Episteme, 9, 2 (2012) 115–137 © Cambridge University Press
doi:10.1017/epi.2012.4

A STATISTICAL APPROACH TO EPISTEMIC DEMOCRACY

MARCUS PIVATO
marcuspivato@trentu.ca

ABSTRACT

We briefly review Condorcet's and Young's epistemic interpretations of preference aggregation rules as maximum likelihood estimators. We then develop a general framework for interpreting epistemic social choice rules as maximum likelihood estimators, maximum *a posteriori* estimators, or expected utility maximizers. We illustrate this framework with several examples. Finally, we critique this program.

INTRODUCTION

Consider the following collective decision problems:

(a) A team of doctors must diagnose a patient, based on her symptoms and medical history.
(b) A team of doctors must identify the treatment for a patient that maximizes her chance of full recovery.
(c) A team of doctors must decide, on the basis of ambiguous evidence, whether a newly identified microbe is the cause of a mysterious disease.
(d) A team of economists must forecast the inflation rate over the coming year.
(e) A team of economists (e.g. in the central bank) must determine the optimal interest rate.
(f) A team of economists must decide whether a proposed change in financial regulations will make markets more volatile or less volatile.
(g) A team of meteorologists must predict the average temperature in Pittsburgh on June 25.
(h) A team of climatologists must estimate the probability that the west Antarctic ice sheet will melt by the year 2050.
(i) A team of scientists and economists must decide which CO_2 emissions control policies will maximize the average annual worldwide economic growth rate, while still keeping atmospheric CO_2 concentrations below 450 ppm for the next 50 years.

In all these examples, a group is trying to collectively determine the correct answer to some factual question, which has an objectively correct (but unknown) answer. Suppose all members of the group have the same values or preferences; the only conflict is over their beliefs about objective facts. Also suppose that they wish to arrive at a collective

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

decision through some broadly democratic process.[1] What is the best way for the voters to reconcile their contradictory beliefs and arrive at a collective decision? We can distinguish between three sorts of questions:

1. A question about the best policy or best action to take.
2. A question about a matter of fact, where there is some obvious background probability distribution over the possible answers.
3. A question about a matter of fact, where there is *no* obvious background probability distribution.

For example, problems (b), (e) and (i) above are Type 1 questions: the team wants to find the optimal treatment, interest rate, or policy. Problems (a), (d), and (g) are Type 2 questions. In these examples, there is a fairly rich body of historical data (e.g. about other patients of the same age, gender, and medical history; about the weather in Pittsburgh in late June; about the behaviour of the macroeconomy over the last 40 years), with which we can construct a background probability distribution. New information (e.g. particular symptoms; barometer data and satelite imagery; recent movements of the consumer price index) can be assessed in the context of this background probability distribution. Problems (c), (f), and (h) are Type 3 questions. There is no obvious choice of background probability distribution for the behaviour of mysterious microorganisms, the consequences of financial regulations, or the long-term behavior of the cryosphere.

In Type 1 questions, the word 'best' suggests that the group wishes to maximize the value of some (universally agreed upon) utility function. When confronting uncertainty (where each action yields a probability distribution over the set of outcomes), the standard approach[2] is to choose the action(s) with the largest *expected* utility. In other words, the social decision should be an *expected utility maximizer* (EUM).

In questions of Type 2 or 3, there is no utility function we wish to maximize – either because the question is of purely 'academic' interest, or because different people have different utility functions, and the group as a whole does not wish to fixate on any particular utility function. In a Type 2 question, our background knowledge can be represented by a *prior probability distribution* over the possible answers; society's goal is to combine this prior distribution with the information provided by the voters to identify the answer which is most likely to be correct, all things considered. If there is a well-defined probabilistic relationship between the possible answers and the possible signals sent by voters, then the correct way to do this is to use Bayes rule to compute a *posterior probability distribution*, and then choose the answer(s) with the largest posterior probability. In the jargon of statistics, the social decision should be a *maximum a posteriori*

1 Cohen (1986) and Estlund (1997) describe this kind of 'purely epistemic' social decision problem as 'epistemic democracy'. This model is in direct contrast to much of social choice theory, which assumes that voters agree about the objective facts, but have irreconcilable differences in their preferences or values.

2 By 'standard', I mean that this approach is almost universally accepted in economics, finance, and management science. Furthermore, it is the unique approach which satisfies certain axioms of 'rationality' or 'consistency', as demonstrated by the well-known theorems of von Neumann and Morgenstern, Savage, etc. Of course, there are other ('non-expected utility') approaches to risky decision-making; these may be better descriptions of actual human behaviour, but they are hard to defend on the grounds of 'rationality'.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

(MAP) estimator.[3] In a Type 3 question, there is no natural prior distribution, and all the possible answers are assumed to be equally likely, *a priori*.[4] In this situation, we begin with the uniformly distributed prior distribution, and again apply Bayes rule to identify the most likely outcome; thus, the social decision will be a *maximum likelihood estimator* (MLE).[5]

Condorcet (1785) was the first to suggest that a social decision could be an MLE; the celebrated *Condorcet Jury Theorem* (CJT) says that, under certain hypotheses, the simple majority voting rule is an MLE when society must answer a yes/no question. Condorcet's hypotheses (that the votes are independent and identically distributed random variables) were highly unrealistic; there are now many variations of the CJT with more realistic hypotheses.[6] Meanwhile Young (1986, 1988, 1995, 1997) has shown that the Kemeny rule and the Borda rule can both be interpreted as MLEs, when society faces a preference-aggregation problem involving more than two alternatives. Conitzer, Xia and their collaborators have extended the Condorcet-Young approach, and asked what other preference aggregation rules can be interpreted as MLEs.[7]

However, preference aggregation is only one kind of social choice problem, and the MLE is only one of the three types of epistemic rules described above. What other voting rules can be interpreted as MLE, MAP, or EUM procedures, given suitable assumptions about the nature of the decision problem and the background knowledge of the voters? This paper will first describe the main results of Pivato (2012), which offer a partial answer to this question. Then I will critique this program.

