

Beyond Numbers: Majority Judgment for Collective Beliefs and Science

Thomas Boyer-Kassem*

Abstract

When attributing a claim to a group, should one consider that the larger number of agents who believe P rather than Q is epistemically relevant, and thus “go with the numbers”? The classical view says one should. The thesis defended here is one should not, as soon as agents are able to assign qualitative grades to propositions. At the group level, the aggregation should be made with the Majority Judgment rule. This paper extends it from preferences to belief and science, with significant consequences for the philosophy of expertise and testimony.

1 Introduction

On a disputed factual question, the number or the proportion of people within a group who believe P rather than not- P , or than Q or R (which are mutually exclusive), is widely believed to be of particular epistemic relevance. The fact that the *majority* of pupils thinks that the surprise exam will take place next week is generally taken to justify the assertion that “the class thinks that the surprise exam will take place next week”. Or knowing that *most* doctors believe that the vaccine does not cause autism is usually considered as relevant premises to infer the safety of the vaccine. A near consensus would give an even superior argument for this conclusion — numbers matter. Philosophers have discussed and qualified this view: if some people or experts have not formed their beliefs independently, they should be counted only once; or if they are not equally reliable, or do not have the same justified confidence, they should be counted accordingly. Philosophers of expertise and testimony agree on the general principle that “numbers matter” or that one should “go with the numbers”, i.e. that all other things being equal, the more people believe that P , the more it gives weight in the testimony in favor of P (e.g. Goldman

*Université de Poitiers, MAPP, Département de Philosophie, 36 rue de la Chaîne, F-86000 Poitiers, France. E-mail: thomas.boyer.kassem@univ-poitiers.fr

2001, Coady 2006, Elga 2010, Beatty and Moore 2010, Kelly 2011, Lackey 2013, Miller 2013).

In a similar spirit, in academic publications, countless papers (implicitly) justify discussing or endorsing some thesis because of the number, or the proportion, of studies or scholars who have defended it. As Beatty and Moore (2010, p. 200) note, “Formal votes may be rare in science, but scientists count other scientists—those for and those against—all the time. This makes sense epistemically.” Bourget and Chalmers (2014 p. 466) note that “In philosophical discussion it is inevitable that some views are presupposed and other views are the focus of attention and argument, while still others are ignored. At a given time in a given community, some views have the status of ‘received wisdom’. [...] Received wisdom is more likely to be determined by what *most people* believe *most people* believe.” (my emphasis)

Counting views is also common practice in scientific expert committees, who regularly advise governments or other institutions on various questions. Many of these expert committees are required to make their decisions, or adopt their reports, by means of a vote with majority rule: the Food and Drug Administration (FDA) in the United States, the European Medicine Agency (EMA), the European Environment Agency (EEA), the European Food Safety Agency (EFSA), the European Chemicals Agency (ECHA), the International Accounting Standards Board (IASB), and the scientific committees advising the European Commission. The fact that a majority of experts think likewise seems implicitly to be a sufficient reason for using the majority rule.

The idea that the number of people who believe P matters has famously received support from a mathematical argument due to Condorcet (1785). Assuming that there is a binary question with one correct answer, that n agents vote and decide by the majority rule, that each agent is more likely to vote for the correct answer than not, and that agents vote independently from each other, the Condorcet Jury Theorem states that larger groups have a larger probability of finding the correct answer, and that infinite groups are infallible. This result has been the starting point of the motto “the wisdom of the crowds” and of epistemic arguments for democracy (Dietrich and Spiekerman (2022)).

Summing up, when faced with mutually exclusive propositions P , Q , ..., is it epistemically relevant that a majority rather than a minority believes that P , or that there is a near consensus for P among the group? In short, are numbers epistemically relevant? The traditional answer is Yes — an explicit Yes according to philosophers of expertise or of testimony, an implicit one according to academics more generally or expert agencies. In this paper, I challenge this widespread view. More specifically, the phrasing “epistemically relevant” can apply to several cases:

- *reporting* a collective belief¹—when deciding which view is the group’s view, should members count how many people favor which view?
- externally *attributing* a belief to the group—should the view externally attributed to the group depend on the number of its members who believe such and such?
- *revising* one’s belief—when considering revising my belief, should I take into account the proportion of agents in the group who believe differently?
- *deferring* to some agents’ beliefs—when considering to defer to experts, should I take into account the proportion of experts who believe such and such?

For simplicity, examples shall be taken from the case of belief attribution, but arguments equally apply to other cases.

Let me be explicit about which problem is *not* considered and which assumptions are *not* made in this paper. The propositions faced by the group, or the question posed to it, are factual ones, typically “ X will occur today” *vs* “ X will occur tomorrow”. They are not preference-related ones, like “The best wine among these is V ” or “Who is to be elected for Presidency?”.² Furthermore, individuals are supposed to be faced with one question and several possible answers, or just several mutually exclusive propositions to be assessed, i.e., alternatives (not arbitrarily propositions that might be connected as in judgment aggregation). Individuals are not supposed to have mathematical, probabilistic assessments of the propositions, or point-like degrees of beliefs, and a Bayesian framework is not assumed. More specifically, it is supposed that not all agents have sharp probabilistic or otherwise numerical judgments on the propositions under study. Instead, agents are assumed to have qualitative views on the possible answers (e.g. “ P is a sound hypothesis”). Similarly, when discussing academic references to beliefs held in the community, I am concerned with qualitative mentions or discussions, not with quantitative meta-analyses.

Given this framework, my thesis is twofold. First, the number of individuals who believe some proposition is not epistemically relevant, contrarily to what many academic papers argue or to what academics or institutions usually implicitly assume. Considering whether “a majority believes that P ”, for instance, is not the right way to report a

¹For simplicity, the concept of “belief” is used throughout the paper, but any cognitive propositional attitude like “acceptance” could be used instead, or even emotive propositional attitudes like “desire”.

²The relevance of drawing a limit between pure epistemic and non-epistemic (value-laden) judgments has been questioned in philosophy (e.g. Rudner 1953, Douglas 2009). This subtlety is ignored as a first approximation but considered in Section 7. Morreau (2021) adopts a broader concept of truth and applies it to the selection of policies. Dietrich and Spiekermann (2022, sec. 1.2) use the concept of correctness-apt decisions, which include beliefs and actions which can be said to be factually correct.

collective belief. What matters, instead, are the qualitative assessments, or qualitative judgments, that all n agents make of all the k propositions (with the standard proviso: dependence between agents, for instance, should be taken into account). If agents were to submit their assessments on sheets of paper, one should not just collect bulletins on which each agent indicates the name of their best-evaluated proposition, but richer bulletins on which the agent evaluates each proposition by a grade (e.g. Very Good, Good, ...), for instance in a tabular with propositions in rows and grades in columns and in which the agent puts a mark on each row. What is epistemically relevant is that “for the group, P is the best proposition”, where “best proposition” is understood in the sense of Majority Judgment (Balinski and Laraki 2007, 2010, 2020a, 2020b), i.e. best evaluated by a majority of agents—to be detailed in Section 5. This makes up the second part of my thesis. We shall say that “the group believes that P ” when P is the best-ranked proposition with the Majority Judgment rule. My argument is an adaptation of those made by Balinski and Laraki in voting theory in favor of Majority Judgment. So far, Majority Judgment has been applied to two domains: elections and competitions, both in the field of social choice theory³ and not in epistemology. Elections and competitions primarily involve non-epistemic judgments or preferences (“How beautiful a jump is it?”, “To what extent would this candidate be a good President?”). I argue here that the reasons for which the majority rule should be abandoned in political science and social choice theory in favor of Majority Judgment can be adapted to the epistemic domain. Thus, the originality of the present paper lies in the fine-grained discussion of the relevance of extending the existing hypotheses, axioms and results to the domain of epistemology.

