

Collective intelligence through aggregation^{*}

Franz Dietrich and Christian List

Suppose a committee, expert panel, or other group is making judgments on some issues, where these may be not just yes/no-questions, such as whether a defendant is guilty, but also variables with many possible values, such as macroeconomic or meteorological variables or travel directions. Furthermore, there may be interconnections between different issues, as in the case of economic or climate variables. How can the group arrive at “intelligent” collective judgments, based on the group members’ individual judgments? We investigate three challenges raised by this judgment-aggregation problem. First, reasonable methods of aggregation (such as defining the collective judgment for each issue as the average or median judgment) can produce inconsistent collective judgments. Secondly, many methods of aggregation are manipulable by strategic voting. Finally, not all methods of aggregation are conducive to tracking the truth on the issues in question. We prove new impossibility or possibility theorems on all three challenges, identifying what it takes to produce collective judgments in a consistent, non-manipulable, and truth-tracking manner and thereby to achieve collective intelligence through aggregation. Overall, the median method, though imperfect, performs reasonably well. We also note the relevance of our analysis for non-human group decisions.

1. Introduction

In a much-cited experiment reported in *Nature* in 1907 [26], the statistician Francis Galton asked the participants of an agricultural fair in rural England to estimate the weight of an ox on display. Around 800 people submitted their estimates. Although the participants were familiar with farm animals, the estimates varied significantly. But the *median* estimate – the one exactly in the middle when all submitted estimates were listed in ascending order – came remarkably close to the true weight. The true weight was 1198lbs, and the median estimate was 1207lbs, an error of only 0.75%. The average

^{*}We thank two anonymous referees and an editor for very helpful comments. Published version in *Philosophical Transactions of the Royal Society B* (2026) 381 (1948): 20240454, available at <https://doi.org/10.1098/rstb.2024.0454>.

individual error was far greater. Galton’s experiment illustrates the idea of “collective intelligence through aggregation”: asking many individuals to express a judgment on some issue and deriving from the answers a collective judgment – for short, *aggregating the individual judgments* – can be a better way to “track the truth” than to ask a single individual alone [25]. Moreover, this can work even if there is significant disagreement among the participants.

In Galton’s study, the method of aggregation was the median rule – taking the median individual judgment as the collective one – but different methods might be used in other cases, such as averaging. Galton’s example is an instance of what we call a “general judgment aggregation problem”:

- A group of individuals – e.g., a committee, an expert panel, or even a non-human group – is making judgments (“decisions”) on one or more issues or variables, and the goal is to arrive at a collective judgment (“decision”) on those issues or variables;
- the issues or variables could be either binary (such as “guilt” versus “innocence” in a criminal jury) or many-valued (such as the weight of an ox); and
- when there are two or more issues or variables, these may be interconnected, in the sense that the judgments (“decisions”) on some of them constrain those that can consistently be made on others.

Galton’s example involves just a single, many-valued issue. Many decision-making groups, however, need to estimate several variables with constraints between them. Think of an expert panel predicting future macroeconomic variables, such as size of the workforce, each worker’s productivity, and GDP, where the last variable is the product of the first two. Or think of an expert panel making judgments about a complex system involving numerous interconnected variables, such as the Earth’s climate system.

Beyond human societies, animal groups sometimes aggregate judgments too, for instance when they choose a travel direction, grazing spot, or nest site from amongst several options, and here, too, the advantages of democratic decision methods have been noted [9, 10, 11, 39]. Finally, judgment aggregation problems also arise in the context of artificial intelligence, for instance when several data bases need to be merged or the inputs of several AI agents need to be combined. Table 1 provides an illustrative taxonomy of different kinds of aggregation problems (see below for citations of relevant literature on different such problems).

Table 1: Examples of general aggregation problems

	One issue/variable	Many issues/variables
Binary judgments	Jury decisions ("guilty", "not guilty")	Making yes/no judgments on logically related propositions
Non-binary judgments	Evaluating or estimating one issue/variable Choosing a grazing spot, nest site, or travel direction	Estimating interconnected economic or climate variables Assigning probabilities or likelihoods to several events

In this paper, we will model such judgment aggregation problems in full generality and investigate three challenges they may raise. First, reasonable methods of aggregation can produce inconsistent collective judgments. Second, many methods of aggregation are manipulable by strategic voting. And third, only some methods successfully track the truth. We will discuss how one might respond to those challenges. Interestingly, the median method, as in Galton’s study, performs better overall than averaging, but it is still threatened by some of the identified challenges.

Our analysis builds on several different literatures: the literature on binary and non-binary judgment aggregation and social-choice theory [1, 2, 5, 8, 14, 22, 23, 24, 30, 33, 37, 40, 42, 44], the literature on Condorcet’s jury theorem [13, 28, 32, 6, 20, 21, 38, 29], and the literature on grade aggregation [3, 4, 35, 29]. All formal proofs and some additional analyses are included in online supplementary material (linked at <https://doi.org/10.1098/rstb.2024.0454>).

2. The single-issue case

We begin by assuming that a group makes a judgment on just a single issue or variable. Formally, we have a set $N = \{1, 2, \dots, n\}$ of two or more individuals (decision-makers, experts, judges, ...) and an issue or variable, denoted x , which takes values in some set V (e.g., the set $\mathbf{R}$ of real numbers). The goal is to assign a value from that set to the given issue/variable (in effect, to “choose” a value for x). We write J_i to denote individual i ’s *evaluation* of that issue/variable, where $J_i \in V$. This expresses i ’s judgment as to which value of x is the right one. In Galton’s study, J_i is the weight estimate submitted by the i^{th} participant. A list (n -tuple) of evaluations across the individuals, written $\langle J_1, \dots, J_n \rangle$, is called a *profile of evaluations*. Examples of issues or variables are:

- the weight of an ox or another real-valued variable, say in economics or climate science, such as GDP or the atmospheric greenhouse gas concentration ($V = \mathbf{R}$);

- the likelihood of an event, such as an earthquake or the melting of a glacier, where this could be either qualitative ($V = \{\text{very unlikely, unlikely, as likely as not, ...}\}$) or quantitative (i.e., a probability in $V = [0, 1]$);
- the grade of an exam paper or an athlete (e.g., $V = \{A, B, C, \dots\}$);
- a candidate to be chosen for some political office ($V = \{\text{Harris, Trump, ...}\}$); or
- a travel direction (e.g., $V = [0^\circ, 360^\circ]$), a grazing spot, or a nest site (e.g., V is a set of locations).