How to Read this Paper

A full and precise understanding of the key ideas in this paper necessarily involves a certain amount of formalism and abstraction. However, the main ideas can be grasped from the concrete examples and informal discussions in the main text. In some places, I have given more mathematically precise exposition in footnotes, which will be of interest to only some readers.

VOTING RULES

Let I represent a set of voters, and let V be the set of possible 'votes' (or 'signals' or 'messages') which could be sent by each voter. A *profile* is a list $\mathbf{v}$ which assigns a signal v_i in V to each voter i in I. Let V^I denote the set of all possible profiles. Let X be a set of

3 A MAP can be regarded as a special kind of EUM, where we obtain a utility of 1 for the correct answer, and a utility of 0 for any incorrect answer.

4 This so-called *Principle of insufficient reason* is only really defensible when no obvious asymmetry exists between different possible answers, so they are effectively interchangeable.

5 Thus, an MLE can sometimes be regarded as a special case of a MAP (with a uniform prior). But this interpretation only makes sense when the uniform distribution is itself a *bona fide* probability measure, which is only the case when the set of possible answers is finite.

6 See e.g. List and Goodin (2001), Nitzan (2010: part III), Hummel (2010), Dietrich and Spiekerman (2011), or Kaniovski (2010).

7 See Conitzer and Sandholm (2005), Conitzer et al. (2009), Xia et al. (2010) and Xia and Conitzer (2011).

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

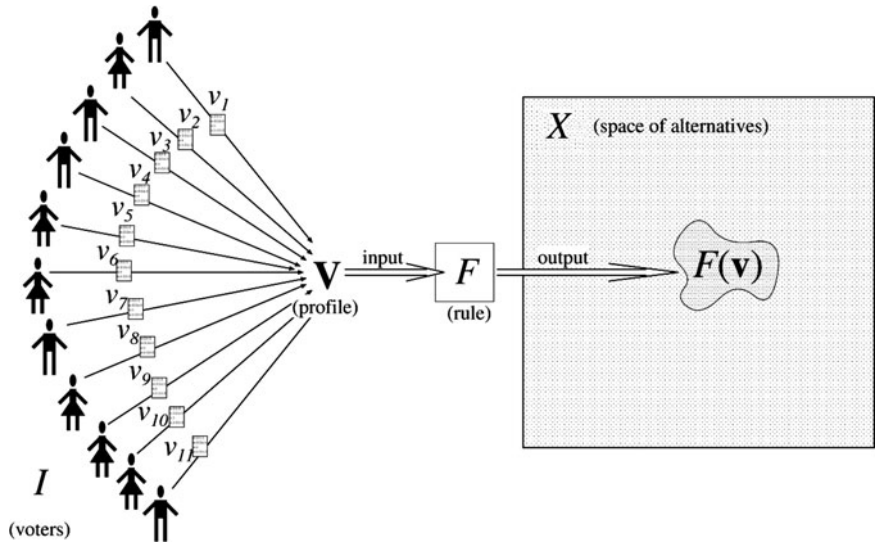


Fig. 1. A voting rule.

alternatives available to society (e.g. possible actions, possible answers to some question, job candidates). A *voting rule* is a correspondence F from V^I to X – that is, for any profile $\mathbf{v}$ in V^I, the rule yields a nonempty subset $F(\mathbf{v})\subseteq X$. Typically (but not always) $F(\mathbf{v})$ will be a singleton. This is illustrated schematically in Figure 1.

For example, suppose Alice, Bob, Chiara, Dave and Emily are trying to decide where to eat dinner. The available restaurants are Indian, Japanese, Korean, Lebanese, and Moroccan. (Thus, in this example, I = {Alice, Bob, Chiara, Dave, Emily}, while X = {Indian, Japanese, Korean, Lebanese, Moroccan}.) The five friends use the following rule: each person names her first choice and her second choice. (Thus, V = {I/J, I/K, I/L, I/M, J/I, J/K, J/L,. . ., M/L}, where, for example, 'I/J' represents a first choice of *Indian* and a second choice of *Japanese*.) A person's *first* choice gets two points, and her *second* choice gets one point – except for Emily. It is Emily's birthday, so her first choice gets three points, and her second choice gets two points. The restaurant which accumulates the most points wins. If the five friends vote as in Table 1, then the Moroccan restaurant wins with five points.