My criticism of the majority rule is distinct from those from both the judgment aggregation literature and the probability aggregation literature. In the former (for reviews, see e.g. List 2012, Pigozzi 2021 or Martini and Sprenger 2017), a proposition arrives as a conclusion of an argument, and agents have categorical beliefs. In the latter (e.g. Dietrich and List 2016), agents have probabilistic beliefs on logically connected propositions. No such hypotheses are made here. Compared to Morreau (2021), the focus is here not made on the probability of tracking the truth.

The paper is structured as follows. Section 2 introduces the model. Section 3 and 4 present two fundamental problems for the view that numbers matter: domination and the independence of irrelevant alternatives, respectively. Section 5 introduces the solution: Majority Judgment. Section 6 discusses objections. Section 7 draws the consequences of using Majority Judgment for the epistemic domain. Section 8 concludes.

³An exception is Laraki and Varloot (2022); but since they assume a different modeling framework with probabilistic estimates for agents, their results cannot be directly reused here.

2 Modeling — a scale of grades

Let us specify the details of the considered problem. Let $\mathcal{P} = \{P_1, P_2, \dots, P_k\}$ be a finite set of k propositions, with $k \geq 2$. These propositions may be competing hypotheses, theses, observations or different answers to an implicit question. For simplicity, it is convenient to assume that they are mutually exclusive, and they can thus be considered as alternatives.⁴ The propositions are factual ones, liable to be true or false. At most one of these answers is correct or true — maybe none is correct, as they may not exhaust the possibility space. Examples are $\mathcal{P} = \{\text{“the hurricane will hit the coast tonight”}, \text{“the hurricane will hit the coast tomorrow”}, \text{“the hurricane will never hit the coast”}\}$.

There is a finite group of n individuals ($n \geq 1$). The group may be defined in various ways: for instance, belonging to it may be common knowledge among its members, or an external observer may define the group as she wishes. Agents are assumed to have some *qualitative* beliefs about each proposition, like “ B is a very good hypothesis, A is really crazy”, not just “Believe B ”/“Does Not Believe A ”. Agents are not supposed to all have precise probability estimates, or precise numerical epistemic assessments, for all P_j .⁵ A typical case is expert judgment: as a specialist in nuclear power plants, Zach is able to assess the soundness of various hypotheses to explain the origin of the incident in the plant, but he does not have any probabilistic model for them, nor any sharp numerical value.

So that the qualitative opinions can be compared across agents, it is convenient to assume the following hypothesis:

(Hypothesis) Ordinal Scale of Grades: there exists a finite set of grades $\mathcal{G} = \{\alpha_1, \dots, \alpha_g\}$ (with $g \geq 2$), arranged as an ordinal scale with a total order, which forms a common language among the group. Each individual i

⁴Although nothing formally prevents from considering propositions that are not mutually exclusive, their interpretation should be made with great caution. For instance, consider the question “When is the eruption most probable?” and answers P_{MT} =“Monday or Tuesday”, P_T =“Tuesday” and P_W =“Wednesday”. Suppose the group assesses P_{MT} better than P_W , and P_W better than P_T . A naive reading could conclude that the group believes the eruption will most probably occur on Monday or Tuesday, and, pressed to guess the very day, would pick Monday — since Tuesday is ranked last. Yet, the group’s assessment is consistent with judging that the most probable day is Wednesday and the least probable one Monday (e.g., assuming sharp probability estimates, 25% for Monday, 30% for Tuesday, 35% for Wednesday, and 10% for other days).

⁵At least one agent does not have a precise numerical estimate for at least one proposition. If all agents had numerical (probability) estimates for all propositions, other axioms than those from Section 5, should be considered (cf. Dietrich and List 2016). A first step to bridge the two is discussed in Section 6.1. Note, however, that not all qualitative scales have the probability scale as an ideal precise limit. For instance, a scale used to assess the *quality* of a hypothesis (like “very good”, “good”, “medium”), or its fruitfulness, seems to be in that case.

assigns a grade from this scale to each proposition, defining a *grading function* $G_i : \mathcal{P} \rightarrow \mathcal{G}$ (let \mathbf{G} be the set of grading functions).

An ordinal scale means that the grades of the scale consist of letters, words, or other symbols, and that all grades can be meaningfully compared by: “equal to”, “greater than” or “less than” (cf. Balinski and Laraki 2010 chap. 8, Morreau 2016). In other words, the grades admit an ordering from lowest to highest. However, it is meaningless to add, subtract or average the grades (e.g., to say that α_2 is twice α_1), as they have no numerical meaning. The grades form a common language within the group: they are the same for all agents, and their meanings are the same for all agents.⁶ The words can be taken from common language and have clear, distinct meanings, or can be provided with a definition. For example, a scale to evaluate the likeliness of a scientific hypothesis may consist of the grades: Very Unlikely, Unlikely, Likely, Very Likely. The scale may be specific to the problem under study, like the Richter earthquake scale, or it may be generic, like: Poor, Acceptable, Good, Very Good, for instance to judge the overall quality of a hypothesis, not just its likeliness. A scale should contain at least two grades to convey some information. A greater number of grades brings greater precision but must be understood and used correctly by all. Grades can be given honestly by agents (they give the grade they privately believe to) or in a strategic way; the latter case is studied in Section 5.1 and it is simpler to assume until then that agents behave honestly.

The hypothesis Ordinal Scale of Grades is not very demanding, as agents just need to pick one grade from a given set. It does not ask for a precise numerical assessment like a probability, or a numerical interval. Is it nevertheless too demanding that *each* agent gives a grade for *each* proposition? The answer depends on the case. For a group of scientific experts advising a government, it is the very least that each expert has some qualitative opinion about the propositions under discussion. Similarly, it is reasonable to assume that scientists have graded opinions on their field’s main hypotheses. However, academic papers often focus on a single hypothesis, making it difficult to infer their stance on others. Consider now testimonial cases. Someone who testifies that “The car went through the red light” because he clearly remembers so would give the grade “Very good” to this proposition and the grade “Poor” to the proposition “The car went through the green light”. To sum up, the hypothesis Ordinal Scale of Grades makes sense for a substantive subset of the cases traditionally considered in epistemology: experts, researchers, and in many cases for laypeople.

