contingent*. The *conjunction* $p \wedge q$ of any propositions p, q is given by their intersection $p \cap q$, the *disjunction* $p \vee q$ by their union $p \cup q$, and the *negation* $\neg p$ of any proposition p is given by its set-theoretic complement $\Omega \backslash p$. An *algebra* is a set of propositions that is closed under these three operations (in fact, closure under both negation and either conjunction or disjunction suffices for closure under all three). Following the standard approach in the theory of probability aggregation, we assume that the set of propositions to which probabilities are assigned – the "agenda" – is an algebra; call it X. We call this algebra *non-trivial* if it contains more than one contingent proposition-negation pair.[2]

For each individual $i \in N$, let Pr_i denote individual i's probability function on X, technically a function from X into the interval $[0, 1]$ satisfying the constraints of probabilistic coherence.[3] For each $p \in X$, $Pr_i(p)$ represents the (subjective) probability, credence, or degree of belief that individual i assigns to p. Let $\mathcal{P}$ denote the set of all possible probability functions on X. An n-tuple $\langle Pr_1, ..., Pr_n \rangle$ of probability functions across the individuals in N is called a *profile*.

A *probability aggregation rule* is a function $F : \mathcal{P}^n \to \mathcal{P}$, which assigns to each profile $\langle Pr_1, ..., Pr_n \rangle$ of individual probability functions a "collective" probability function Pr. A key aim of the theory of probability aggregation is to identify reasonable probability aggregation rules, which could plausibly be used in expert panels, policy-making committees, and other contexts of peer disagreement.

Our focus here is on the attainment of "non-manipulability": probability aggregation should not be vulnerable to the strategic misrepresentation of individual probability

[2] Any algebra that is not of the forms $\{\emptyset, \Omega\}$ or $\{\emptyset, \Omega, p, \neg p\}$ is non-trivial.

[3] Formally, a *probability function* on X is a function $Pr : X \to [0, 1]$ such that $Pr(\Omega) = 1$ and $Pr(p \vee q) = Pr(p) + Pr(q)$ whenever p and q have empty intersection. On a more demanding definition, the finite additivity condition is replaced by a countable additivity condition. Our main result holds in both cases.

assignments ("strategic voting"). Let us first define what it means for an individual to manipulate the collective probability for some proposition p. We say that individual i *can manipulate* the collective probability for p at the profile $\langle Pr_1, ..., Pr_n \rangle$ if

(i) i disagrees with the probability that would be collectively assigned to p if i were to submit to the aggregation rule his or her sincere probability function Pr_i, and

(ii) i can move the collective probability for p in the direction of his or her own probability for p by submitting a different ("strategically misrepresented") probability function Pr_i', while all others' probability functions are held fixed.

In such a situation, individual i has the opportunity to manipulate the group's probability for p in the direction of his or her own probability assignment, by pretending to hold the probability function Pr_i' rather than Pr_i. Formally, i *can manipulate* the collective probability for p at the profile $\langle Pr_1, ..., Pr_n \rangle$ if

(i) $Pr(p) \neq Pr_i(p)$, and

(ii) there exists a probability function Pr_i' such that

$$Pr'(p) \begin{cases} < Pr(p) & \text{if } Pr_i(p) < Pr(p), \\ > Pr(p) & \text{if } Pr_i(p) > Pr(p), \end{cases}$$

where $Pr = F(Pr_1, ..., Pr_n)$ and $Pr' = F(Pr_1, ..., Pr_i', ..., Pr_n)$.

Now "non-manipulability" is simply the following requirement on a probability aggregation rule F:

Non-manipulability: No individual $i \in N$ can manipulate the collective probability for any proposition $p \in X$ at any profile $\langle Pr_1, ..., Pr_n \rangle \in \mathcal{P}^n$.

This condition is a direct generalization of the condition of *non-manipulability* in binary ("true/false") judgment aggregation (Dietrich and List 2007b). There, we called an aggregation rule *non-manipulable* if no individual can manipulate the collective judgment on any proposition at any profile of judgments, and we said that individual i *can manipulate* the collective judgment on proposition p at a given profile of judgments if

(i) i disagrees with the judgment that would be collectively made on p if i were to submit to the aggregation rule his or her sincere set of judgments, and

(ii) i can bring about a collective judgment on p that matches his or her sincere judgment on p by submitting a different ("strategically misrepresented") set of judgments, while all others' sets of judgments are held fixed.