This is a simple example of a *scoring rule*. In a general scoring rule, there is a *score function, S*, which, for any voter i in I and any signal v in V, assigns a real-valued

Table 1. The restaurant problem

Voter	1st/2nd choice	Indian	Japanese	Korean	Lebanese	Moroccan
Alice	I/J	2	1	0	0	0
Bob	K/L	0	0	2	1	0
Chiara	M/I	1	0	0	0	2
Dave	J/L	0	2	0	1	0
Emily	M/K	0	0	2	0	3
Total score		3	3	4	2	5

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

'score' $S(i,v,x)$ to each alternative x in X. (For example, in the restaurant voting rule, $S(i,$ J/K, Japanese$) = 2$ and $S(i,$ J/K, Korean$) = 1$ if i is any voter *except* Emily. Meanwhile $S($Emily, J/K, Japanese$) = 3$ and $S($Emily, J/K, Korean$) = 2$.) If there are N voters, then it will be convenient to represent the voters with the set $I = \{1,2,3,\ldots,N\}$. (In the restaurant example, $N = 5$.) For any profile $\mathbf{v}$ in V^I, let $F_S(\mathbf{v})$ be the set of all element(s) x in X which maximize the sum $S(1,v_1,x) + S(2,v_2,x) + \ldots + S(N,v_N,x)$. The resulting correspondence F_S from V^I into X is the *scoring rule* defined by S. Many common voting rules can be represented as scoring rules. For example:

- In the *plurality rule*, $V = X$, and, for every voter i in I, every alternative x in X, and every signal v in V, we have $S(i,v,x) = 1$ if $v = x$, whereas $S(i,v,x) = 0$ if $v \neq x$. (Thus, each voter names her favourite choice. The alternative named by the most voters wins.)
- In the *weighted plurality rule*, $V = X$, and, for every voter i in I, every alternative x in X, and every signal v in V, we have $S(i,v,x) = w_i$ if $v = x$, whereas $S(i,v,x) = 0$ if $v \neq x$. (Here, w_i is the 'weight' of voter i, which roughly measures her influence over the outcome.)
- In the *antiplurality rule*, $V = X$, and for every voter i in I, every alternative x in X, and every signal v in V, we have $S(i,v,x) = -1$ if $v = x$, whereas $S(i,v,x) = 0$ if $v \neq x$ (i.e. each voter votes *against* her least-prefered alternative, implicitly endorsing all the other alternatives).
- In *approval voting*, V is the set of all *subsets* of X, and, for every voter i in I, every alternative x in X, and any signal v in V (i.e. any subset v of X), we have $S(i,v,x) = 1$ if x is an element of v, whereas $S(i,v,x) = 0$ if x is not an element of v (Brams and Fishburn 1983).
- In *range voting*, V is the set of all functions mapping each element of X into a real number between 0 and 1, and, for every voter i in I, every alternative x in X, and every signal v in V, we have $S(i,v,x) = v(x)$ (Smith 2001).
- In the *Borda rule*, V is the set of all preference orders over X, and, for every voter i in I, every alternative x in X, and every signal v in V, we have $S(i,v,x) = r$ if x is ranked rth from the bottom according to the preference order v.
- In the *Kemeny* (1959) *rule*, both V and X are the set of all preference orders over some set A of alternatives, and, for every voter i in I, every alternative x in X, and every signal v in V, the score $S(i,v,x)$ is the number of pairwise comparisons on which the preference orders v and x agree.

Table 2 illustrates the computation of the plurality, antiplurality, approval, Borda, and range voting rules. In this example, there are three alternatives, x, y and z, and five voters, Alice, Bob, Chiara, Dave and Emily. Each alternative has a utility for each voter which is between 0 and 1. We assume these utilities determine the outcome of the range voting rule, and we also assume that each voter 'approves' of an alternative if and only if it has a utility for her of at least 0.5. As shown in Table 2, different rules can produce different winners from the same profile.

Next, Table 3 illustrates the Kemeny rule. We have three alternatives (say, $A = \{a,b,c\}$), so the space X contains six preference orders (i.e. $X = \{a > b > c,\ b > c > a,\ c > a > b,\ c > b > a,\ b > a > c,\ a > c > b\}$). This time, we suppose there are seven voters (say, $I = \{$Alice, Bob, Chiara, Dave, Emily, Fred, Gwenn$\}$). Each voter has a preference order over the alternatives. Each preference order on $\{a,b,c\}$ is entirely described by three binary comparisons: a vs. b, b vs. c, and c vs. a. (For example, the preference order $a > b > c$ corresponds to the

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

Table 2. Five scoring rules

	Utility (and range voting score)			Plurality score			Antiplurality score			Approval score			Borda score		
Voter	x	y	z	x	y	z	x	y	z	x	y	z	x	y	z
Alice	1.0	0.6	0	1	0	0	1	1	0	1	1	0	2	1	0
Bob	0	1.0	0.3	0	1	0	0	1	1	0	1	0	0	2	1
Chiara	0.7	0	1.0	0	0	1	1	0	1	1	0	1	1	0	2
Dave	0.4	0	1.0	0	0	1	1	0	1	0	0	1	1	0	2
Emily	0.6	1.0	0	0	1	0	1	1	0	1	1	0	1	2	0
Total	2.7	2.6	2.3	1	2	2	4	3	3	3	3	2	5	5	5
Winner	x			y or z			x			x or y			x, y, or z		

Table 3. The Kemeny rule

Voter	**Actual preferences**	**Score of $a>b>c$**	**Score of $b>c>a$**	**Score of $c>a>b$**	**Score of $c>b>a$**	**Score of $b>a>c$**	**Score of $a>c>b$**
Alice	$a>b>c$	3	1	1	0	2	2
Bob	$b>c>a$	1	3	1	2	2	0
Chiara	$c>a>b$	1	1	3	2	0	2
Dave	$c>a>b$	1	1	3	2	0	2
Emily	$b>a>c$	2	2	0	1	3	1
Fred	$b>a>c$	2	2	0	1	3	1
Gwenn	$c>b>a$	0	2	2	3	1	1
Total score		10	**12**	10	11	11	9

assertions '$a>b$', '$b>c$', and '$a>c$'.) Thus, for any voter i in I, and any two preference orders x and v, the Kemeny score $S(i,v,x)$ is 3 if $v=x$, and it is 0 if v is the exact reverse of x; otherwise it is either 1 or 2. The winning preference order[8] is $b>c>a$, because it gets the highest total Kemeny score (namely, 12).