⁶Cf. Balinski and Laraki 2010, chap. 8. As a contrast, Morreau (2021) interestingly studies the advantages that a diversity in grading standards can bring.

Let us consider the following objection: assuming that not all individuals have sharp probability estimates is unrealistic, and qualitative degrees of belief can and should be replaced with probabilistic assessments. My answer is that the epistemic assessment with the qualitative scale does not necessarily evaluate the *probability* or *likeliness* of the proposition. Some epistemic dimensions are not amenable to a probability (or sharp numerical) judgment (cf. footnote 5 p. 5). According to Kuhn (1977), scientists choose between theories using criteria or values like accuracy, consistency, scope, simplicity, fruitfulness. While accuracy may relate to probability (e.g. in a Bayesian spirit), others cannot: for instance, there is no uncontroversial quantitative measure of simplicity, in particular because there are several ways to understand simplicity itself, as Kuhn noted. This remark can be extended to propositions in general: our epistemic assessments are not limited to judging probabilities or likeliness, and it is relevant to consider a qualitative scale in a large range of cases.

3 A problem for counting views—domination

Following the Ordinal Scale of Grades hypothesis, each agent i assigns a grade to each proposition through a grading function $G_i \in \mathbf{G}$. Let a *grading profile* be the grading functions of all individuals, $(G_1, \dots, G_n) \in \mathbf{G}^n$ —it gathers the grades given by all individuals on all propositions. The next step is to aggregate this grading profile, so that the group itself can be said to select one proposition—let an *aggregation rule* be a function $f : \mathbf{G}^n \rightarrow \mathcal{P}$ —to rank propositions—let a *ranking rule* be a function $h : \mathbf{G}^n \rightarrow \mathbf{R}$, where \mathbf{R} is the set of binary relations \succeq among the propositions—or to assign grades to propositions—let a *grading aggregation rule* be a function $f : \mathbf{G}^n \rightarrow \mathbf{G}$.

With two propositions, an example of an aggregation rule is the majority rule: one says that “the group of experts believe that P ” in case at least a majority of experts believe that P . Another rule is a super-majority, which requires instead that at least m experts believe that P , with $m > n/2$. With more than two propositions, the majority rule can be generalized by selecting the proposition with the largest number of agents who believe it — it is called “going with numbers”, or “most votes count”, or “plurality voting”. Going with numbers can define both an aggregation rule and a ranking rule.

Is the majority rule a good aggregation rule? To assess aggregation or ranking rules, an axiomatic approach is often privileged. It consists of first formulating general desiderata before looking for rules that satisfy them. An intuitive requirement is that if proposition A receives on the whole better grades than proposition B — say that the grades of A *dominate* those of B —, then the ranking rule should rank A in front of B , and

agent	grade				vote
	Excellent	Good	Fair	Poor	
Alice			B	A	B
Bob		B	A		B
Carmen	A			B	A

Table 1: Grades given by scientists to hypotheses A and B , and scientists’ votes. The majority rule outputs B , while the Domination axiom requires A .

the aggregation rule should not select B . For instance, if hypothesis P receives better grades than hypothesis not- P , then the group should be said to believe P , not not- P . Formally, denote by $(\alpha_1, \dots, \alpha_n)$ the grades received by proposition A , reindexed through a permutation so that they are now ordered from the highest grade to the lowest (thus, not indexed according to the arbitrary numbers of the agents who give them). Similarly for proposition B , with $(\beta_1, \dots, \beta_n)$. The domination⁷ requirement can be formally stated as:

(Axiom) Domination: If the grades of A *dominate* those of B ,
i.e. if $\forall i, \alpha_i \geq \beta_i$ and for some $j, \alpha_j > \beta_j$, then A is ranked above B .

This requirement is epistemically justified: to attribute a belief to a group, one should respect the hierarchy induced by the grades given by the individuals; better individual grades should mean something at the group level — be ranked above. Assume agents grade honestly and are not strategic (this hypothesis is relaxed in Section 5).

The problem is that the majority rule, or “go with the numbers”, does not respect this Domination axiom. Finding a counterexample is easy. Three scientists, Alice, Bob, and Carmen, assess two hypotheses, A and B . On the scale {Excellent, Good, Fair, Poor}, hypothesis A receives the grades {Excellent, Fair, Poor}, while hypothesis B receives the grades {Good, Fair, Poor}. Domination requires that the group rank A above B , and can be said to believe A . Yet these grades are compatible with Table 1. If scientists decide to take a majority vote, each scientist votes for their best evaluated hypothesis (last column of Table 1). With two B ’s and one A , the majority rule declares that B is the best hypothesis. Contradiction. The majority rule does not respect Domination (cf. Balinski and Laraki 2020a). The example can be extended to $n > 3$ agents with a scale of at least $n + 1$ grades, so a super-majoritarian voting rule would not solve the problem.

⁷Domination is also called first-order stochastic dominance in the literature. It is called Detailed-Anonymity by Boyer-Kassem (2019), in a different grade modeling framework though.

Counting the votes for or against a proposition, or “going with the numbers”, cannot be a sound epistemic rule.

In the Domination axiom, the grades received by a proposition are considered at the group level, while the reference to which agent gave which grade is irrelevant. On the contrary, the majority rule (or any “going with the numbers” rule) asks that the comparison be made at the individual level first, so that an agent votes for one proposition. This latter way of doing does not make sense from an epistemic viewpoint because what should matter is a proposition’s overall epistemic assessment, in such a way that the same result be given for identical sets of grades. Suppose now there are $2n$ agents, with n grading one proposition and n grading the other. Then, the result of the aggregation rule would better be the same—there would be no sense in arbitrarily pairing these agents two by two to first compare their grades and derive one vote for one proposition. Yet this is what the majority rule asks. It amounts to making comparisons twice — first, at the individual level with grades, second, at the group level with the number of votes. This is a sure way to lose relevant information and be led astray. On the contrary, respecting Domination requires keeping the grades until the group level. A theorem by Balinski and Laraki (2010, p. 182) establishes that for a ranking method to satisfy a few desirable axioms (and avoid Arrow’s and Condorcet’s paradoxes), “*who gave what grade cannot be taken into account: only the sets of grades themselves may be taken into account*” [emphasis original].