It is easy to see that, if we restrict the range of every probability function – individual and collective – to the extremal values 0 and 1 (thereby mimicking the case of "true/false" judgments), our probability-theoretic non-manipulability condition reduces to its judgment-theoretic counterpart. In the 0/1 case, "moving the collective probability for p in the direction of one's own probability for p" simply means "turning what was previously a mismatch between one's own judgment on p and the collective judgment into a match".

Although the present non-manipulability condition, like its counterpart in judgment aggregation, is inspired by the classic condition of *strategy-proofness* in social choice theory (Gibbard 1973, Satterthwaite 1975), there is a subtle interpretational difference. Strategy-proofness, as defined by Gibbard and Satterthwaite, is the absence of *incentives* to manipulate the outcome by voting strategically, where those incentives are defined relative to the voters' preferences. An aggregation rule is strategy-proof if it never gives any voters an incentive to misrepresent their preference orderings, so as to achieve an outcome they prefer by the lights of their true preference orderings. By contrast, non-manipulability, as defined here, is the absence of *opportunities* to manipulate the outcome by voting strategically, where an opportunity to manipulate occurs whenever someone can move the collective probability or judgment for any proposition in the direction of his or her own probability or judgment by misrepresenting his or her probability function or judgment set. This definition does not refer to any preferences at all. The focus on opportunities (rather than incentives) to manipulate is well-justified by the informational nature of the probability aggregation framework (or similarly by the nature of the judgment aggregation framework). Incentives can be analyzed only relative to certain assumptions about the individuals' preferences over various outcomes. Unlike the framework of preference aggregation, however, the frameworks of judgment or probability aggregation do not include any preferences. If one wanted to analyze incentives to manipulate, one would first have to enrich those frameworks, by explicitly introducing preferences over probability functions or over judgment sets, as discussed by Dietrich and List (2007b) in judgment aggregation (for related results, see Nehring and Puppe 2002) and by Laraki and Varloot (2022) in probability aggregation. By contrast, our present definition of non-manipulability applies to the probability-aggregation framework in its original form, without requiring the additional introduction of preferences.

In what follows, we will show that, under two very mild conditions on the aggregation rule, non-manipulable probability aggregation is impossible.

3 Main result

We introduce two minimal and uncontroversial conditions on any probability aggregation rule:

Consensus preservation: When all individuals have the same probability function, this becomes the collective probability function. Formally, for every unanimous profile $\langle Pr, ..., Pr \rangle \in \mathcal{P}^n$, $F(Pr, ..., Pr) = Pr$.

Non-dictatorship: There exists no individual whose probability function always becomes the collective one. Formally, there is no $i \in N$ such that, for every profile $\langle Pr_1, ..., Pr_n \rangle \in \mathcal{P}^n$, $F(Pr_1, ..., Pr_n) = Pr_i$.

Surprisingly, an impossibility result holds:

Theorem 1: For any non-trivial algebra X, there exists no probability aggregation rule $F : \mathcal{P}^n \to \mathcal{P}$ satisfying non-manipulability, consensus-preservation, and non-dictatorship.

In other words, any probability aggregation rule will satisfy at most two of these three conditions, but never all three. We state a proof in the next section.