Aside from the voting rule in the restaurant example and the weighted plurality rule, all of the scoring rules described above are *anonymous*, which means that every voter has exactly the same influence over the outcome as every other voter.[9] Also, the above rules are all *neutral*, which means that these rules do not favor any alternative in X over any

8 Note: the output of the Kemeny rule is a *preference order* over the set A, not an element of A itself. Some might argue that we are really only interested in the top element of the winning preference order (in this case, b). However, as argued by Arrow (1963), Young (1986, 1988, 1995), and others, there are situations where it makes more sense to regard the social decision as the entire preference order $b>c>a$, and not just as b. E.g. if we want to regard the group as a rational agent, then (according to the standard economic concept of 'rationality') we must impute a *preference order* to the group, which will 'rationalize' its choices. This is especially important if the group belatedly discovers that b is not feasible; then it needs a contingency plan – in this case, c. (And if c is also not feasible, then we need a second contingency plan – in this case, a – and so on.)

9 Technically, a scoring rule is *anonymous* if $S(i,v,x)=S(j,v,x)$ for all voters i and j in I, every signal v in V, and every alternative x in X.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

Table 4. Predicting tomorrow's weather

$p(v=y\|w_t=x)$: the conditional probability that a voter sends signal y if the true weather is x.		x			
		S	*C*	*F*	*R*
y	*S*	0.6	0.2	0.1	0.05
	C	0.2	0.5	0.2	0.05
	F	0.1	0.2	0.5	0.2
	R	0.1	0.1	0.2	0.7

other alternative.[10] If $S(i,v,x) = -\infty$ for some voter i in I, some signal v in V, and some alternative x in X, then voter i can effectively 'veto' the choice x by sending the signal v. A rule has *no vetos* if this is never the case. All the above rules have no vetos.

STATISTICAL ESTIMATORS

I will begin with an example. Suppose Alice, Bob and Chiara are three independent, equally competent meteorologists, who are trying to predict tomorrow's weather. For simplicity, suppose the weather can be in one of four states: **S**unny, **C**loudy, **F**oggy, or **R**aining. Assume that all three meteorologists have the predictive accuracy described in Table 4. For example, if the true state of tomorrow's weather is cloudy (C), then the C column of Table 4 says there is a 50% chance that a given meteorologist will correctly predict this fact, a 20% chance that she will incorrectly think it will be sunny (S), a 20% chance that she will incorrectly predict fog (F), and a 10% chance that she will incorrectly predict rain (R).

Another way to read Table 4 is in terms of *likelihoods*. If Alice predicts that it will be sunny, then the S row of Table 4 says that there is a 0.6 likelihood that this prediction is correct, a 0.2 likelihood that the weather will actually be cloudy, a 0.1 likelihood that it will be foggy, and a 0.05 likelihood of rain. (Likelihoods are *not* probabilities; observe that $0.6 + 0.2 + 0.1 + 0.05 \neq 1$.)

A *profile* is a list (v_a,v_b,v_c), summarizing the forecasts of the three meteorologists; we will denote it with the symbol $\mathbf{v}$, as before. For example, suppose that Alice predicts that it will be sunny, Bob predicts that it will be cloudy, and Chiara predicts that it will be raining. Then, we have $v_a = S$, $v_b = C$, and $v_c = R$, yielding the profile $\mathbf{v} = (S,C,R)$.

Suppose tomorrow's weather will actually be sunny. Then, according to Table 4, the conditional probability that Alice predicts sunshine is given by $p(v_a = S|w_t = S) = 0.6$, while the conditional probability of Bob's prediction is $p(v_b = C|w_t = S) = 0.2$, and the conditional probability of Chiara's prediction is $p(v_c = R|w_t = S) = 0.1$. If Alice, Bob and Chiara make *independent* random errors, then the conditional probability of seeing the profile (S,C,R), given that tomorrow's weather is actually sunny, is the *product* of the

10 Technically, a scoring rule is *neutral* if, for every voter i in I, and every pair of alternatives x and y in X, and every signal v in V, there is another signal w in V such that $S(i,v,x) = S(i,w,y)$. Thus, for any level of support which i can express for x with one signal, there is another signal allowing i to express the same level of support for y.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

conditional probabilities of their three individual predictions. In symbols: $\text{Prob}[\mathbf{v} = (S,C,R)|w_t = S] = p(v_a = S|w_t = S) \times p(v_b = C|w_t = S) \times p(v_c = R|w_t = S)$. More generally, the conditional probability of the profile (S,C,R) under the four possible weather states is as follows:[11]