4 Another problem — independence of irrelevant alternatives

Whether or not a proposition is finally withdrawn from the set of alternatives considered by the agents should not change the relative ranking of other alternatives. Otherwise, this would give some undue weight to an irrelevant alternative. For instance, which scientific hypothesis is considered the best by scientists cannot depend on whether some additional minor hypothesis is considered. An alternative C may be said to be present or absent: in the former case, C belongs to the set of possible alternatives on which agents can express their opinions; in the latter case, C does not. Consider the following axiom (Balinski and Laraki 2020a, Axiom 8):

(Axiom) Independence of Irrelevant Alternatives (IIA):

$A \succ B$ when C is present if and only if $A \succ B$ when C is absent.

expert	grade			vote	
	Good	Fair	Poor	C present	C absent
#1 to #4	A	C	B	A	A
#5 to #7	B	C	A	B	B
#8 and #9	C	B	A	C	B

Table 2: Grades given by scientists to hypotheses A , B and C , and their votes. When C is present, $A \succ B$, while when C is absent, $B \succ A$.

The majority rule, or any other rule that relies on the number of votes for alternatives, does not respect IIA. To see it, suppose an expert panel with nine experts is faced with three hypothesis between A , B , and C . They would give the grades as in Table 2. With the majority rule, an expert votes for the proposition to which he or she gives the highest grade. When C is present, A receives 4 votes, B receives 3, and C receives 2. If one “goes with the numbers”, A is selected. When C is absent, B receives 5 votes, and A receives 4, which now makes B be selected. So, removing the least ranked hypothesis has changed the winning one, and IIA is violated. This echoes well-known issues in voting theory: in politics as in epistemology, violating IIA is a highly undesirable feature.

5 The solution—Majority Judgment, the epistemic version

Our question “Is the number of agents who believe P_j epistemically relevant?” has received a partial answer so far: it suffers from severe limitations. A more complete answer should show that another aggregation rule is epistemically better. This Section does this by arguing that a better ranking and aggregation rule exists—Majority Judgment, the epistemic version. The axiomatic method is continued.

5.1 Satisfying epistemic axioms

Let us note $A \succeq B$ when the ranking rule places A above B . The following axioms parallel those of Balinski and Laraki (2020a).

(Axiom 1) Grades: The ranking rule takes as input the $n \times k$ grades given by the n agents to the k propositions.

Thus, all information is kept up to the group level (it should not be *a priori* forbidden to do so).

(Axiom 2) Unrestricted Domain: Agents' opinions are unrestricted (i.e. any grade can be given to any proposition).

The ranking rule should be able to give outcomes for any epistemic situation.

(Axiom 3) Anonymity: Permuting the agents (i.e. their indexes in the formal model) does not change the outcome.

The ranking rule should not be unbalanced or discriminating towards some agents just because they have some particular arbitrary indexes. With Axiom 1, it says that who gave which grade should not matter, only the set of grades should. However, there might be good epistemic reasons to give unequal weights to agents, typically if some are more experts than others, and a generalization is considered at the end of Section 5.2.

(Axiom 4) Neutrality: Permuting the propositions does not change the outcome.

The ranking rule should be neutral towards the epistemic content of the propositions.

(Axiom 5) Monotonicity: If $A \succeq B$ and one (or more) agent increases the grade they give to A , then $A \succ B$.

The ranking rule should be sensitive to better epistemic evaluations.

(Axiom 6) Completeness: For any A and B , $A \succeq B$ or $B \succeq A$. If both, then $A \approx B$ (i.e. they have the same rank).

The ranking rule should be able to compare any propositions.

(Axiom 7) Transitivity: If $A \succeq B$ and $B \succeq C$, then $A \succeq C$.

Transitivity might seem obvious, but it is violated by some ranking rules under the name of the Condorcet paradox, leading to inconsistent rankings or no ranking at all.

(Axiom 8) Independence of Irrelevant Alternatives (IIA):
 $A \succ B$ when C is present if and only if $A \succ B$ when C is absent.

One may be tempted to add the Domination axiom, but this is superfluous:

(Theorem 3, Balinski and Laraki 2020a): many⁸ ranking rules satisfy Axioms 1 to 8. They rank propositions only depending on the grades each proposition receives, and respect domination.

⁸The number may be considered as finite or infinite, depending on how a ranking is defined (cf. Balinski and Laraki (2020a), p. 441 and 445).

Thus, there is a plethora of suitable ranking rules. This positive result should be contrasted with Arrow’s (1951) famous theorem which, replacing Axiom 1 with the demand that inputs are agents’ *rankings* of propositions, states that there is no ranking rule (cf. Balinski and Laraki 2020a, Sec. 5). In other words, asking agents to give grades instead of rankings opens huge possibilities.

Examples of suitable functions of the above Theorem are point-summing methods: a numerical value is associated with each grade (e.g. 3 for Good, 2 for Fair, and 1 for Bad), and the propositions are ranked according to the sum or to some average of the received values. These methods display two major drawbacks: first, the associated numerical values have no meaning beyond stipulation—recall that the scale is an ordinal, not an interval one. Consequently, the sum or the average of these values have no meaning either, and adopting different numerical correspondences leads to inconsistent rankings (which one is the right one?). Second, point-summing methods are highly manipulable, that is, agents have incentives not to express their true evaluations of propositions because the group ranking will be closer to their own private rankings if they distort or exaggerate some grades of some propositions. Manipulability is to be avoided in many epistemic contexts: scientific experts might have conflicts of interest or be corrupted by interested parties (see Urfalino and Costa 2015 on the FDA), or agents may just have ego and aim at imposing their views in general. This suggests adding the following axiom:

(Axiom 9) Strategy-proofness in ranking: If $A \succeq B$, then no agent who grades B above A can raise B ’s group grade, or lower A ’s group grade.

Someone who disagrees with an aspect of a ranking should not be able to invert it by misrepresenting their views; the best strategy should be to express their own private views. Unfortunately, this is too demanding a requirement:

(Theorem 5, Balinski and Laraki 2020a): no ranking rule satisfies Axioms 1 to 9.

That is, no ranking rule (which satisfies reasonable demands) can be fully strategy-proof. Instead, one may require that agents cannot act strategically on a large set of cases, if not all. Say that two propositions are polarized when the higher agents evaluate one proposition, the lower they evaluate the other.⁹ If Axiom 9 is required only on the domain of polarized pairs of candidates instead of the full domain of any kind of candidates, then the impossibility result disappears:

⁹Formally, A and B are polarized when, for any two agents, if agent #1 gives A a higher (resp. lower) grade than agent #2, then agent #1 gives B a lower (resp. higher) or equal grade than agent #2.

(Theorem 6, Balinski and Laraki 2020a): There is only one ranking rule which satisfies Axioms 1 to 8, and Axiom 9 on the limited domain of polarized pairs of candidates — call it Majority Judgment.

Thus, Majority Judgment is characterized as the unique function that satisfies a set of desirable epistemic axioms, including a demanding strategy-proof axiom. Note that Majority Judgment is of course also defined for unpolarized candidates because it is a universal function (Axiom 2); the polarization condition is only used to select it among other possible functions. Other axiomatic characterizations of Majority Judgment exist in terms of consensus, crankiness, or error minimization (Balinski and Laraki 2007, 2010, 2020a). Laraki and Varloot (2022) consider some refinements of Majority Judgment (unequal weights to agents, or probability distributions on grades).