To illustrate that any two of the three conditions can indeed be satisfied, we give three examples. First, consider a *dictatorial aggregation rule*, where, for every $\langle Pr_1, ..., Pr_n \rangle \in \mathcal{P}^n$, $F(Pr_1, ..., Pr_n) = Pr_i$ for some antecedently fixed individual $i \in N$ (the "dictator"). This is non-manipulable (insofar as neither the dictator nor anyone else can manipulate the outcome by misrepresenting their probability function) and consensus-preserving (insofar as any consensus will be shared by the dictator too), but it obviously violates non-dictatorship. Secondly, consider an *imposed aggregation rule*, where, for every $\langle Pr_1, ..., Pr_n \rangle \in \mathcal{P}^n$, $F(Pr_1, ..., Pr_n) = Pr$ for some antecedently fixed probability function Pr. This is also non-manipulable (insofar as no individual can affect the outcome at all) and non-dictatorial (since there is no dictator), but not consensus-preserving. Indeed, it is completely unresponsive to the individual probability functions that are being fed into it and always produces the same fixed outcome. Finally, consider *linear averaging*, where, for every $\langle Pr_1, ..., Pr_n \rangle \in \mathcal{P}^n$, $F(Pr_1, ..., Pr_n) = \frac{1}{n}(Pr_1 + ... + Pr_n)$. This is clearly consensus-preserving and non-dictatorial, and it looks more reasonable than a dictatorial or imposed aggregation rule, but it can be manipulated. Any individual whose probability for some proposition p is higher (or lower) than the anticipated collective probability can manipulate the collective probability in his or her own direction by overstating (or understating) his or her probability for p, assuming it is not already maximal (or minimal). We now summarize the proof of our theorem.

4 Proof

Our proof proceeds via a series of lemmas. We first establish that non-manipulability implies the following two conditions:

Propositionwise independence: The collective probability for any proposition p depends only on the individual probabilities for p, not on individual probabilities for other propositions. Formally, for any profiles $\langle Pr_1, ..., Pr_n \rangle, \langle Pr'_1, ..., Pr'_n \rangle \in \mathcal{P}^n$ and any proposition $p \in X$, if, $Pr_i(p) = Pr'_i(p)$ for all $i \in N$, then $Pr(p) = Pr'(p)$, where $Pr = F(Pr_1, ..., Pr_n)$ and $Pr' = F(Pr'_1, ..., Pr'_n)$.

Reinforcement monotonicity: The collective probability for any proposition does not change if an additional individual comes to accept that collective probability. Formally, for any individual $i \in N$, proposition $p \in X$, profile $\langle Pr_1, ..., Pr_n \rangle \in \mathcal{P}^n$, and alternative individual probability function Pr'_i, if $Pr_i(p) \neq Pr(p)$ and $Pr'_i(p) = Pr(p)$, then $Pr'(p) = Pr(p)$ where $Pr = F(Pr_1, ..., Pr_n)$ and $Pr' = F(Pr_1, ..., Pr'_i, ..., Pr_n)$.

In the Appendix, we prove:

Lemma 1: Non-manipulability implies propositionwise independence.

Lemma 2: Non-manipulability implies reinforcement monotonicity.

We further observe that, in the presence of propositionwise independence, consensus preservation implies:

Zero-preservation: For any profile $\langle Pr_1, ..., Pr_n \rangle \in \mathcal{P}^n$ and any proposition $p \in X$, if $Pr_i(p) = 0$ for all $i \in N$, then $Pr(p) = 0$ where $Pr = F(Pr_1, ..., Pr_n)$.

Lemma 3: Consensus preservation and propositionwise independence jointly imply zero-preservation.

By lemmas 1 and 3, a probability aggregation rule that satisfies our theorem's three conditions will also satisfy propositionwise independence and zero-preservation. We can now make use of a classic characterization theorem.

Lemma 4 (Aczél and Wagner 1980, McConway 1981): For any non-trivial algebra, a probability aggregation rule $F : \mathcal{P}^n \to \mathcal{P}$ satisfies propositionwise independence and zero-preservation if and only if it is a linear pooling rule, i.e., there exist weights $w_1, w_2, ..., w_n \geq 0$ with $w_1 + w_2 + ... + w_n = 1$ such that, for every profile $\langle Pr_1, ..., Pr_n \rangle \in \mathcal{P}^n$,

$$Pr = w_1 Pr_1 + w_2 Pr_2 + ... + w_n Pr_n,$$

where $Pr = F(Pr_1, ..., Pr_n)$.

Combining all four lemmas yields the following:

Intermediate conclusion: For any non-trivial algebra, a probability aggregation rule satisfies non-manipulability and consensus preservation *only if* it is a linear pooling rule satisfying reinforcement monotonicity.

We finally observe that no non-dictatorial linear pooling rule satisfies reinforcement monotonicity; i.e., the only way to satisfy reinforcement monotonicity under linear pooling is to give one individual a weight of 1 and to give all others a weight of 0.