(B1)

$$\begin{aligned}
&\text{Prob}[\mathbf{v} = (S, C, R)|w_t = S] = p(v_a = S|w_t = S) \times p(v_b = C|w_t = S) \times p(v_c = R|w_t = S)\\
&\quad = 0.6 \times 0.2 \times 0.1 = 0.012,\\
&\text{Prob}[\mathbf{v} = (S, C, R)|w_t = C] = p(v_a = S|w_t = C) \times p(v_b = C|w_t = C) \times p(v_c = R|w_t = C)\\
&\quad = 0.2 \times 0.5 \times 0.1 = 0.01,\\
&\text{Prob}[\mathbf{v} = (S, C, R)|w_t = F] = p(v_a = S|w_t = F) \times p(v_b = C|w_t = F) \times p(v_c = R|w_t = F)\\
&\quad = 0.1 \times 0.2 \times 0.2 = 0.004, \text{ and}\\
&\text{Prob}[\mathbf{v} = (S, C, R)|w_t = R] = p(v_a = S|w_t = R) \times p(v_b = C|w_t = R) \times p(v_c = R|w_t = R)\\
&\quad = 0.05 \times 0.05 \times 0.7 = 0.00175.
\end{aligned}$$

We refer to these values as the *group likelihoods* of the four states, given the profile (S,C,R).[12] The *maximum likelihood estimator*, given the profile (S,C,R), is the state with the highest group likelihood. In this case, it is state S, which has a group likelihood of 0.012. Thus, if our *only* information about the weather was the forecasts of Alice, Bob and Chiara, then we would judge it *more likely* that tomorrow's weather will be sunny than that it will be any of the other three states. Thus, tomorrow might be a good day for a trip to the beach.

However, suppose that, based on historical data, we *also* know that the prior probability distribution of weather at this time of year is given by: $\text{Prob}(S) = 0.2$, $\text{Prob}(C) = 0.4$, $\text{Prob}(F) = 0.2$, and $\text{Prob}(R) = 0.2$. In that case, we compute the *posterior probabilities* of the four states as follows. First, define

(B2)

$$\begin{aligned}
&R[S|S, C, R] := \text{Prob}(S) \times \text{Prob}[\mathbf{v} = (S, C, R)|w_t{=}S] = 0.2 \times 0.012 = 0.0024,\\
&R[C|S, C, R] := \text{Prob}(C) \times \text{Prob}[\mathbf{v} = (S, C, R)|w_t{=}C] = 0.4 \times 0.01 = 0.004,\\
&R[F|S, C, R] := \text{Prob}(F) \times \text{Prob}[\mathbf{v} = (S, C, R)|w_t{=}F] = 0.2 \times 0.004 = 0.0008, \text{ and}\\
&R[R|S, C, R] := \text{Prob}(R) \times \text{Prob}[\mathbf{v} = (S, C, R)|w_t{=}R] = 0.2 \times 0.00175 = 0.00035,
\end{aligned}$$

Now let $\underline{R} := R[S \mid S,C,R] + R[C \mid S,C,R] + R[F \mid S,C,R] + R[R \mid S,C,R] = 0.00755$. Then *Bayes rule* says that the *posterior probabilities* of the four states are as follows:[13]

11 It is not necessary to follow the computations in (B1), (B2) and (B3) in detail to understand the main idea.

12 Note that these likelihoods do *not* form a probability distribution on X, because $0.012 + 0.01 + 0.004 + 0.00175 \neq 1$.

13 We divide by $\underline{R}$ to ensure that the four values sum up to 1, thereby producing a probability distribution.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

(B3)

$$\text{Prob}[w_t = S|\mathbf{v} = (S, C, R)] = R[S|S, C, R]/\underline{R} = 0.0024/0.00755 \times 0.318,$$
$$\text{Prob}[w_t = C|\mathbf{v} = (S, C, R)] = R[C|S, C, R]/\underline{R} = 0.004/0.00755 \times 0.530,$$
$$\text{Prob}[w_t = F|\mathbf{v} = (S, C, R)] = R[F|S, C, R]/\underline{R} = 0.0008/0.00755 \times 0.106, \text{ and}$$
$$\text{Prob}[w_t = R|\mathbf{v} = (S, C, R)] = R[R|S, C, R]/\underline{R} = 0.00035/0.00755 \times 0.0463.$$

The maximum *a posteriori* estimator, given this scenario and profile, is the state with the largest posterior probability. In this case, it is *C*, which attains a posterior probability of approximately 0.530. Perhaps tomorrow is not a good day for the beach, after all.

The example of the three meteorologists illustrates a much more general method. Suppose that X represents a set of possible 'states of nature'. The true state is unknown. As before, let V be a set of possible 'signals' or 'messages' which could be sent by each voter. Suppose that each voter receives some partial information about the true state, which determines the way she votes. We can mathematically model these assumptions by specifying a prior probability distribution over X, and, for each voter i in I and every state x in X, specifying a probability distribution over V (called the *error model*), which describes the sort of signal which voter i is likely to send if the true state of nature is x. (In the example of the three meteorologists, we had $V = X = \{S,C,F, R\}$, and the error model was given by Table 4.) I will call a combination of a prior and an error model a *scenario*. I will say this scenario is *anonymous* if all voters are equally competent, and receive the same quantity and quality of information (i.e. for any state x in X, we have the same probability distribution over signals in V for every voter i in I). If the conditional probability of voter i sending signal v is *zero*, given state of nature x, then the scenario says certain events are 'impossible'. If this never occurs, for any voter i in I, any signal v in V, and any state x in X, then I will say the scenario has *no impossibilities*. (For example, the error model in Table 4 is anonymous, and has no impossibilities.)