5.2 Majority Judgment

Majority Judgment simultaneously provides an aggregation function (a group’s view can be attributed for every proposition), a ranking rule (propositions can be ranked from first to last), and a grading aggregation rule (a group’s grade can be attributed to each proposition). The grade the group attributes to a proposition is called the *majority-grade*. It is the middle-most grade, or median grade, of the n grades it has received: rearranging the grades received by a proposition by decreasing order, the majority-grade is the middle grade.¹⁰ Consider the scale Excellent (E), Very Good (VG), Good (G), Fair (F), and Poor (P). If A receives the grades {E, E, VG, F, P}, its majority-grade is VG. The qualifier “majority” comes from the fact that a majority of agents think that the proposition deserves *at least* this grade, and another majority of agents think that it deserves *at most* this grade—there is a majority agreement on the grade.

Propositions are ranked according to their majority-grade. For instance, A with majority-grade VG is ranked above B which receives the grades {E, G, G, G, F} and has majority-grade G. What if two propositions have the same majority-grades? There are several equivalent ways of expressing the algorithm. One is the following. The middle-most grade is the same, so start from this middlemost grade and consider one additional grade on the left and one on the right, forming an interval of three middlemost grades. There are three possible cases: (i) within this interval, A ’s grades dominate B ’s (or vice versa); in this case, A is ranked above B (or vice versa); (ii) within this interval, A ’s right

¹⁰If n is odd, the $(n + 1)/2$ th grade; if n is even, the $(n + 2)/2$ th grade. That is, when n is even and the two central grades are different, the majority-grade is the lowest of these two central grades (cf. Balinski and Laraki (2011) p. 4-5).

grade dominates B 's right grade, and B 's left grade dominates A 's left grade (or vice versa for A and B); here, A 's grades are less extreme, or more consensual, than B 's, leading A to be ranked above B (or vice versa); (iii) the grades within the interval are equal; in that case, extend the middlemost interval once more by considering one additional grade on each side, and determine again which of the three cases applies. Suppose we want to compare B with C which has grades $\{\text{VG}, \text{G}, \text{G}, \text{G}, \text{G}\}$. Both have majority-grade G , so we consider the interval of the three middlemost grades, here $\{\text{G}, \text{G}, \text{G}\}$ for both B and C . They are equal (case (ii)), so we consider the five middlemost grades, here $\{\text{E}, \text{G}, \text{G}, \text{G}, \text{F}\}$ for B and C , and $\{\text{VG}, \text{G}, \text{G}, \text{G}, \text{G}\}$ for C . On the left, grade E dominates grade VG while on the right, grade G dominates grade F , so we are in case (ii), C 's grades are more consensual and C is ranked above B .

Another formulation of the algorithm is to consider for each proposition the grades different from the majority-grade, either above or below. The larger number of grades among these four quantities (and not the values of the grades) makes the balance tip. In case two propositions have the same larger number of grades different from the majority-grade, one looks at the value of these most numerous grades. For instance, compare again B and C . The grades different from the majority-grade G are in bold:

- B has grades $\{\mathbf{E}, \text{G}, \text{G}, \text{G}, \mathbf{F}\}$
- D has grades $\{\mathbf{VG}, \mathbf{VG}, \text{G}, \text{G}, \mathbf{P}\}$.

B has 1 grade above its majority-grade, 1 grade below. D has 2 grades above G , 1 below. The largest of these four quantities is 2, for the 2 grades above G of D . So D is above B .

6 Objections and answers

Before developing the consequences of an epistemic version of Majority Judgment (Section 7), we consider in this Section two objections that could be raised, about the grades agents give, and about the Condorcet Jury Theorem.¹¹

6.1 Are there not some constraints on the grades?

Contrary to Section 2, suppose for a moment that all agents had probabilistic views P_j^i on the propositions. Since $\sum_{j=1}^k P_j^i = 1$, if the probability of a proposition increases, the probability of another proposition decreases. One can consider that the grades agents

¹¹Objections that are about the algorithm of Majority Judgment itself, and not its epistemic aspects, have been discussed in Balinski and Laraki 2010, 2020a, 2020b, in particular.

give are rough or fuzzy estimates of possible (or “hidden”) probabilities. For instance, the Intergovernmental Panel on Climate Change uses an explicit correspondence table to qualify the claims of its reports, e.g. Very Likely used for a likelihood between 90 and 100% (Stocker et al 2013, 142). Thus, the objection goes, the constraint $\sum_{j=1}^k P_j^i = 1$ leads to a corresponding constraint on the grades: the higher a grade an agent gives to a proposition, the lower they should give to some other proposition. This just means that the propositions are polarized (cf. Section 3). Now, Theorem 4 in Balinski and Laraki (2020a) precisely shows that the majority rule coincides with Majority Judgment for polarized alternatives. So Majority Judgment brings actually nothing new, except a more complex voting apparatus.

My reply first focuses on polarization. The mentioned theorem actually has more demanding hypotheses: it is that *all* propositions are polarized *by pairs*, i.e., for any two propositions, the higher agents evaluate one proposition, the lower they evaluate the other. This is actually not implied by the constraint $\sum_{j=1}^k P_j^i = 1$. For instance, with three propositions, Alice may give the probabilities 60% to A , 30% to B and 10% to C , and Bob 40% to A , 40% to B and 20% to C . Assume that these probabilities (10%, 20%, 40%, 60%) are associated with different grades, i.e. that the language of grades is sufficiently rich. Then A and B are polarized, but B and C are not. So the above-mentioned theorem cannot get off the ground. In addition, one may question the fact that the scale of grades is in correspondence with some probability scale (cf. Section 2). Overall, the constraint-implies-polarization objection does not apply in general, and its validity may only be claimed in the case of a binary, exhaustive choice: A or not- A .

Let us consider this special case in more detail, assuming the scale {Very Good, Good, Fair, Poor, Really Bad}. In our model, agents are not supposed to have precise probabilistic beliefs, so their assessments typically have uncertainties. Alice may be unsure whether A is true and consider that both A and not- A are still Good hypotheses at this point of her inquiry, in the sense that both should receive substantial future research time and efforts.¹² Suppose Bob has less uncertainty, and is able to assess A better than not- A , but only slightly so: he gives them the grades just above and just below some threshold, namely Fair to A and Poor to not- A . In this example, A and not- A are not polarized, since the grades given by Alice are both higher. In addition, agents may find the question ill-posed and prefer to express themselves on the set $\{A, \text{not-}A, \text{“Cannot conclude at this point”}\}$, giving a high grade to the third proposition (and with three propositions, polarization is not granted). Hence, the objection fails for most interesting

¹²Research efforts are generally not symmetric: trying to establish A does not amount to trying to refute not- A (think of classical statistics).

cases. It holds for some, but then the problem is that agents may not be in a position to know that it does and that majority rule could be used.