Lemma 5: A linear pooling rule satisfies reinforcement monotonicity *only if* it is dictatorial.

Theorem 1 now follows inmediately. This completes our proof.

5 Is this result surprising?

One might wonder whether we should be surprised by the present impossibility result. The answer is yes and no. On the one hand, the result is surprising because the central impossibility result for binary judgment aggregation – the analogue of Arrow's theorem – does not carry over to probability aggregation. So, one might have expected that the impossibility of non-manipulable judgment agregation – the analogue of the Gibbard-Satterthwaite theorem – also fails to carry over to probability aggregation. This expectation would not have been unreasonable, since, in standard social choice theory, Arrow's theorem and the Gibbard-Satterthwaite theorem are known to be intimately connected. On the other hand, our present impossibility result may look less surprising once we recognize that it perfectly corresponds to a matching impossibility result on non-manipulable binary judgment aggregation. Let us now explain this in more detail.

Let $\mathcal{J}$ be the set of all probability functions on the algebra X that take only the extremal probability values 0 and 1. The set $\mathcal{J}$ is a very restricted subset of $\mathcal{P}$. We can think of a restricted probability aggregation rule $F : \mathcal{J}^n \to \mathcal{J}$ as representing a binary ("true/false") judgment aggregation rule.[4] The following is a version of what we have elsewhere called "Arrow's theorem in judgment aggregation" (Dietrich and List 2007a, Dokow and Holzman 2010; relatedly, see Nehring and Puppe 2010):

Impossibility of propositionwise independent judgment aggregation: For any non-trivial algebra, every aggregation rule $F : \mathcal{J}^n \to \mathcal{J}$ (a "binary judgment aggregation rule") satisfying consensus preservation and propositionwise independence is dictatorial.

[4]Whenever we refer to an aggregation rule $F : \mathcal{J}^n \to \mathcal{J}$ and apply our conditions to it, we re-interpret these as referring only to probability functions in $\mathcal{J}$ rather than $\mathcal{P}$, i.e., essentially to binary judgments.

This is an impossibility result insofar as dictatorial aggregation rules are unattractive.[5] By contrast, a version of the seminal theorem of Aczél and Wagner (1980) and McConway (1981), stated as lemma 4 above, is the following possibility result:

Possibility of propositionwise independent probability aggregation: For any non-trivial algebra, every aggregation rule $F : \mathcal{P}^n \to \mathcal{P}$ (a "probability aggregation rule") satisfying consensus preservation and propositionwise independence is a linear pooling rule.

So, the move from the binary to the probabilistic format turns an impossibility result into a possibility result. However, contrast this pair of results with the following pair. The first result in this second pair is an impossibility theorem on non-manipulable judgment aggregation, an analogue of the Gibbard-Satterthwaite theorem (Dietrich and List 2007b); the second is our present theorem.

Impossibility of non-manipulable judgment aggregation: For any non-trivial algebra, there exists no aggregation rule $F : \mathcal{J}^n \to \mathcal{J}$ (a "binary judgment aggregation rule") satisfying non-manipulability, consensus preservation, and non-dictatorship.[6]

Impossibility of non-manipulable probability aggregation: For any non-trivial algebra, there exists no aggregation rule $F : \mathcal{P}^n \to \mathcal{P}$ (a "probability aggregation rule") satisfying non-manipulability, consensus preservation, and non-dictatorship.

This suggests that the impossibility of non-manipulable judgment aggregation – the analogue of the Gibbard-Satterthwaite theorem – is more persistent than the impossibility of propositionwise independent judgment aggregation – the analogue of Arrow's theorem – which goes away when we move from the binary to the probabilistic format.

6 Escape routes

Are there any escape routes from the impossibility of non-manipulable probability aggregation? The first thing to note is that, since our theorem's other two conditions on the aggregation rule – consensus preservation and non-dictatorship – are extremely compelling and uncontroversial, dropping one of them does not seem very plausible. Dictatorial aggregation rules are not just undemocratic, but they are also degenerate

[5] The original theorem applies to a more general class of "agendas". Specifically, the set X on which judgments are made need not be an algebra, but could be any negation-closed set of propositions with two combinatorial properties ("non-affineness" and "path-connectedness"). It so happens that any non-trivial algebra has those combinatorial properties.