An *aggregation rule* is a function, denoted F , that maps each profile $\langle J_1, \dots, J_n \rangle$ in some domain of admissible profiles (e.g., all logically possible profiles) to a collective evaluation $J = F(J_1, \dots, J_n)$, where J stands for the group's evaluation of the issue/variable x , with $J \in V$. Examples of aggregation rules are the following:

- *Plurality rule* (or *mode rule*): The collectively assigned (“chosen”) value for issue/variable x is the value most often assigned by the individuals (and if two or more values are tied, it is one of those values, selected by some tie-breaking criterion).
- *Median rule*: The collectively assigned value is the median value assigned by the individuals, i.e., the value exactly in the middle (“middlemost”) when all evaluations are listed in ascending order (with repetitions whenever the same value is assigned more than once), assuming an odd number of individuals.¹
- *Quantile rule with parameter p* (where $0 \leq p \leq 1$): The collectively assigned value is the p -quantile value assigned by the individuals, i.e., the value at position $m = \lceil pn \rceil$ in the ascending order of evaluations, where $\lceil pn \rceil$ is the smallest number in $\{1, 2, \dots, n\}$ greater than or equal to pn . For instance, the 0-quantile rule selects the lowest value assigned by the individuals; the 1-quantile rule selects the highest value; and the 0.5-quantile rule selects the median value (assuming n is odd).²
- *Averaging rule*: The collectively assigned value is the arithmetic average of the values assigned by the individuals.

¹When n is even, there is a *median pair* of values (e.g., if $n = 4$, the second and third values in the ascending order). If these coincide, we get a unique value. If not, we can extend the definition of the median as follows: we could either (i) apply a tie-breaking criterion (e.g., choose the lower of the two values), which would give us a non-neutral rule; or (ii) deem all values in the interval weakly between the lower and the higher value admissible, which would lead to a multi-function generating a set of outputs; or (iii) randomize between the two values, which would yield a non-deterministic rule.

²Quantile rules (sometimes also called “order functions”) have been discussed in detail in [3, 29].

- *Distance-based rule*: The collectively assigned value is a value J that minimizes the sum-total distance from the individual values $\sum_{i=1}^n d(J, J_i)$, where d is some distance function on the set V . If V is an interval of real numbers (possibly all of $\mathbf{R}$) the distance between any two values J, J' could be defined as $d(J, J') = |J - J'|^q$ for some fixed $q \geq 1$. For $q = 1$, the corresponding distance-based rule becomes the median rule (assuming n is odd), and for $q = 2$, it becomes the averaging rule. For other values of q , it is distinct.
- *Dictatorship*: The collectively assigned value is the value assigned by some antecedently fixed individual, where the same individual determines the collectively assigned value for every profile.

Although, in human groups, it is natural to think of an aggregation rule as a formal or explicit mechanism for generating collective judgments based on individual judgments (“explicit voting”), the concept of an aggregation rule, understood as a function from individual-level inputs to collective-level outputs, can also accommodate informal or implicit mechanisms (“implicit voting” or “voting with one’s feet”), as in the example of animal groups making collective decisions in a participatory manner [9].

What are the properties – especially the advantages and disadvantages – of different aggregation rules? The first point to note is that different rules rely on different assumptions about the nature of the issue/variable to be evaluated. The plurality and dictatorial rules are always formally well-defined, irrespective of whether we are estimating a real-valued variable (such as the weight of an ox), assigning a qualitative grade to an exam paper (such as A, B, C etc.), choosing one of several electoral candidates (such as Harris or Trump), or choosing a nest-site or grazing spot. Those rules do not require any particular structure of the value set; we can think of its elements simply as options among which a choice is to be made. The median, quantile, and averaging rules, by contrast, require a value set with a certain structure. For the median and quantile rules, the set of possible values must be linearly ordered from lowest to highest, and for the average, we need an arithmetic scale (e.g., the real numbers), with a value set that is closed under averaging. This makes the averaging rule inapplicable, for instance, to grade assignments such as A, B, C etc. or to qualitative probability assignments such as “unlikely”, “likely” etc., where the average is undefined. The median and quantile rules can still be applied here, since the values can be linearly ordered. A distance-based rule, finally, requires a value set for which a distance function is defined.

Moreover, although averaging seems a particularly obvious method of aggregation, it is meaningful only if the scale is sufficiently non-arbitrary. If it were arbitrary whether

to measure the severity of an earthquake on a linear scale or on a logarithmic one, then averaging experts' estimates of earthquake severity would have no meaning independently of the chosen scale. If, on the other hand, the scale is unique up to positive linear transformations of the form $ax + b$, with $a > 0$, then averaging does have a scale-independent meaning (technically, averaging commutes with rescaling). Temperature is a good example. Suppose Celsius, Kelvin, and Fahrenheit are the only admissible scales. Then averaging several temperature estimates in Celsius and converting the result into Fahrenheit or Kelvin yields the same aggregate estimate as first converting the estimates into Fahrenheit or Kelvin and then averaging them. The median and quantile rules, by contrast, do not require an arithmetic scale. An ordering of the values suffices (on the significance of scales, see, e.g., [43] and [42]).

The median rule further admits a majoritarian rationalization [3, 4]: the median value (where n is odd) is the unique value for which a majority judges that it is not too high *and* a majority judges that it is not too low. No other value has this “democratic” property. Finally, the median rule is robust to outlier judgments. Had a few participants in Galton’s study erroneously estimated the weight of the ox to be zero or a million lbs, this would not have changed the median. The averaging rule lacks this robustness. Quantile rules can be rationalized if we wish to find a value that only a minority considers too high (if $p < 0.5$) or only a minority considers too low (if $p > 0.5$).

3. The challenge of inconsistency

To see how the aggregation of judgments becomes more complicated when there are multiple issues or variables, such as in expert panels or policy-making committees, consider a simple example (for earlier related examples, see [7, 8]). Suppose a three-member expert panel seeks to estimate three macroeconomic variables for a medium-sized country for some future year.

x : workforce productivity, i.e., GDP per worker;

y : employment level, i.e., number of workers;

z : GDP.

It is agreed that $z = x \times y$. Suppose the experts’ estimates are as shown in the first three rows of Table 2.

Table 2: Experts' estimates

	Workforce productivity (x) (EUR)	Employment level (y) (number of workers)	GDP (z) (EUR)
Expert 1	116 thousand	25 million	2900 billion
Expert 2	87 thousand	39 million	3393 billion
Expert 3	106 thousand	26 million	2756 billion
Median	106 thousand	26 million	2900 billion
Average	103 thousand	30 million	≈ 3016 billion

Although the experts differ in their estimates for x , y , and z , their individual estimates are consistent with the agreed constraint $z = x \times y$. Strikingly, however, if the expert panel aggregates these estimates via the median or the averaging rule, it reaches collective estimates that are inconsistent with the agreed constraint. The median estimates for x and y would imply a value of 2756 billion for z , while the median estimate for z is 2900 billion. The average estimates for x and y would imply a value of 3090 billion for z , while the average estimate for z is ≈ 3016 billion. So, although the median rule or the averaging rule can be applied to each variable *in isolation*, the simultaneous application of one of these rules to *all* variables yields an inconsistent result [7, 8].

The observation that the aggregation of judgments via the median or average may lead to inconsistencies resembles a classic problem concerning the aggregation of binary (true/false) judgments, where majority voting does not guarantee consistent collective judgments even when all individual judgments are consistent [30, 33, 34]. To illustrate, suppose three experts make judgments on three propositions about the year 2050:

p : The atmospheric greenhouse gas concentration will exceed 650 ppm CO₂e.

$p \rightarrow q$: If the atmospheric greenhouse gas concentration exceeds 650 ppm CO₂e, then the Artic ocean will be ice-free.

q : The Artic ocean will be ice-free.