6.2 What about the Condorcet Jury Theorem?

Another objection is the Condorcet Jury Theorem (CJT): with two alternatives and a majority vote, if agents select the right alternative better than random and with independent probabilities, then larger groups have a larger probability of finding the truth and infinite groups are infallible (cf. Dietrich and Spiekermann 2022). This looks like a strong epistemic argument in favor of the majority rule. What is more, the domination objection from Section 3 could be escaped by the CJT: even if *some* instances of the use of the majority rule do not respect domination, the CJT guarantees convergence to the truth, so domination is respected statistically speaking. And with two propositions, the problem of independence of irrelevant alternatives from Section 4 does not apply anymore. So, the objection goes, the two criticisms which motivated Majority Judgment vanish, while the CJT is an argument in favor of the majority rule.

My answer is threefold. Firstly, the probabilistic independence between agents assumed by the CJT is rarely verified in practical cases, because a causal dependence often exists between agents: in science, researchers share theories or hypotheses and rely on common articles or data; in committees, experts exchange arguments before they make a collective decision; in countries, people rely on media or social networks. Hence, agents generally have correlated views, their probabilities of finding the truth are not probabilistically independent, and CJT cannot apply. One may object that dependencies can be technically conditionalised away, and a more realistic Jury Theorem be established. But it also requires relaxing the competence condition so as not to arrive at the unrealistic conclusion of the asymptotic infallibility of groups (Dietrich and Spiekermann 2013, 2022 Sec. 2).

Secondly, Majority Judgment too can claim to enable groups to find the truth. Moreau (2021) establishes an analog of the CJT for Majority Judgment, called the grading-jury theorem: when n agents assign grades on some ordinal scale, under some competence condition for agents,¹³ the probability that the median grade of the group is the right one increases with n , and converges to 1 when n goes to infinity. This remains true in

¹³Formally: $|o - q| < p < 1$, where p is the individual probability that an agent assigns the right grade to the proposition, o and q are respectively the individual probability of assigning either the right grade or a lower (respectively higher) one. All these individual probabilities are assumed to be probabilistically independent.

situations where agents are individually quite incompetent.¹⁴ In other words, median grading is forgiving. Since Morreau’s median grading amounts to Majority Judgment when there is no tie, the grading-jury theorem establishes the truth-tracking property of Majority Judgment in the absence of a tie, i.e. on a very large and relevant part of the situations. Pivato (2017) offers another generalization of the CJT, which covers both the median rule and the majority rule, this time with partially correlated voters and two or more propositions. So, the majority rule is not the only rule that converges to the truth and its superior nature cannot simply be made by appealing to the CJT.

Thirdly, the requirement that each agent i is better than random at finding the truth ($p_i > 0.5$) actually assumes another illegitimate probabilistic dependence, which can be seen in the language of grades. Call G_A^i and G_B^i the grades that agent i assigns to the propositions A and B , respectively. Say that A is the true one. If the language of grades is rich enough, agents will select the right hypothesis by attributing a higher grade to A than to B : $p_i > 0.5$ is equivalent to $p(G_A^i \succ G_B^i) > 0.5$. This latter inequality puts a condition on how the grades assigned by an individual should be related; this is in stark contradiction with the Unrestricted Domain axiom (Section 5.1) that the grades can be of any value, and with the Domination axiom (Section 3) that the comparison of grades be made at the group level, without constraint on which grades are given by which individuals. In other words, the CJT considers a limited subset of cases than Majority Judgment.

Overall, the Condorcet Jury Theorem rests on questionable premises, and when it applies, it cannot be used to argue for the majority rule over Majority Judgment.

7 Consequences

Let us now draw the consequences of applying Majority Judgment to epistemic contexts—first general ones, then real-life applications, and finally the philosophical literature and academia. The main general consequence is that Majority Judgment should prevail in the epistemic domain as soon as agents can assign qualitative grades to epistemic alternatives and cannot all assign precise probabilities or numerical assessments. All agents should give grades for all propositions, the majority-grade should be identified for each proposition, and the propositions should be ranked accordingly. Counting people, hands, articles, or ballots in favor of the true answer should become an abandoned practice. To report a collective belief, to attribute a belief to a group, to revise one’s be-

¹⁴They are less competent than random to pick the right grade, but balanced in their errors.

liefs, or to defer to others' beliefs, it is not epistemically relevant that "most people think that X ". What matters is which proposition receives the highest majority-grade. The right question to ask is: "How do you evaluate each proposition?" or "According to you, which grade best applies to each hypothesis?", not "According to you, what is the right or true hypothesis?", or "How do you rank these propositions?". All propositions should be assessed in an absolute, not comparative, way through a common scale, whether the aim is to rank all propositions or identify the first (true) one.

These morals give hints about the *deliberation* process that often precedes decisions. Its role should be to help agents attribute grades to propositions, which suggests that propositions should be discussed in turn and assessed in an absolute, not comparative, way. The grade each proposition deserves may even be explicitly debated in the group.

Two remarks are in order on the scope of application of the present conclusions. First, they rely on the hypothesis that agents can assign qualitative grades, but not precise probabilities. Arguably, this is often the case. On one side, it is rare to have only categorical Yes/No beliefs. On the other side, some agent often cannot meaningfully assign a pointed, precise probability to each proposition. The Bayesian literature rests on a theoretical idealization which is known not to be always met in practice. What is more, not all epistemic assessments bear on the probability or likeliness of propositions. For the simplicity, the fruitfulness or the overall quality of a hypothesis, it is meaningless to hope for numerical assessments, and a qualitative framework is required. So, the present framework offers an interesting model in-between, which applies to a wide variety of cases, from scientific expert knowledge to jury members and laypersons. Second, questions which involve preferences, like "Who should be elected for President?", have not been considered here (but Balinski and Laraki 2007, 2010, 2020a, 2020b have already argued for Majority Judgment in these cases). Other cases may be categorized as mixed, involving both preferences and matters of facts, like "Does this research project deserve a grant?" or "Which sanitary measures should be recommended?". Indeed, they require to answer both preference-related questions, like "Should the agency prioritize this topic?" or "Is the closure of schools preferable to a $x\%$ of economic growth loss?", and factual questions, like "Does the research project offer novel hypotheses or ideas?" or "Will this measure lead to $x\%$ of economic growth loss?". Because Majority Judgment has been argued to apply both in preference and epistemic contexts, a natural conclusion is that it should also apply to mixed cases like these.