[6] Again, the original theorem in Dietrich and List (2007b) applies to a more general class of "agendas"; X needs to be merely negation-closed and path-connected. Any non-trivial algebra has those properties.

limiting cases of aggregation rules. And aggregation rules that violate consensus preservation are equally unattractive. If all individuals agree on the assignment of probabilities to *all* propositions in the given algebra, what reason could there be to overrule such a consensus? Furthermore, the impossibility result would continue to hold if consensus preservation were weakened to the following, even less demanding condition.

Uncertainty-free consensus preservation: When all individuals have the same uncertainty-free probability function, this becomes the collective probability function. Formally, for every unanimous and uncertainty-free profile $\langle Pr, ..., Pr \rangle \in \mathcal{P}^n$ (where Pr assigns only the extremal values 0 and 1 to all propositions in X), $F(Pr, ..., Pr) = Pr$.

This treats as "sacrosanct" only those cases of a consensus in which all individuals are certain about all propositions in X. In such cases, it is even harder to image a reason for overruling the consensus in question. Since uncertainty-free consensus preservation still implies zero-preservation in the presence of propositionwise independence, the proof of our theorem continues to go through if consensus preservation is weakened to it.

It seems, then, that the only way to allow for non-manipulable probability aggregation is to restrict one's consideration to cases where the algebra of propositions is trivial, i.e., where it is of the form $\{\emptyset, \Omega, p, \neg p\}$ for some contingent proposition p; we can set aside the even more trivial case $\{\emptyset, \Omega\}$. In such trivial cases, probabilities are assigned at most to a single proposition-negation pair. A non-manipulable probability aggregation rule that will work for such an algebra is the one that assigns to each proposition the median probability assigned to it by the individuals in N (assuming the number of individuals is odd). Even if the individuals were willing to misrepresent their probability assigments to achieve the closest possible match between their individually assigned probabilities and the collective ones, their best way to achieve this outcome would be to reveal their own probabilities truthfully. Reporting a probability for p that is lower than the sincerely held probability for p would only run the risk of shifting the median probability below one's own probability for p, and reporting a probability for p that is too high would only run the risk of raising the median probability above one's own probability. Indeed, a median aggregation rule is a paradigmatic example of an aggregation rule that is reinforcement-monotonic.

It is also illuminating to see why such a median rule is not available for a non-trivial algebra. The reason is that if we assigned the median probability to every proposition $p \in X$, the collective probability assignment could easily become probabilistically incoherent. Trivial algebras are the only ones for which the median rule is a well-defined probability aggregation rule, i.e., one that guarantees probabilistic coherence. But this excludes collective probability assignments to complex issues, a severe limitation.

Interestingly, however, we can find something that looks like a partial escape route from our impossibility. Consider an algebra based on a set of worlds Ω that can be "decomposed" into binary characteristics, such as whether global warming by 2050 will exceed 1.5 °C compared to preindustrial levels, whether CO_2 emissions will decrease by at least 50% by 2030, and so on. For concreteness, suppose the algebra is the power set of $\Omega = \{0,1\}^k$ for some $k > 1$. So, worlds are each represented by k-tuples of zeros and ones, corresponding to k distinct characteristics that characterize any such world. We can then identify k logically independent propositions $p_1, ..., p_k$, where, for each j, p_j stands for "the jth characteristic is 1" and $\neg p_j$ stands for "the jth characteristic is 0". If we are also willing to stipulate that the j characteristics are probabilistically independent, we can assign collective probabilities to all propositions in the given algebra by taking the collective probability for any proposition in $\{p_1, ..., p_k, \neg p_1, ..., \neg p_k\}$ to be the median individual probability for that proposition (assuming the number of individuals is odd) and completing the rest of the collective probability function uniquely as required for probabilistic coherence. This will be a form of "premise-based" probability aggregation (as defined in Dietrich and List 2017b), where the propositions in $\{p_1, ..., p_k, \neg p_1, ..., \neg p_k\}$ serve as the "premises" (to which a median rule is applied) and all other propositions are treated as "conclusions" (for which the collective probability is derived by implication, relative to the constraints of probabilistic coherence, under the stipulation of probabilistic independence across the k characteristics). While this aggregation rule is manipulable (and it preserves consensus only for profiles in which the individuals themselves treat the j characteristics as probabilistically independent), it does not permit manipulation of the collective probabilities for any of the premise propositions. If we were willing to relax non-manipulability to a form of non-manipulability that is restricted to the premises alone (on the grounds that those are the most important propositions), then we would be able to achieve this weakened form of non-manipulability.