If individual judgments are as shown in Table 3, for instance, then majorities judge that p is true, that $p \rightarrow q$ is true, and yet that q is false, an inconsistent set of judgments.

Table 3: Experts' judgments

	p	$p \rightarrow q$	q
Expert 1	True	True	True
Expert 2	True	False	False
Expert 3	False	True	False
Majority	True	True	False

Is the failure of plausible aggregation rules to guarantee consistent collective judgments an artifact of the averaging or median rules, or does this problem occur more generally? To address this question, we introduce a general model (earlier related modelling strategies can be found in [7, 8, 16, 23, 24]).

As before, we have a set $N = \{1, 2, \dots, n\}$ of two or more individuals. Now there is a finite and non-empty set $X = \{x, y, z, \dots\}$ of issues or variables. Again, these take values in some set V (e.g., the set $\mathbf{R}$ of real numbers). Each individual i 's judgments are represented by an *evaluation* J_i , formally a function from the set X into the set V , which assigns to each issue or variable x in X a value $J_i(x)$ in V , expressing i 's judgment about the value of x .

To represent the constraints between different issues or variables, we introduce a non-empty set $\mathcal{J}$ of evaluations (functions from X into V) which count as *consistent*. In the example of the three macroeconomic variables, the set $\mathcal{J}$ consists of those evaluations J which satisfy the constraint $J(z) = J(x) \times J(y)$. Valuations breaching that constraint are excluded from $\mathcal{J}$. If, in a more complicated case, the constraints between different issues or variables were given by an entire system of equations or inequalities, then the set $\mathcal{J}$ would consist of those evaluations J that respect all those equations or inequalities. In general, the set of issues or variables X together with the set V of possible values and the set $\mathcal{J}$ of consistent evaluations represents the *evaluation problem* under consideration.

To see that this model can capture not only disagreements about the values of the variables but also disagreements about the constraints between the variables, note that one could include in X a further variable c (for “constraint”), which takes the value $c = 1$ if the constraint between x , y , and z is $z = x \times y$ and the value $c = 2$ if that constraint is, for example, $z = x + y$. An evaluation J (a function from X into V) would then count as consistent (and thereby be included in $\mathcal{J}$) if

either $c = 1$ and $J(z) = J(x) \times J(y)$

or $c = 2$ and $J(z) = J(x) + J(y)$.

This can be generalized to capture any possible disagreements about the constraints between different variables.

As before, an n -tuple of evaluations across the individuals, $\langle J_1, \dots, J_n \rangle$, is called a *profile of evaluations*, and an *aggregation rule* is a function, F , that maps each profile $\langle J_1, \dots, J_n \rangle$ in some domain of admissible profiles to a collective evaluation $J = F(J_1, \dots, J_n)$. For any issue or variable $x \in X$, $J(x)$ denotes the group's evaluation of that issue/variable x . We have already mentioned several examples of aggregation rules, such as the plurality, averaging, and median rules etc. These can clearly be extended

from one issue or variable to many. The plurality rule in fact admits two distinct extensions. One could either define the *collective value for each issue/variable* as the value most often assigned to it by the individuals (the *issuewise plurality rule*). Or one could define the *collective evaluation in its entirety* as the evaluation that is submitted most often (the *global plurality rule*).³ Moreover, it may make sense to define an aggregation rule that uses different criteria for producing the collective judgments for different issues or variables, for instance averaging for issues such as x (productivity) and y (employment) in the macroeconomic example and the plurality criterion for issues such as c (the relevant constraint). For a distance-based rule, finally, one must introduce a distance function on the set $\mathcal{J}$ of consistent evaluations; for each profile, the collective evaluation is then one that minimizes the sum-total distance from the individual evaluations.⁴

Can we find compelling aggregation rules that avoid the problem of collective inconsistency? We will now introduce some basic desiderata that one might expect a good aggregation rule F to meet (inspired by social-choice and judgment-aggregation theory, e.g., in [2, 14, 33, 34, 37]). Later, one might consider relaxing some of them.

Universality: Any profile of consistent evaluations is admissible as input to F . Formally, the domain of F is the set of all possible profiles of consistent evaluations $\mathcal{J}^n$.

Collective rationality: The output of F is a consistent evaluation. Formally, for any admissible profile $\langle J_1, \dots, J_n \rangle$, $F(J_1, \dots, J_n) \in \mathcal{J}$.

Anonymity: F treats different evaluators equally. Formally, $F(J_1, \dots, J_n) = F(J'_1, \dots, J'_n)$ whenever two admissible profiles $\langle J_1, \dots, J_n \rangle$, $\langle J'_1, \dots, J'_n \rangle$ are permutations (reorderings) of each other.

For the next desideratum, we need a preliminary definition. For any two evaluations J and J' and any subgroup $C \subseteq N$, we say that J *supported by C beats J' supported by $N \setminus C$* if, for the profile $\langle J_1, \dots, J_n \rangle$ in which all $i \in C$ have $J_i = J$ and all $i \notin C$ have $J_i = J'$, the collective evaluation $F(J_1, \dots, J_n)$ is J (provided this profile is admissible). In such a profile, the set of individuals N is divided into two subgroups, one of which (C) supports evaluation J while the other ($N \setminus C$) supports evaluation J' .

Pairwise neutrality: The aggregation rule is neutral between any two evaluations between which the group is divided. Formally, for any consistent evaluations J, J' and any subgroup $C \subseteq N$, if J supported by C beats J' supported by $N \setminus C$, then J' supported by C beats J supported by $N \setminus C$.

³In each case, one might need a tie-breaking criterion.

⁴For each profile $\langle J_1, \dots, J_n \rangle$, the collective evaluation J must be chosen to minimize $\sum_{i=1}^n d(J, J_i)$. To be *well-behaved*, a distance function $d : \mathcal{J} \times \mathcal{J} \rightarrow [0, \infty)$ must satisfy (i) for all J, J', J'' , $d(J, J') + d(J', J'') \geq d(J, J'')$, (ii) for all J, J' , $d(J, J') = 0 \Leftrightarrow J = J'$, and (iii) for all J, J' , $d(J, J') = d(J', J)$.

Issuewise independence: The collective evaluation for each issue/variable depends only on individual evaluations for that issue/variable. Formally, for any admissible profiles $\langle J_1, \dots, J_n \rangle$, $\langle J'_1, \dots, J'_n \rangle$ and any issue/variable $x \in X$, if $J_i(x) = J'_i(x)$ for all $i \in N$, then $J(x) = J'(x)$, where $J = F(J_1, \dots, J_n)$ and $J' = F(J'_1, \dots, J'_n)$.

Reinforcement monotonicity: The collective evaluation does not change if an additional individual comes to accept it as his or her own evaluation. Formally, for any individual $i \in N$ and any admissible profile $\langle J_1, \dots, J_n \rangle$, if $J'_i = F(J_1, \dots, J_n)$, then $F(J_1, \dots, J'_i, \dots, J_n) = F(J_1, \dots, J_n)$ (assuming $\langle J_1, \dots, J'_i, \dots, J_n \rangle$ is also admissible).