As a consequence, research grant agencies should stop asking their referees with numerical marks to make averages, since the given marks are generally not precise and function as ordinal grades, and use grades and Majority Judgment. Expert committees

should not issue advice with a majority rule vote. The inner rules of the FDA, the EMA, the EEA, the EFSA, the ECHA, the IASB, and many others should switch to Majority Judgment. Experts should take the habit, for instance in their public outreach, to express themselves not by advocating the hypothesis or recommendation they think is true or best, but by expressing their views on several alternatives.

Let us now turn to more specific consequences for the academic literature. The arguments developed here directly apply to the line of epistemology papers that defend the “go with the numbers” view or the majority rule about expertise or testimony (e.g., Goldman 2001, Coady 2006, Elga 2010, Beatty and Moore 2010, Kelly 2011, Lackey 2013, Miller 2013). For instance, Goldman (2001, p. 103) claims that the number of experts who believe P is epistemically relevant when experts are independent and reliable. But even if one grants the epistemic importance of independence and reliability, expressing the condition in terms of the *number of experts* who believe such and such is missing the point twice: what matters are the grades given by the experts to propositions (not just whether they categorically believe P), and then which proposition has the highest majority-grade at the aggregate level (not the number of experts which believe it). Goldman, like many others after him, is not considering the right quantity and/because he is not asking the right question. The same remark can be made to Lackey (2013, p. 244): she argues for replacing Goldman’s independence and reliability conditions with that of justified confidence, but gives the same undue importance to counting the number of experts who support such or such view. Against Coady (2006, p. 76), “going by the numbers” is not the only rational procedure often available to novices: they often have access to the degree of support of the various propositions. Against Kelly, the symmetrical situation where “every peer on one side of the issue has one and only one counterpart on the other side” (Kelly 2011, p. 203) is not what should drive the peers to suspend judgment; instead, they should look for a symmetry in the grades received. Similarly, higher-order evidence (in Kelly’s picture) should not be assessed in terms of the numbers of peers, but in terms of a set of grades, or a majority-grade. The general approach taken by the field of philosophy of testimony should be amended.

The literature about another epistemic concept, consensus, should be revised too. Consensus is generally regarded as a marker of epistemic authority (Beatty and Moore 2010, Miller 2013), but this concept can be misleading in that it invites considering the number of agents who hold such or such view. Indeed, “consensus” often means a very high percentage of agents who hold the same beliefs (and it is vulnerable to the domination paradox, cf. Section 3). Instead, what matters is the majority-grade each view receives. Miller (2013), who questions on what conditions a consensus is a mark of knowledge,

starts with the mistaken hypothesis that the number of people who believe P matters. It would be more relevant to build a concept of consensus in terms of difference in majority-grades, or the homogeneity of the grades received. Beatty and Moore (2010) point out that a lack of unanimity is not in itself a reason not to trust a group, but still defend the idea of counting scientists. Instead, grades are what should be counted and aggregated.

Similarly, Manin offers the following defense of decisions with the majority rule: “The approval of the greatest number reflects, in that context, the greater strength of one set of arguments compared to others.” (1987, p. 359, quoted by Beatty and Moore 2010, p. 203). His argument takes the agent as a unit: arguments convince agents, and better arguments convince more agents. The present work shows that finer considerations are required: not just for which side one agent is tipping, but to what extent she assesses the opposing arguments. Then, the majority-grades are the right way to compare which set of arguments has “the greater strength”. Although perhaps primarily counter-intuitive, the unit in epistemology as in politics should not be the agent, and judgments should be detached from the agents.

Academic discussions rightly position themselves according to what is perceived as the common view or the ‘received wisdom’ (Bourget and Chalmers 2014, p. 466). If papers only mention the preferred view of their authors, then counting papers is probably the only reasonable way to proceed. My suggestion is for the future: authors could not only mention which view they finally support (or consider as more probable or closer to the truth), but rather to which degree the discussed view, or alternative ones, is supported. Let me start myself: in an epistemic context where qualitative judgments are available, Majority Judgment is a very good rule, while the majority rule is a fair rule (good in a binary question context).

Finally, let me be clear on a few non-consequences. First, when grades cannot be assigned to propositions, then Majority Judgment cannot start, and “going with numbers” may remain the only solution. For instance, it is the case if one inquires about the collective view of a past scientific community, with texts that only mention the theories scientists endorse, not what they think of other theories. In case an agent has not heard of some proposition, it makes sense for the agent to give it the lowest grade on the scale (Balinski and Laraki 2010). Second, the fact that the model rests on a qualitative scale does not mean that agents or experts should not strive to acquire quantitative knowledge; if they have probabilities for instance, then there are theories on how to aggregate them (cf. Cooke 1991, List 2012, Pigozzi 2021 or Martini and Sprenger 2017). Third, the present work only compares Majority Judgment with the majority rule, so it says nothing

for or against other decision rules like non-opposition or apparent consensus.¹⁵ In this sense, the arguments advanced here are not a blank check for Majority Judgment. Fourth, Majority Judgment would be ill-suited for a particular class of epistemic assessments: grading over time. Suppose Majority Judgment is used to give students a grade for the academic year based on the grades they receive through exams they pass each month. After the middle of the year, a student who has only received “Insufficient” grades knows for sure that it will be her final majority-grade. In another situation, a student can know that whether he will receive only “Very Bad” or only “Good” for the rest of the year will equally lower his majority-grade from “Excellent” to “Very Good”. The origin of the problem here is the dynamic characteristic; so majority judgment is of course still recommended in static educational settings, like Ph.D. jurys.

8 Conclusion

I have argued that, in cases agents can assess the propositions under study with qualitative grades (and cannot all assign sharp numerical values to all propositions), the number of agents who believe some proposition is not epistemically relevant. Whether it is a majority, a small number or a near consensus, one should not take numbers of individuals into account to report, attribute, defer to a collective belief, or revise one’s beliefs. What matters, instead, are the grades all agents assign to all propositions. The aggregation should then be made with the concept of majority-grade, which is the highest grade for which a majority of agents think the proposition deserves at least that grade. One should not ask “Which proposition do you believe?” or “Is there a majority of experts for P ?”, but “Is P the best evaluated option?”. This paper has thus presented the first extension of Majority Judgment from social choice theory to epistemology. The standard qualifications, noted by the literature which defends the numbers view—e.g. dependence, reliability, justified confidence—apply here similarly. Typically, a grade given by an individual who is half dependent on another individual should be counted half. Hopefully, this can be accommodated by Majority Judgment, which can receive a formulation with non-integer votes or percentages, like in the case of large electorates.