The present partial escape route from our impossibility – recognizing that premise-based aggregation can be non-manipulable on the premise propositions, though not non-manipulable across the board – mirrors a parallel escape route from the impossibility of non-manipulable binary judgment aggregation. There too, one can achieve a restricted form of non-manipulability on the premises by using a premise-based aggregation rule. Under such an aggregation rule, the group takes majority votes only one some logically independent "premises", such as propositions p and q in the example of Table 1(a), and derives the collective judgments on all other propositions, such as $p \wedge q$ in our example, by logical inference. Premise-based aggregation is non-manipulable on the premises, but it is vulnerable to manipulation on other propositions.

7 Concluding remarks

As soon as probabilities are assigned to more than one proposition and its negation, non-manipulability is essentially impossible to achieve in probability aggregation: for any non-trivial algebra of propositions, any probability aggregation rule can satisfy at most two of (1) non-manipulability, (2) consensus preservation, and (3) non-dictatorship. Since the latter two conditions are hard to give up, it seems that we must live with the manipulability of any probability aggregation rule that we use in practice.

One consolation may be that the existence of an opportunity for an individual to manipulate does not automatically translate into an actual act of manipulation. The individual in question must be able to identify that opportunity, which requires sufficient information about the probability functions of others and a sufficient computational capacity to figure out how to manipulate effectively. These two epistemic conditions – informational and computational – are not always met in real-world conditions. Moreover, the individual must be inclined and willing to manipulate – a motivational condition. As has been noted in discussions of strategic voting within the theory of democracy, misrepresenting one's views comes at certain cognitive and possibly also reputational costs, especially when voting is preceded by a period of group deliberation in which the participants share their views with one another. Moreover, according to theories of "expressive" rather than "instrumental" voting (Brennan and Lomasky 1993), truthful voting can be entirely rational, since voters may care about their own vote as an end in itself (for instance, as an expressive act) and not just as a means for changing the election outcome. Here, opportunities for manipulation will not be seized; they will not translate into acts of manipulation. One can therefore hope that, in real-world settings of probability aggregation, opportunities for manipulation will not always be acted upon.

References

Aczél, J., and C. Wagner (1980). A characterization of weighted arithmetic means. *SIAM Journal on Algebraic and Discrete Methods* 5(1): 1–8.

Arrow, K. J. (1953/1963). *Social Choice and Individual Values.* New York (Wiley).

Brennan, G., and L. Lomasky (1993). *Democracy and Decision: The Pure Theory of Electoral Preference.* Oxford (OUP).

Dietrich, F., and C. List (2007a). Arrow's theorem in judgment aggregation. *Social Choice and Welfare* 29(1): 19–33.

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

Dietrich, F., and C. List (2016). Probabilistic opinion pooling. In C. Hitchcock and A. Hajek (eds.), *Oxford Handbook of Probability and Philosophy.* Oxford (OUP).

Dietrich, F., and C. List (2017a). Probabilistic opinion pooling generalized. Part one: general agendas. *Social Choice and Welfare* 48(4): 747–786.

Dietrich, F., and C. List (2017b). Probabilistic opinion pooling generalized. Part two: the premise-based approach. *Social Choice and Welfare* 48(4): 787–814.

Dokow, E., and R. Holzman (2010). Aggregation of binary evaluations. *Journal of Economic Theory* 145(2): 495–511.

Freeman, R., D. M. Pennock, D. Peters, and J. W. Vaughan (2021). Truthful aggregation of budget proposal. *Journal of Economic Theory* 193(April): 105234.

Genest, C., and J. V. Zidek (1986). Combining Probability Distributions: A Critique and Annotated Bibliography. *Statistical Science* 1(1): 113–135.

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

Laraki, R., and E. Varloot (2022). Level-Strategyproof Belief Aggregation and Application to Majority Judgment under Uncertainty. Working paper, arXiv:2108.04705v4.