The examples of aggregation rules mentioned earlier satisfy several, but not all, of these desiderata. All of them – plurality, median, quantile, average, distance-based, and dictatorial rules – satisfy universality. All satisfy pairwise neutrality, with the exception of quantile rules distinct from the median rule.⁵ All except the global plurality and distance-based rules satisfy issuewise independence. All of the given rules further satisfy anonymity, except for dictatorial ones, and all satisfy reinforcement monotonicity, except for the averaging rule.⁶ However, only dictatorial rules, the global plurality rule, and distance-based rules generally satisfy collective rationality. Can we find aggregation rules that satisfy all of the desiderata at once? We will now see that the answer is negative whenever an evaluation problem involves connections between three or more issues or variables. We call such evaluation problems *non-simple*.⁷

Theorem 1: For any non-simple evaluation problem, there exists no aggregation rule satisfying universality, collective rationality, anonymity, pairwise neutrality, issuewise independence, and reinforcement monotonicity.

This theorem stands in the tradition of earlier impossibility theorems in binary and many-valued judgment-aggregation theory [14, 22, 23, 24, 33, 37, 40, 44]. However, unlike earlier theorems (such as [23, 24, 12]), it applies to a wide class of realistic evaluation problems with non-binary issues/variables and not merely to ones with highly structured interconnections between the issues/variables. Our theorem applies, for instance, to our macroeconomic example, where x , y , and z are related via the constraint $z =$

⁵Distance-based rules satisfy pairwise neutrality, provided distance is well-behaved.

⁶This includes well-behaved distance-based rules. To see that averaging violates reinforcement monotonicity, suppose half of the individuals assign the value 10 to x and the other half 20, so the average value is 15. If one individual now changes the evaluation for x from 10 to 15, the average will change.

⁷To define *non-simplicity* formally, let a *partial evaluation* be a function from some subset $X' \subseteq X$ to V . A partial evaluation is *consistent* if it can be extended to a consistent evaluation in $\mathcal{J}$, and *inconsistent* if not. It is *minimal inconsistent* if it is inconsistent but becomes consistent when restricted to any proper subset of X' . An evaluation problem, given by X , V , and $\mathcal{J}$, is *non-simple* if it admits a minimal inconsistent partial evaluation of more than two issues (i.e., with $|X'| > 2$). This concept generalizes a related concept for binary judgment aggregation discussed in [14, 22, 37].

$x \times y$, and to many other cases where one variable or issue is a function of two or more others. In practice, therefore, when a group seeks to aggregate judgments on several variables or issues, it has no choice but to give up at least one of the six desiderata. Arguably, universality, collective rationality, anonymity, and pairwise neutrality are hard to drop, because they are each very plausible and undemanding. Issuewise independence and reinforcement monotonicity are more contestable. Why, for instance, should the collective judgments on variable z depend only on the individual judgments on z , as issuewise independence demands, and not on the individual judgments on x and y , which seem relevant too? And why should we rule out averaging just because it violates reinforcement monotonicity? After all, averaging may seem independently plausible.

As we will now see, a rationale for requiring issuewise independence and reinforcement monotonicity comes from the significance of these two desiderata for tackling our second challenge, namely the challenge of manipulation.

4. The challenge of manipulation

When we aggregate several individuals' judgments on some variables or issues into collective judgments, we would like the individuals to have an incentive to reveal their judgments truthfully [27, 41]. For instance, we would not like the members of a committee or expert panel to misrepresent their judgments for strategic purposes. Aggregation by averaging, for instance, may incentivize some individuals to misrepresent their judgments. Suppose an expert panel is estimating the value of some variable, and the panel uses the averaging rule. If one of the members thinks the others' estimates are too low, this member can move the average estimate in his or her preferred direction by reporting an estimate that is higher than what he or she truly believes. And if the member thinks the others' estimates are too high, he or she can move the average estimate in the preferred direction by reporting a lowered estimate. In each case, the member will have "manipulated" the collective judgments by misrepresenting his or her own judgments.

To model strategic voting (drawing on the modelling strategy in [15, 18], cf. [5]), we will say that an individual has an *opportunity to manipulate* the collective evaluation for a particular issue/variable x at a particular profile of evaluations if the following holds:

- If (holding others' evaluations fixed) this individual submits his or her sincere evaluation for all issues/variables, the collective evaluation of issue/variable x will differ from the individual's sincere evaluation of x , and
- if he/she suitably misrepresents his/her evaluation for some of the issues/variables,

the collective evaluation of issue/variable x will move in the direction of this individual's sincere evaluation of x .

For the notion of “moving in the direction of a given individual's evaluation” to be well-defined, we must assume that the set V of values is linearly ordered, so that we can make “less than” ($<$) or “greater than” ($>$) comparisons between different values. Formally, we now say that a given aggregation rule F is *manipulable* by some individual $i \in N$ on some issue/variable $x \in X$ at some admissible profile $\langle J_1, \dots, J_n \rangle$ if

- $J(x) \neq J_i(x)$, and
- there exists some J'_i (where $\langle J_1, \dots, J'_i, \dots, J_n \rangle$ remains admissible) such that

$$J'(x) \begin{cases} < J(x) & \text{if } J_i(x) < J(x), \\ > J(x) & \text{if } J_i(x) > J(x), \end{cases}$$

where $J = F(J_1, \dots, J_n)$ and $J' = F(J_1, \dots, J'_i, \dots, J_n)$. Averaging, for instance, is manipulable on any issue/variable x by any individual i who finds the average evaluation $J(x)$ too high or too low and who can then move the average in the desired direction by reporting an evaluation $J'_i(x)$ that is lower or higher than i 's sincere evaluation $J_i(x)$.

Now the desideratum of non-manipulability demands that none of the individuals should ever have such a manipulation opportunity.

Non-manipulability: F is not manipulable by any individual $i \in N$ on any issue/variable $x \in X$ at any admissible profile.

Can we find non-manipulable aggregation rules? Here is where the significance of issuewise independence and reinforcement monotonicity lies (for related results on binary judgment aggregation or probability aggregation, see [15, 18, 36]).

Theorem 2: For an aggregation rule F satisfying universality, non-manipulability implies issuewise independence and reinforcement monotonicity.

In a nutshell, if a rule violates independence or reinforcement monotonicity at a given issue, then some individual(s) can sometimes manipulate the collective evaluation of this issue in their direction, by misrepresenting their evaluation of that issue (if reinforcement monotonicity fails) or by misrepresenting their evaluation of other issues (if independence fails). The present theorem gives us necessary conditions for non-manipulability. In supplementary material, we derive the full necessary and sufficient conditions, which are slightly more demanding and more technical and which we here state only informally:

Issuewise independence: As before.

Approach monotonicity: The collective evaluation of an issue/variable does not change if some individual’s evaluation of that issue/variable comes closer to it.

Non-overshoot: If an individual who agrees with the collective evaluation of an issue/variable at a given profile unilaterally changes his or her evaluation of that issue/variable, the collective evaluation does not move more strongly in the same direction.

The median and quantile rules, unlike the averaging rule, satisfy all of these conditions, and so they are non-manipulable (as previously noted for the single-issue case in [3, 4]). For example, none of the participants in Galton’s experiment could have moved the median weight estimate in a preferred direction by submitting a misrepresented weight estimate. The participants each had the best chance of bringing about a median estimate in line with their own estimate by revealing what they genuinely thought. Quantile rules share this non-manipulability, but unlike the median rule, they are not pairwise neutral. The median rule thus emerges once again as potentially attractive.