Methodologically, representing beliefs within a qualitative scale, beyond a categorical one, has brought new strong results. My intuition is that it is a fruitful framework for social epistemology more generally. For instance, one may extend the literature of judgment aggregation, which considers categorical beliefs, to beliefs within a qualitative

¹⁵Deliberation within the group enables to reach a state where some proposition is not publicly opposed by any group member, which thus displays an apparent consensus (cf. Urfalino 2007, 2014).

scale. Future work could also investigate how the bridge can be made between categorical and qualitative beliefs on one side, and qualitative and probabilistic beliefs on the other side (but this may not be easy, given the difficulty in bridging categorical and probabilistic beliefs, cf. Dietrich and List 2018, Dietrich 2022, Wang 2024). Probability intervals could be considered, and axioms of strategy-proofness should be particularly examined. Another valuable future research path could be to compare the relative merits of Majority Judgment with decisions by non-opposition.

References

- Arrow, Kenneth J. 1951. *Social Choice and Individual Values*. Yale University Press, New Haven, CT. 2nd edition 1963, Wiley, New York, NY.
- Balinski, Michel, and Rida Laraki. 2007. “A theory of measuring, electing and ranking”. *Proc Natl Acad Sci USA* 104(21):8720–8725.
- Balinski, Michel, and Rida Laraki. 2010. *Majority Judgment: Measuring, Ranking, and Electing*. Cambridge, MA: MIT Press.
- Balinski, Michel, and Rida Laraki. 2020a “Majority Judgment vs Majority Rule”, *Social Choice and Welfare*, 54: 429–461.
- Balinski, Michel, and Rida Laraki. 2020b “Majority Judgment vs Approval Voting”, *Operations Research*, 70(3):1296–1316.
- Beatty, John and Alfred Moore. 2010. “Should We Aim For Consensus?”. *Episteme: A Journal of Social Epistemology* 7(3): 198–214.
- Bourget, David and David J. Chalmers. 2014. “What do philosophers believe?”. *Philosophical Studies*, 170:465–500.
- Boyer-Kassem, Thomas. 2019. “Scientific Expertise and Risk Aggregation”. *Philosophy of Science*, 86: 124–144.
- Coady, David. 2006. “When Experts Disagree”. *Episteme: A Journal of Social Epistemology*, 3(1-2): 68–79.
- Condorcet le Marquis de. 1785. *Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix*. L’Imprimerie Royale, Paris.
- Cooke, Roger M. 1991. *Experts in Uncertainty. Opinion and Subjective Probability in Science*. Oxford University Press.
- Dietrich, Franz. 2022. “Categorical Versus Graded Beliefs”. *Frontiers in Psychology* 13:817940.
- Dietrich, Franz, and Christian List. 2016. “Probabilistic Opinion Pooling”. In A. Hájek and C. Hitchcock, eds, *Oxford Handbook of Probability and Philosophy*, Oxford

- University Press.
- Dietrich, Franz, and Christian List. 2018. “From Degrees of Belief to Beliefs: Lessons from Judgment-Aggregation Theory”. *The Journal of Philosophy* 115: 225-270.
- Dietrich, Franz and Kai Spiekermann. 2013. “Epistemic Democracy with Defensible Premises”. *Economics and Philosophy*, 29(1): 87–120.
- Dietrich, Franz and Kai Spiekermann. 2022. “Jury Theorems”. The Stanford Encyclopedia of Philosophy (Summer 2022 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2022/entries/jury-theorems/>>.
- Douglas, Heather E. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.
- Elga, Adam. 2010. “How to Disagree about How to Disagree”. In Richard Feldman and Ted A. Warfield (eds.), *Disagreement*, Oxford: Oxford University Press, pp. 175–86.
- Goldman, Alvin I. 2001. “Experts: Which Ones Should You Trust?”. *Philosophy and Phenomenological Research*, 63(1): 85–110.
- John, Stephen D. 2011. “Expert Testimony and Epistemological Free-Riding: The MMR Controversy”. *The Philosophical Quarterly* (61): 496–517.
- Kelly, Thomas, 2011. “Peer Disagreement and Higher Order Evidence”. In Alvin I. Goldman and Dennis Whitcomb (eds.), *Social Epistemology: Essential Readings*, New York: Oxford University Press, pp. 183–217.
- Kuhn, Thomas S. 1977. “Objectivity, Value Judgment, and Theory Choice”. In T. S. Kuhn, *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago: The University of Chicago Press, pp. 320-339.
- Lackey, Jennifer. 2013. “Disagreement and Belief Dependence: Why Numbers Matter”, in David Christensen and Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press: 243–268.
- Laraki, Rida and Estelle Varloot. 2022. “Level-Strategypooof Belief Aggregation and Application to Majority Judgment under Uncertainty”. *Proceedings of the 23rd ACM Conference on Economics and Computation*. New York, NY, USA: Association for Computing Machinery. p. 335-369. <https://doi.org/10.1145/3490486.3538309>
- List, Christian. 2012. “The theory of judgment aggregation: an introductory review.” *Synthese* 187:179–207.
- Manin, Bernard. 1987. “On Legitimacy and Political Deliberation”. *Political Theory* 15(3): 338–68.
- Martini, Carlo, and Jan Sprenger. 2017. “Opinion aggregation and individual expertise.” In *Scientific collaboration and collective knowledge*, ed. by T. Boyer-Kassem, C. Mayo-Wilson and M. Weisberg, Oxford University Press.

- Miller, Boaz. 2013. “When is consensus knowledge based? Distinguishing shared knowledge from mere agreement”. *Synthese* 190(7): 1293–1316.
- Morreau, Michael. 2016. “Grading in groups”. *Economics and Philosophy* 32: 323–352.
- Morreau, Michael. 2021. “Democracy without Enlightenment: A Jury Theorem for Evaluative Voting”. *The Journal of Political Philosophy* 29(2): 188–210.
- Pigozzi, Gabriella. 2021. “Belief Merging and Judgment Aggregation”. The Stanford Encyclopedia of Philosophy (Spring 2021 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2021/entries/belief-merging/>
- Pivato, Marcus. 2017. “Epistemic democracy with correlated voters”. *Journal of Mathematical Economics*, 72, 51–69.
- Resnik, David B. 2021. *Precautionary Reasoning in Environmental and Public Health Policy*. Cham (Switzerland): Springer.
- Rudner, Richard. 1953. “The Scientist qua Scientist Makes Value Judgments”. *Philosophy of Science* 20: 1–6.
- Stocker, Thomas F. et al. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Urfalino, Philippe. 2007. “La décision par consensus apparent. Nature et propriétés”, *Revue européenne des sciences sociales*, XLV-136. URL : <<http://ress.revues.org/86>>.
- Urfalino, Philippe. 2014. “The Rule of Non-Opposition: Opening Up Decision-Making by Consensus.” *The Journal of Political Philosophy* 22(3): 320–341.
- Urfalino, Philippe, and Pascaline Costa. 2015. “Secret-Public Voting in FDA Advisory Committees.” In *Secrecy and Publicity in Votes and Debates*, ed. Jon Elster, 165–94. Cambridge: Cambridge University Press.
- Wang, Minkyung. 2024. Aggregating individual credences into collective binary beliefs: An impossibility result. *Theory and Decision* (97):39-66.