Assume that the signals sent by different voters are conditionally independent random variables, for any state of nature. (Informally, this means that each voter forms her opinion in isolation, without communicating with any of the other voters; thus, the errors she makes are not correlated with the errors of other voters.) Given any scenario *Sc*, and any profile $\mathbf{v}$ in V^I, we can use Bayes rule to compute the posterior probability distribution over X, conditional on $\mathbf{v}$. (In the meteorology example, this computation was performed in equations (B1), (B2), and (B3).) Let MAP(*Sc*,$\mathbf{v}$) denote the element(s) of X which obtain maximal probability in this posterior distribution (the *maximum a posteriori estimator* determined by scenario *Sc* and profile $\mathbf{v}$). This is the state which we deem to be *most probable*, given the information in the profile $\mathbf{v}$, together with the background knowledge encoded in the scenario *Sc*. Thus, if we had to make a bet about the true state of nature, then our 'best bet' would be MAP(*Sc*,$\mathbf{v}$). (In the meteorology example in Table 4, we had MAP(*Sc*,$\mathbf{v}$) = {*C*}.) The computation of the maximum *a posteriori* estimator is schematically illustrated in Figure 2 above.

In particular, if we assume that the prior probability is uniformly distributed over X (i.e. all states are equally probable, *a priori*), then we obtain the *maximum likelihood estimator* determined by scenario *Sc* and profile $\mathbf{v}$; this is denoted by MLE(*Sc*,$\mathbf{v}$). This is the state which we deem to be *most likely*, given the information in the profile $\mathbf{v}$, together

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

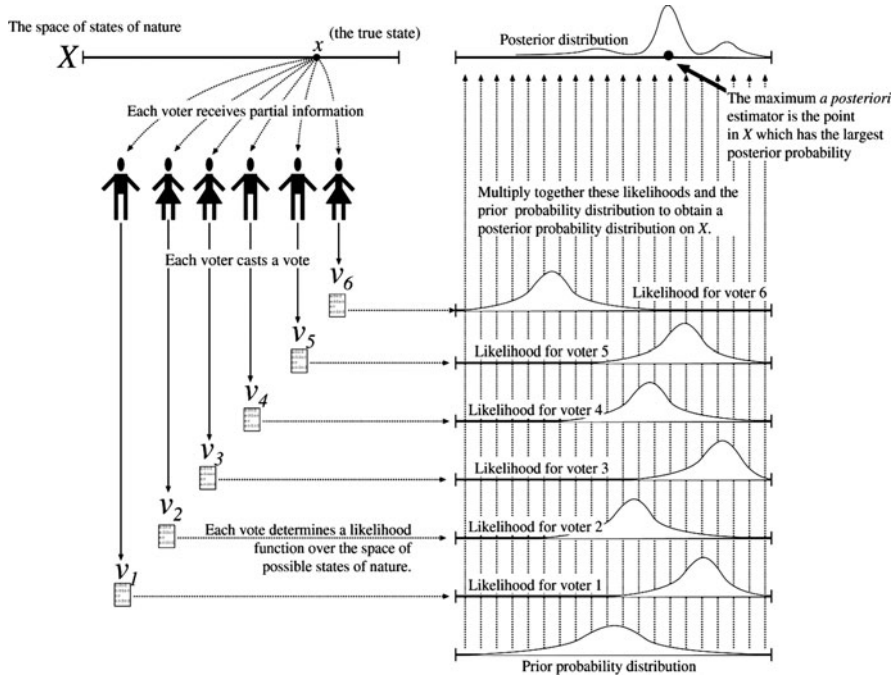


Fig. 2. A maximum a posteriori estimator.

with the voter error model encoded in the scenario *Sc*, but in the absence of any particular prior probability measure. (In the meteorology example in Table 4, we had MLE(*Sc*,**v**) = {*S*}.)

STATISTICAL RATIONALIZATION OF VOTING RULES

Let us say that a voting rule *F* is *MAP-rationalizable* if there is some error model (describing the fallibility of the voters' judgments) and some *a priori* probability distribution on *X (*describing our prior knowledge of the situation) such that, for *any* profile **v** of signals from the voters, the outcome *F*(**v**) is the maximum *a posteriori* estimator in *X*, given the data in **v**, according to this error model and prior distribution.[14] In other words, the decisions made by *F* can always be 'rationalized' as the outcome of a statistical procedure, based on some scenario describing the information available to the voters.

Let us say that the rule *F* is *MLE-rationalizable* if this rationalization can be made with a *uniformly distributed* prior probability distribution. This means that there is some error model (describing the fallibility of the voters) such that, for any profile **v** of signals, the rule *F* computes the maximum likelihood estimator in *X*, given the data in **v**. Finally, let us say

14 Formally, *F* is *MAP-rationalizable* if there exists some scenario *Sc* such that *F*(**v**) = MAP(*Sc*,**v**) for *every* profile **v** in V^I.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

that the rule F is *anonymously* MAP-rationalizable if the rationalizing error model is anonymous. Pivato (2012) proves the following result.

> **Rationalization Theorem:** A voting rule is MAP-rationalizable if and only if it is a scoring rule. Furthermore, it is anonymously MAP-rationalizable if and only if it is an anonymous scoring rule. The rule has no vetoes if and only if the corresponding error model has no impossibilities. Finally, if the rule is neutral, then it is MLE-rationalizable.[15]

In particular, this theorem implies that the (anti)plurality, Borda, Kemeny, approval, and range-voting rules are all anonymously MLE-rationalizable. The restaurant voting rule (Table 1) and the weighted plurality rule are (non-anonymously) MLE-rationalizable. Thus, the application of each of these voting rules to epistemic social decisions can be justified by positing a suitable error model, describing the fallibility of the voters. (Of course, sometimes this error model is not very plausible.)