The observation that issuewise independence and reinforcement monotonicity are necessary for non-manipulability, however, also heralds some bad news. As we have seen, the median rule, which satisfies these desiderata, does not guarantee consistent collective judgments. This problem generalizes: if we wish to attain collective rationality and non-manipulability, we run into an impossibility result in the tradition of earlier social-choice-theoretic impossibility results on strategic voting [5, 15, 18, 27, 41, 36].

Theorem 3: For any non-simple evaluation problem, there exists no non-manipulable aggregation rule satisfying universality, collective rationality, anonymity, and pairwise neutrality.

This theorem is a corollary of Theorems 1 and 2. Since the last four desiderata may be non-negotiable in many committees and decision-making groups, it may be impossible in practice to achieve non-manipulability. When judgments are to be aggregated on three or more issues or variables with some constraints between them, some individuals may have an opportunity to manipulate the collective judgments by misrepresenting their own judgments. Whether they will always act on such an opportunity is a different matter. This will depend on how much information these individuals have about the judgments of others and whether they care only about the outcome or also about their truthfulness. However, the designers of mechanisms for aggregating judgments should be aware of this defect of many real-world aggregation rules – a defect that echoes a classic impossibility result in social choice theory, the so-called Gibbard-Satterthwaite theorem [27, 41], and related results concerning the manipulability of binary judgment aggregation [15, 36].

5. The challenge of truth-tracking

The goal is often to arrive at collective judgments that not just reflect the individual judgments but also “track the truth” [6, 21, 28, 32, 35, 38]. In Galton’s example, the aim is to estimate the true weight of the ox. Expert panels similarly seek to track the truth. In an animal group, the aim is to navigate in the right direction, which leads to food, safety, or a good nest-site. To analyze this demand of “track-tracking”, we will now suppose that, for each issue/variable, there is an independent fact of the matter as to what its true value is. For example, there is a fact about the true weight of the ox, the current level of atmospheric CO₂, the true GDP under a given economic scenario, or which travel direction leads to a desired destination. Of course, that fact is often hard to know, but this is why we require groups of individuals to judge what it is.

To model the truth-tracking performance of different aggregation rules, we will consider potentially infinitely many individuals $i = 1, 2, 3, \dots$ from whom different-sized groups of evaluators could be drawn. For any number n , the first n individuals will then form a group of size n . We retain our earlier definitions of the set X of variables or issues under consideration, the set V of possible values, and the consistency constraints encoded by $\mathcal{J}$. However, we now define an *aggregation rule* as a function that maps any admissible profile of evaluations $\langle J_1, \dots, J_n \rangle$ of any length n to a collective evaluation $J = F(J_1, \dots, J_n)$. Technically, F has the “variable-group-size universal domain”, $\bigcup_{n=1}^{\infty} \mathcal{J}^n$. Our earlier examples of aggregation rules can easily be redefined in this variable-group-size way.⁸

We model the facts to be tracked and the judgments about these facts as random variables relative to some underlying probability space with probability measure P . The facts can be modelled as a random variable $\mathbf{S}$, which we will call the *situation* and which encodes the factually true values of the issues or variables in X . As the situation fixes the true evaluation of those issues/variables, we assume that $\mathbf{S}$ takes values from the set $\mathcal{J}$. For each issue/variable x , $\mathbf{S}(x)$ represents the true value of x . (Later we will generalize the definition of $\mathbf{S}$.) We further assume that every value S taken by $\mathbf{S}$ – for short, every *situation* – has positive probability, i.e., $P(S) > 0$. The individual evaluations can be modelled as random variables too, namely $\mathbf{J}_1, \mathbf{J}_2, \mathbf{J}_3, \dots$, taking values in $\mathcal{J}$.

This model allows us to represent, for instance, the probability that individual i ’s judgment about the situation is correct, $P(\mathbf{J}_i = \mathbf{S})$, or the probability that i ’s judgment on issue or variable x is too low or too high, $P(\mathbf{J}_i(x) < \mathbf{S}(x))$ and $P(\mathbf{J}_i(x) > \mathbf{S}(x))$. We thus have a framework for investigating the probability of obtaining correct collective

⁸For aggregation rules that are defined only for an odd group size n , like the median rule, the variable-group-size universal domain must be defined as $\bigcup_{n=1,3,5,\dots} \mathcal{J}^n$.

judgments under various aggregation rules and various scenarios.

For our first baseline result, we make the idealizing assumptions that the individuals are *independent*, *judgmentally homogeneous*, and *competent* on every issue/variable. *Independence* means that the random variables $\mathbf{J}_1, \mathbf{J}_2, \dots$ are probabilistically independent, given the situation $\mathbf{S}$. *Judgmental homogeneity* means that these random variables are identically distributed, given $\mathbf{S}$. *Competence* on an issue/variable x means that each individual i has a probability greater than $\frac{1}{2}$ of not underestimating the value of x and also a probability greater than $\frac{1}{2}$ of not overestimating that value, regardless of the situation that obtains.⁹ In an idealized version of Galton’s experiment, the participants estimating the weight of the ox might be independent, judgmentally homogeneous, and competent. They each have a probability greater than $\frac{1}{2}$ of not overestimating the weight of the ox, and also a probability greater than $\frac{1}{2}$ of not underestimating it. Note that this permits a low probability of judging the weight with perfect accuracy. Under these idealized assumptions, the following holds:

Theorem 4: If individuals are independent, judgmentally homogeneous, and competent on every issue/variable, the median rule is increasingly accurate and asymptotically infallible on every issue/variable.

Here, being *increasingly accurate* on some issue/variable x means that the probability that the aggregation rule evaluates this issue correctly increases with increasing group size n ,¹⁰ and being *asymptotically infallible* on some issue/variable x means that this probability converges to 1 as n tends to infinity. Mathematically, the present theorem is a multi-issue version of a jury theorem for grade aggregation, where different evaluators each assign grades such as A, B, C etc. to a given option and the collective grade is defined as the median grade [35].¹¹

Theorem 4 can be viewed as a generalization of the classic “jury theorem” that goes back to Nicolas de Condorcet in the 18th century [28, 32, 21]. Condorcet had shown that,

⁹Formally, we define individual i ’s *non-underestimation competence* on issue/variable x as $\mathbf{C}_{i,x}^{\geq} = P(\mathbf{J}_i(x) \geq \mathbf{S}(x) | \mathbf{S})$ and i ’s *non-overestimation competence* on x as $\mathbf{C}_{i,x}^{\leq} = P(\mathbf{J}_i(x) \leq \mathbf{S}(x) | \mathbf{S})$. Both are random variables, because the probability of not underestimating or overestimating the value of a given issue/variable can vary from situation to situation. *Competence* on issue/variable x then means that $\mathbf{C}_{i,x}^{\geq} > \frac{1}{2}$ and $\mathbf{C}_{i,x}^{\leq} > \frac{1}{2}$, regardless of the situation.