List, C., and P. Pettit (2004). Aggregating sets of judgments: Two impossibility results compared. *Synthese* 140(1–2): 207–235.

McConway, K. J. (1981). Marginalization and Linear Opinion Pools. *Journal of the American Statistical Association* 76(374): 410–414.

Moulin, H. (1980). On Strategy-Proofness and Single Peakedness. *Public Choice* 35(4): 437–455.

Nehring, K., and C. Puppe (2002). Strategyproof Social Choice on Single-Peaked Domains: Possibility, Impossibility and the Space Between. Working paper, UC Davis.

Nehring, K., and C. Puppe (2010). Abstract Arrovian aggregation. *Journal of Economic Theory* 145(2): 467–494.

Satterthwaite, M. (1975). Strategy-proofness and Arrow's conditions. *Journal of Economic Theory* 10(2): 187–217.

Appendix: remaining proofs

This appendix contains the proofs of Lemmas 1 and 2, the two key results which (jointly with the classic linearity result) imply our theorem via the argument presented in the main text. We follow the formalism of the main text.

Proof of Lemma 1. Assume non-manipulability. To show propositionwise independence, fix $p \in X$ and $(Pr_1, ..., Pr_n), (Pr'_1, ..., Pr'_n) \in \mathcal{P}^n$ such that $Pr_i(p) = Pr'_i(p)$ for all i. Writing $Pr = F(Pr_1, ..., Pr_n)$ and $Pr' = F(Pr'_1, ..., Pr'_n)$, we must show $Pr(p) = Pr'(p)$. We can assume without loss of generality that the two profiles differ only at one individual, since any profile in $\mathcal{P}^n$ can be transformed into any other one in n steps, by replacing first individual 1's probability function, then individual 2's probability function, etc., where any two consecutive profiles differ only at one individual. Let i be the only individual such that $Pr_i \neq Pr'_i$. Hence, $(Pr'_1, ..., Pr'_n) = (Pr_1, ..., Pr'_i, ..., Pr_n)$. There are three cases:

Case 1: $Pr(p) < Pr_i(p)$. Then $Pr'(p) \leq Pr(p)$; otherwise i could manipulate the collective probability for p at $(Pr_1, ..., Pr_n)$ by submitting Pr'_i. As also $Pr(p) < Pr_i(p)$, we have $Pr'(p) < Pr_i(p)$. Thus $Pr(p) \leq Pr'(p)$; otherwise individual i could manipulate the collective probability for p at $(Pr_1, ..., Pr'_i, ..., Pr_n)$ by submitting Pr_i. Note that $Pr'(p) \leq Pr(p)$ and $Pr(p) \leq Pr'(p)$. Thus $Pr(p) = Pr'(p)$.

Case 2: $Pr_i(p) < Pr(p)$. Again $Pr(p) = Pr'(p)$, for reasons like in Case 1.

Case 3: $Pr_i(p) = Pr(p)$. Then also $Pr'_i(p) = Pr(p)$, as $Pr_i(p) = Pr'_i(p)$. Thus $Pr(p) = Pr'(p)$; otherwise individual i could manipulate the collective probability for p at $(Pr_1, ..., Pr'_i, ..., Pr_n)$ by submitting Pr_i, in fact even to the extent of fully enforcing his or her own genuine probability $Pr'_i(p) = Pr(p)$. ■

Proof of Lemma 2. Assume non-manipulability. To show reinforcement monotonicity, fix an individual i, proposition $p \in X$, profile $(Pr_1, ..., Pr_n) \in \mathcal{P}^n$, and alternative probability function $Pr'_i \in \mathcal{P}$. Writing $Pr = F(Pr_1, ..., Pr_n)$ and $Pr' = F(Pr_1, ..., Pr'_i, ..., Pr_n)$, we assume that $Pr_i(p) \neq Pr(p)$ and $Pr'_i(p) = Pr(p)$, and must show that $Pr'(p) = Pr(p)$. This holds because otherwise individual i could manipulate the collective probability for p at profile $(Pr_1, ..., Pr'_i, ..., Pr_n)$ by submitting Pr_i, which in fact fully enforces his or her own probability $Pr'_i(p) = Pr(p)$. ■