If a voting rule *cannot* be represented as a scoring rule, then the Rationalization Theorem tells us that it *cannot* be MAP-rationalized. But some voting rules which do not initially look like scoring rules might admit some obscure scoring representation. Thus, the Rationalization Theorem is not very useful unless there is a simple way to determine whether a voting rule is a scoring rule. One way is to look for a property called *reinforcement*. A voting rule F satisfies this property if the following is true: given two disjoint populations, if each of these populations, in isolation, selects an alternative x when it uses the rule F, then the *combined* population should *also* select the alternative x when it uses the rule F. All scoring rules satisfy reinforcement. (Indeed, results of Myerson (1995) and Pivato (2011) imply that, amongst neutral voting rules, scoring rules are essentially the *only* rules which satisfy reinforcement.) Thus, a rule which violates reinforcement *cannot* be MAP-rationalizable. This eliminates many popular voting rules, such as *plurality vote with runoff* (used in the French presidential election), *instant runoff* (used in the Irish and Indian presidential elections), *successive elimination* rules (used in many sports tournaments, and also used by the US Congress to decide between a bill and its amendments), the *Copeland rule* (choose the alternative which beats the most other alternatives in a two-way vote), the *Simpson-Kramer rule* (choose the alternative with the smallest majority opposing it in a two-way vote), the *Schulze rule*, and the *Slater rule*. This makes it very hard to defend the use of any of these voting rules in an epistemic context.

METRIC VOTING RULES AND METRIC ERROR MODELS

The fact that a voting rule *can* be rationalized by some scenario does not imply that this scenario is a particularly realistic description of the epistemic problem facing society. What would a realistic description look like? Typically, the space X of alternatives has a sort of 'geometry'; there is a sense in which some elements of X are 'close together'

15 Pivato (2012: Theorem 2.2). Note that neutrality is sufficient, but not *necessary*, for MLE-rationalizability. The necessary and sufficient condition for MLE-rationalizability is a weak form of neutrality called *balance;* its definition is somewhat technical, so it is omitted.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

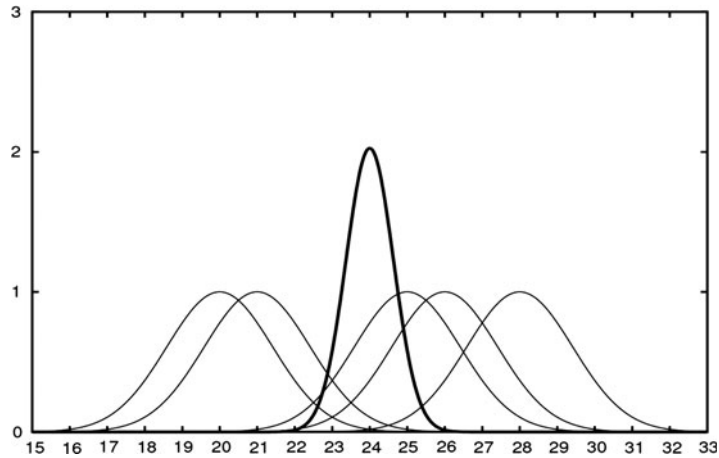


Fig. 3. Predicting tomorrow's temperature with Gaussian random errors.

(i.e. very similar), whereas other elements are 'far apart' (very dissimilar).[16] Suppose each voter's signal is a guess about the correct element in X (so in this model, $V = X$). If the voters are not completely unreliable, then they are more likely to guess an answer which is close to the right answer than one which is far away.

For example, suppose five fallible, but independent and equally competent meteorologists want to predict tomorrow's temperature. Alice predicts a temperature of 20 degrees Celsius, Bob predicts 21 degrees, Chiara predicts 25 degrees, Dave predicts 26 degrees, and Emily predicts 28 degrees. Suppose each meteorologist's prediction is the true temperature, plus a random error. Suppose these errors are independent, identically distributed normal (i.e. Gaussian) random variables.[17] Then the maximum likelihood temperature, given the profile (20,21,25,26,28), is 24 degrees – the *mean* of the five estimates. In Figure 3, the (Gaussian) likelihood functions of the five meteorologists are the five smaller thin curves. The thick curve is product of these five functions (rescaled); it is the group likelihood function. Note that it obtains its maximum at 24. Note also that the group likelihood function has a much narrower peak than the individual likelihood functions. In other words, it has a smaller *variance*, indicating a greater degree of certainty. Indeed, if we had aggregated the estimates from a *larger* group of (equally competent, statistically independent) meteorologists, then the variance would be even *smaller* (indicating even greater certainty). In fact, the variance approaches zero (perfect certainty) as the number of meteorologists grows to infinity. This illustrates the *wisdom of crowds* effect, first observed by Galton (1907), and recently popularized by Surowiecki (2004).

Now suppose instead that the meteorologists' errors are independent, identically distributed *exponential* random variables.[18] Then the maximum likelihood estimator is 25 degrees – the *median* of the five estimates. In Figure 4, the (exponential) likelihood

16 Formally, this means X is a *metric space*.

17 E.g. suppose that, if the true temperature is x, then the probability of predicting temperature v is given by $p(v|x) = C \exp[-(v-x)^2/4]$ (where C is some normalization constant).

18 E.g. suppose that, if the true temperature is x, then the probability of predicting temperature v is given by $p(v|x) = C\,(\frac{1}{2})^{|v-x|}$ (where C is some normalization constant).