¹⁰Formally, this probability is $P(F(\mathbf{J}_1, \dots, \mathbf{J}_n)(x) = \mathbf{S}(x))$. The increase is *weak*, meaning that the probability either increases or remains equal as n goes up. Furthermore, to avoid complications due to ties, the group size n is restricted to odd numbers (1, 3, 5, ...).

¹¹The latter result (from [35]) can be extended to quantile rules, which outperform the median rule when individuals have an upside or downside bias in their evaluations [29]. In [29], individuals’ non-overestimation and non-underestimation competence need not both exceed $1/2$, as long as their sum exceeds 1. The median rule must then be replaced by a p -quantile rule with p calibrated so as to compensate for any upside or downside bias in individual evaluations. Our present theorem could be analogously extended to quantile rules, so as to allow for individual evaluation biases.

in a collective decision on a single yes/no question (such as whether a defendant in a trial is guilty) with decision-makers who make correct judgments with independent and identical probabilities greater than $\frac{1}{2}$, the probability of a correct majority judgment increases and converges to 1 as the group size increases. Our result extends this to non-binary judgments (in line with [35]) and to any number of issues/variables. While we focus on the median rule here, we briefly address the truth-tracking properties of averaging in online supplementary material. (Recall again that averaging has several disadvantages compared to the median rule: it rests on more demanding measurability assumptions, is not robust to outlier judgments, and violates non-manipulability.)

A consequence of Theorem 4 is that, under its idealized assumptions, the probability of an inconsistency in the collective judgments made by the median rule (i.e., a breach of the constraints encoded by $\mathcal{J}$) converges to 0 as n tends to infinity, because the truth that is being accurately tracked is consistent. So, the median rule seems to be able to address *all* three challenges we have discussed, at least asymptotically: the challenges of collective inconsistency, manipulation, and truth-tracking.

Unfortunately, however, the assumptions of Theorem 4 are very unrealistic. To begin with, the assumption of judgmental homogeneity is unrealistic. If there is heterogeneity across the individuals, they may not all be equally competent. This limitation, however, can still be addressed relatively easily. The conclusion of increasing accuracy in Theorem 4 can be restored by introducing randomness about the identity of additional group members; and the conclusion of asymptotic infallibility continues to hold under the assumption that individuals are competent *on average* on every issue/variable x , i.e., that the average probability of not underestimating the value of x and the average probability of not overestimating it both exceed $\frac{1}{2}$, regardless of the situation that obtains.¹² We can thus treat judgmental homogeneity as a convenient simplifying assumption.

The other limitations of Theorem 4, however, are more serious (for reasons similar to those discussed in [13, 20, 21]). First of all, different individuals' judgments are not normally independent from each other in the strong sense that, holding fixed the true values of the relevant issues or variables, there are no probabilistic dependencies between them. The individuals may have communicated, relied on shared or partially shared evidence, and be affected by shared environmental conditions. In consequence, there may be common causes affecting all of their judgments. Secondly, the assumption of issuewise competence is unrealistic, even if we make the simplifying assumption of

¹²Using the formalism from a previous footnote, the average non-underestimation competence on issue/variable x in a given group of size n is $\frac{1}{n} \sum_{i=1}^n \mathbf{C}_{i,x}^{\geq}$, and the average non-overestimation competence is $\frac{1}{n} \sum_{i=1}^n \mathbf{C}_{i,x}^{\leq}$. *Competence on average* on x means that $\frac{1}{n} \sum_{i=1}^n \mathbf{C}_{i,x}^{\geq}$ and $\frac{1}{n} \sum_{i=1}^n \mathbf{C}_{i,x}^{\leq}$ each exceed $\frac{1}{2}$ in the limit (or more generally limit inferior) as $n \rightarrow \infty$, regardless of the value of $\mathbf{S}$.

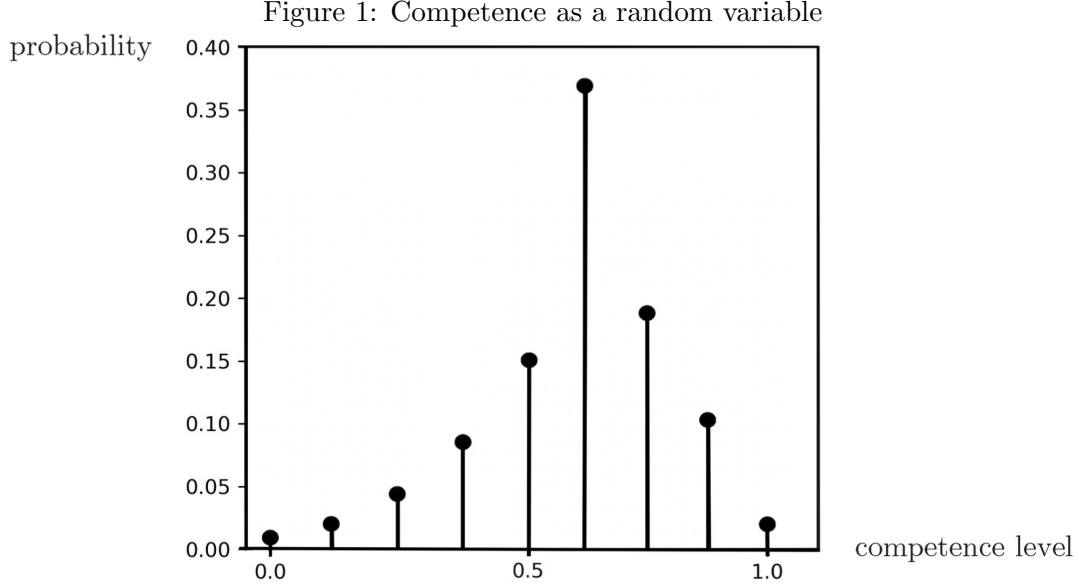
homogeneity. If the situation includes common causes and other factors influencing the individual judgments, it is unrealistic to assume that *regardless of the situation* each individual *always* has a probability greater than $\frac{1}{2}$ of not underestimating the value of any issue/variable and also a probability greater than $\frac{1}{2}$ of not overestimating it. The evidence or other factors may have a misleading influence and make it sometimes more likely that individuals under- or overestimate the value of some issues or variables.

To address those limitations, we must amend our model. First, we must specify the situation $\mathbf{S}$ more richly, so as to include not only the true values of the variables or issues but also other possible common causes, from shared deliberation and shared evidence to any other shared influences that may compromise independence in the original sense. Formally, we introduce hypothetical issues or variables beyond those in the set X (building on the modelling strategy in [20, 21]). So, we now introduce a possibly larger set X^+ of issues or variables (a superset of X) and a possibly larger set V^+ of values (a superset of V) and define a *value configuration* to be a function from X^+ to V^+ . We write $\mathcal{J}^+$ to denote the (non-empty) set of all value configurations that are deemed possible. We assume that a value configuration is possible (i.e., in $\mathcal{J}^+$) only if its restriction to the issues or variables in X is consistent (i.e., in $\mathcal{J}$). The *situation* in this generalized sense is a random variable $\mathbf{S}$ with values in $\mathcal{J}^+$.¹³ With this generalization in place, independence becomes plausible again: if we conditionalize on all common causes of the different individuals' judgments, there are no further dependencies between them.

Secondly, we must capture the fact that some situations may be misleading in the sense that they make it more likely that individuals over- or underestimate the value of some issues or variables. In what follows, we require only a *tendency* for individuals to be competent on those issues or variables, not a *guaranteed* competence [20, 21]. The important point is that the individuals' probabilities of not overestimating or not underestimating the value of a given issue or variable – for short, their “competence levels” – are not constant but random variables that depend on the situation. In some situations, competence levels may be higher, in others lower. We say that individual i *tends to be competent* on an issue or variable x if i 's probability of not underestimating the value of x *tends to exceed* $\frac{1}{2}$ and i 's probability of not overestimating that value also *tends to exceed* $\frac{1}{2}$. Formally, this means that each of these probabilities (“competence levels”) takes the value $\frac{1}{2} + \epsilon$ with at least as high a probability as the value $\frac{1}{2} - \epsilon$, for every $\epsilon > 0$. Figure 1 (adapted from [21]) shows an example of what such a probability distribution over competence levels might look like. In this example, nine competence

¹³Again, we assume that $P(S) > 0$ for each situation S , i.e., each value of $\mathbf{S}$. What we previously defined is the special case where $X = X^+$, $V = V^+$, and $\mathcal{J} = \mathcal{J}^+$.

levels are possible (0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1), and each competence level above $\frac{1}{2}$ is more probable than the opposite level below $\frac{1}{2}$. A *tendency to be*



competent thus permits the occurrence of situations in which the individuals’ competence levels are less than $\frac{1}{2}$. In Figure 1, for instance, it can easily happen that individual competence is only 0.375 or even lower. But such a competence level is less likely than the opposite competence level of 0.625 or higher. The previous definition of “guaranteed” competence is the special case where competence levels are *always* above $\frac{1}{2}$.

Theorem 5: If individuals are independent and judgmentally homogeneous, the median rule is increasingly accurate on any issue/variable on which individuals tend to be competent, and asymptotically fallible on it unless individuals are always competent on it.

This theorem (which generalizes a result from [20] to non-binary evaluations) captures a less unrealistic form of “collective intelligence”. First, the theorem’s independence assumption is stated against the background of a sufficiently rich specification of the situation, which can include common causes of the individual judgments. Second, the individuals only need to have a tendency to be competent – and only on certain issues or variables. The median rule performs well on those issues, in the sense that larger groups of decision-makers are more reliable than smaller groups. But we can no longer derive asymptotic infallibility; the collective reliability is typically bounded above by some threshold below 1, due to a non-zero probability that the situation is misleading.

Crucially, however, the median rule retains its key merits when restricted to issues or variables on which individuals tend to be competent.

6. Concluding remarks

Groups must often aggregate judgments on non-binary issues or variables. The problem arises not only in human decision-making committees or expert panels, but also in animal groups choosing one option among many: travel directions, grazing spots, or nest sites.

It is surprisingly difficult to find an aggregation rule that simultaneously addresses the challenges of inconsistency, manipulation, and truth-tracking. While the challenges of inconsistency and manipulation are arguably most relevant to the human case, the truth-tracking challenge is highly relevant for animal group decisions too.¹⁴ Overall, the median rule performs reasonably well: it is non-manipulable and has increasing truth-tracking accuracy on those issues/variables on which individuals tend to be competent, such as the weight of an ox or the best travel route to some destination. But it does not generally overcome the challenge of inconsistency, except asymptotically when the unrealistic assumptions of our initial Condorcet-inspired jury theorem are met.

How, then, can we achieve collective intelligence through aggregation? We briefly consider two proposals that retain some of the median rule’s merits while avoiding inconsistencies across different issues. One proposal is to drop the universality requirement and to restrict the domain of admissible inputs of the median rule to those profiles $\langle J_1, \dots, J_n \rangle$ that have the special property of *comonotonicity* with respect to the given linear order on V : for all individuals i, j , either (i) for every $x \in X$, $J_i(x) \leq J_j(x)$, or (ii) for every $x \in X$, $J_i(x) \geq J_j(x)$. Comonotonicity means that, for any two individuals i and j , the move from i ’s to j ’s evaluation goes in the same direction for every issue. When restricted to comonotonic profiles of consistent evaluations, the median rule (for odd n) satisfies collective rationality, while still satisfying all the other desiderata from Theorem 1. The reason is that, for any comonotonic profile $\langle J_1, \dots, J_n \rangle$, there exists an individual i whose evaluation J_i lies in the middle on *every* issue, and the median rule selects this evaluation, J_i , which is consistent. For a similar reason, any quantile rule is also consistent for comonotonic profiles, but of course not generally pairwise neutral.¹⁵

¹⁴The inconsistency challenge is less relevant to non-human animals to the extent that animal group decisions do not involve multiple interconnected issues. It is an intriguing question whether there could be “strategic voting” in animal groups that use a manipulable aggregation rule, such as averaging.

¹⁵When all issues/variables are binary, comonotonicity becomes a special case of a known property called “unidimensional alignment” [31]. The full generalization of unidimensional alignment in the multi-valued case is more complicated. A profile $\langle J_1, \dots, J_n \rangle$ is *unidimensionally aligned* if the individuals can be linearly ordered on an axis from “left” to “right” such that, for each issue/variable $x \in X$, *either* $J_i(x)$

A second proposal is to give up issuewise independence and to aggregate judgments in a “premise-based way” (building on similar ideas in binary judgment and probability aggregation; e.g., [6, 17, 19, 30, 33]). A *premise-based aggregation rule* works as follows:

1. We designate some issues/variables as “premises”, such as x (productivity) and y (employment level) in our earlier example.
2. We aggregate the judgments for those issues/variables via the median criterion.
3. We use some relevant inference rules, such as the constraint $z = x \times y$, to derive the collective judgments for other issues/variables, such as z (GDP).

The resulting judgments will always be consistent. This approach seems plausible in particular when the issues/variables can be subdivided into independent and dependent variables (or “basic” and “derived” variables) and when the judgments on the former are given priority over those on the latter. The idea is that the values of the dependent (“derived”) variables are a function of the values of the independent (“basic”) variables (the “premises”), and we can obtain collective estimates of the dependent variables by aggregating the individuals’ estimates of the independent variables and then applying the relevant function to derive the resulting values of the dependent variables.

The premise-based approach makes the collective judgments non-manipulable on the “premise” variables, and it has a reasonable truth-tracking performance on them as long as the individual evaluators tend to be competent on them. The downside is that the collective judgments will be sensitive to the choice of the premise issues or variables. This may be no problem when there is a clear criterion for distinguishing between premises and non-premises, for instance when it is clear which variables count as independent and which count as dependent. But if we are dealing with interconnected variables in a complex system, such as the economy or the climate, such a clearcut criterion may not exist. Decision-making groups should be aware of the challenges identified here.

Finally, our analysis can accommodate two distinct cases of many-valued judgments: first, cases where the issues/variables to be estimated are non-binary (such as the weight of an ox, the level of atmospheric CO₂, the travel direction); and second, cases where the issues/variables themselves are fundamentally binary (such as whether a particular scientific hypothesis is true or whether a particular event will happen) but the evaluators’

weakly increases as i moves from left to right or $J_i(x)$ weakly decreases as i moves from left to right. Comonotonicity is the special case that leaves out one of the two disjuncts. For odd n , the median rule is also consistent for every unidimensionally aligned profile $\langle J_1, \dots, J_n \rangle$. By contrast, other quantile rules are generally consistent only for comonotonic profiles. We thank a referee for prompting us to consider comonotonicity as an escape route from the impossibility result of Theorem 1.

attitudes towards them are non-binary (such as degrees of belief or degrees of desire or approval). Although our discussion has focused mostly on the first case, our model and results can equally accommodate the second. This flexibility, along with the ubiquity of general aggregation problems in many decision-making groups – human and non-human – should make our analysis widely applicable.

References

- [1] Aczél, J., and Wagner, C. (1980). A characterization of weighted arithmetic means. *SIAM Journal on Algebraic and Discrete Methods* 1: 259–260.
- [2] Arrow, K. (1951/1963). *Social Choice and Individual Values*. New York: Wiley.
- [3] Balinski, M., and Laraki, R. (2007). A theory of measuring, electing, and ranking. *Proceedings of the National Academy of Sciences* 104: 8720–8725.
- [4] Balinski, M., and Laraki, R. (2011). *Majority Judgment: Measuring, Ranking, and Electing*. Cambridge/MA: MIT Press.
- [5] Basili, M., Savaglio, E., and Vannucci, S. (2021). Strategy-Proof Aggregation of Approximate and Imprecise Judgments. Working paper, University of Siena.
- [6] Bovens, L., and Rabinowicz, W. (2006). Democratic answers to complex questions – an epistemic perspective. *Synthese* 150: 131–153.
- [7] Claussen, C.A., and Røisland, Ø. (2005). Collective economic decisions and the discursive dilemma. Working paper, Norges Bank.
- [8] Claussen, C.A., and Røisland, Ø. (2010). A quantitative discursive dilemma. *Social Choice and Welfare* 35: 49–64.
- [9] Conradt, L., and C. List (2009). Group decisions in humans and animals: a survey. *Philosophical Transactions of the Royal Society B* 364: 719–742.
- [10] Conradt, L., and T. J. Roper (2003). Group decision-making in animals. *Nature* 421: 155–158.
- [11] Couzin, I. D., J. Krause, N. R. Franks, and S. A. Levin (2005). Effective leadership and decision-making in animal groups on the move. *Nature* 433: 513–516.
- [12] Dietrich, F. (2015). Aggregation theory and the relevance of some issues to others. *Journal of Economic Theory* 160: 463–493.

- [13] Dietrich, F., and List, C. (2004). A model of jury decisions where all jurors have the same evidence. *Synthese* 142: 175–202.
- [14] Dietrich, F., and List, C. (2007). Arrow’s theorem in judgment aggregation. *Social Choice and Welfare* 29: 19–33.
- [15] Dietrich, F., and List, C. (2007). Strategy-proof judgment aggregation. *Economics and Philosophy* 23: 269–300.
- [16] Dietrich, F., and List, C. (2010). The aggregation of propositional attitudes: towards a general theory. *Oxford Studies in Epistemology* 3: 215–234.
- [17] Dietrich, F., and List, C. (2017). Probabilistic opinion pooling generalized. Part two: the premise-based approach. *Social Choice and Welfare* 48: 787–814.
- [18] Dietrich, F., and List, C. (2025). The impossibility of non-manipulable probability aggregation. *British Journal for the Philosophy of Science*, forthcoming.
- [19] Dietrich, F., and Mongin, P. (2010). The premise-based approach to judgment aggregation. *Journal of Economic Theory* 145: 562–582.
- [20] Dietrich, F., and Spiekermann, K. (2013). Epistemic Democracy with Defensible Premises. *Economics and Philosophy* 29: 87–120.
- [21] Dietrich, F., and Spiekermann, K. (2023). Jury Theorems. In E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition).
- [22] Dokow, E., and Holzman, R. (2010). Aggregation of binary evaluations. *Journal of Economic Theory* 145: 495–511.
- [23] Dokow, E., and Holzman, R. (2010). Aggregation of non-binary evaluations. *Advances in Applied Mathematics* 45: 487–504.
- [24] Duddy, C., and Piggins, A. (2013). Many-valued judgment aggregation: Characterizing the possibility/impossibility boundary. *Journal of Economic Theory* 148: 793–805.
- [25] Elster, J., and Landemore, H. (eds.) (2012). *Collective wisdom: Principles and Mechanisms*. Cambridge: Cambridge University Press.
- [26] Galton, F. (1907) Vox Populi. *Nature* 75: 450–451.

- [27] Gibbard, A. (1973). Manipulation of voting schemes: a general result. *Econometrica* 41: 587–601.
- [28] Grofman, B., Owen, G., and Feld, S. L. (1983). Thirteen theorems in search of the truth. *Theory and Decision* 15: 261–278.
- [29] Kilinc, B. (2025). A skewed jury theorem: more theorems in search of the truth. *Synthese* 206: 108.
- [30] Kornhauser, L. A., and Sager, L. G. (1986). Unpacking the court. *Yale Law Journal* 96: 82–117.
- [31] List, C. (2003). A possibility theorem on aggregation over multiple interconnected propositions. *Mathematical Social Sciences* 45: 1–13.
- [32] List, C., and Goodin, R. E. (2001). Epistemic Democracy: Generalizing the Condorcet Jury Theorem. *Journal of Political Philosophy* 9: 277–306.
- [33] List, C., and Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy* 18: 89–110.
- [34] List, C., and Puppe, C. (2009). Judgment aggregation: A survey. In P. Anand et al., *Oxford Handbook of Rational and Social Choice*. Oxford: Oxford University Press.
- [35] Morreau, M. (2021). Democracy without Enlightenment: A Jury Theorem for Evaluative Voting. *Journal of Political Philosophy* 29: 188–210.
- [36] Nehring, K., and Puppe, C. (2002). Strategyproof Social Choice on Single-Peaked Domains: Possibility, Impossibility and the Space Between. Working paper, UC Davis.
- [37] Nehring, K., and Puppe, C. (2010). Abstract Arrovian aggregation. *Journal of Economic Theory* 145: 467–494.
- [38] Pivato, M. (2017). Epistemic Democracy with Correlated Voters. *Journal of Mathematical Economics* 72: 51–69.
- [39] Prins, H. H. T. (1996). Selecting grazing grounds: a case of voting. In *Ecology and Behaviour of the African Buffalo*, pp. 218–236. London: Chapman & Hall.
- [40] Rubinstein, A., and Fishburn, P. C. (1986). Algebraic aggregation theory. *Journal of Economic Theory* 38: 63–77.

- [41] Satterthwaite, M. (1975). Strategy-proofness and Arrow's conditions. *Journal of Economic Theory* 10: 187–217.
- [42] Sen, A. K. (1977). On Weights and Measures: Informational Constraints in Social Welfare Analysis. *Econometrica* 45: 1539–1572.
- [43] Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science* 103: 677–680.
- [44] Wilson, R. (1975). On the theory of aggregation. *Journal of Economic Theory* 10: 89–99.