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

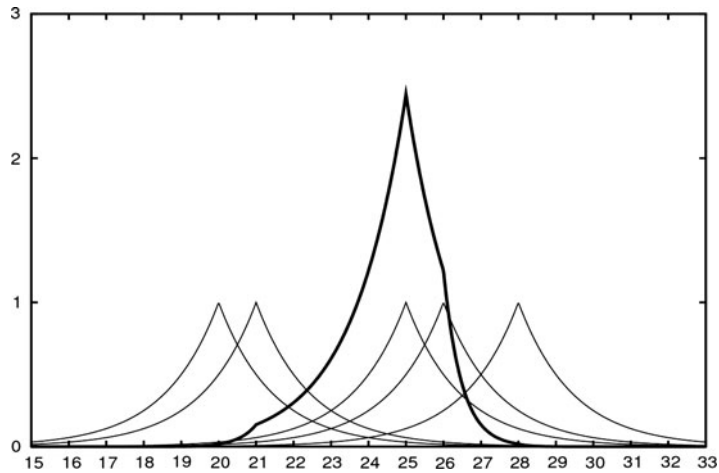


Fig. 4. Predicting tomorrow's temperature with exponential random errors.

functions of the five meteorologists are the five thin curves. The thick curve is product of these five functions (rescaled); it is the group likelihood function. Note that it obtains its maximum at 25.

In both of these examples, the meteorologists' mistakes were described by a *metric* error model. Formally, this means that, for every voter i in I, there is a (decreasing) real-valued function E_i such that, if the true state is x, then for any other state v in X, the probability that voter i guesses v is given by $E_i[d(x,v)]$, where $d(x,v)$ is the 'distance' between x and v in X.[19] Metric error models seem quite plausible; they capture the idea that reasonably competent voters are less likely to make large errors than small errors. If the space X of states of nature has a natural geometric structure (e.g. a temperature scale), then almost any reasonable explanation of how the voters make their mistakes will result in a metric error model.

In both the temperature examples, the MLE can be described as the outcome of a scoring rule. In the Gaussian error model, the score function is given by $S(i,v,x) = -(v-x)^2$, for every meteorologist i and all possible temperatures v and x. (For example, $S(i, 22, 25) = -3^2 = -9$.) In the exponential error model, the score function is given by $S(i,v,x) = -|v-x|$, for every meteorologist i and all possible temperatures v and x. (For example, $S(i, 22, 25) = -|22-25| = -3$.) Note that, in both cases, $S(i,v,x)$ depends only on the *distance* between v and x; furthermore, the score decreases as the distance increases. A scoring rule with this structure is called a *metric* voting rule: for each voter i in I and all states x and v in X, the value of $S(i,v,x)$ is a decreasing function of the distance between v and x. (In both temperature examples, the score function was *anonymous*, but this need not be the case in general.)

19 E.g. in the Gaussian error model shown in Figure 3, we had $E_i(r) = C \exp(-r^2/4)$ for any distance $r \geq 0$ and every meteorologist i. In the exponential error model shown in Figure 4, we had $E_i(r) = C\,(1/2)^r$, for any distance $r \geq 0$ and every meteorologist i. In both the temperature examples, the error model was *anonymous*, but this need not be the case in a general metric error model.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

These examples illustrate a general pattern: the MLE or MAP for any metric error model is a metric voting rule. Conversely, any metric voting rule can be MAP-rationalized in terms of a metric error model. Thus, the Rationalization Theorem becomes much more plausible and compelling in the setting of metric voting rules.

In particular, suppose, in the metric error model, that the functions E_i are *exponentially decaying* for all voters i in I.[20] Then it can be shown that (for a suitable prior probability distribution) the MAP estimator is the *weighted median rule*.[21] For any profile $\mathbf{v}$, the weighted median rule picks the element(s) of X which *minimize* the sum $w_1\ d(x,v_1) + w_2\ d(x,v_2) + \ldots + w_N\ d(x,v_N)$, where $w_1,w_2,\ldots,w_N$ are nonnegative 'weights' assigned to the voters. In other words, the weighted median rule chooses the state in X which has the *smallest weighted average distance* to the opinions of the voters. The rule is *anonymous* (i.e. treats all voters the same) if and only if $w_1 = w_2 = \ldots = w_N = 1$;[22] this yields the (unweighted) *median rule*, which minimizes the sum $d(x,v_1) + \ldots + d(x,v_N)$. (Thus, the median rule chooses the state in X which has the *smallest average distance* to the opinions of the voters.) The temperature estimation example in Figure 4 illustrated the median rule on the set $\mathbb{R}$ of all real numbers (i.e. all possible temperatures). A space X is called *homogeneous* if the geometry of X 'looks the same' around each alternative x in X.[23] For example, the space $\mathbb{R}$ is homogeneous. In this case, as we saw above, the median rule is the MLE for any anonymous exponential error model.[24]

For another example, let X be the space of all strict preference orders on some set A of alternatives. For any x and y in X, let $d(x,y)$ be the number of pairs in A on which the orders of x and y disagree. Then the structure described by X and d is a homogeneous space. The median rule on X is the *Kemeny rule*. To illustrate this, consider a set of three alternatives (i.e. $A = \{a,b,c\}$). Recall that each preference order on $\{a,b,c\}$ is entirely described by three binary comparisons: a vs. b, b vs. c, and c vs. a. Thus, for any two preference orders x and y, we have $d(x,y) = 0$ if $x = y$, while $d(x,y) = 3$ if x is the exact reverse of y; otherwise $d(x,y) = 1$ or 2. Table 5 portrays the same seven-voter profile which I earlier used to illustrate the Kemeny rule in Table 3. The median preference order is $b > c > a$, which attains the smallest average distance (of 9/7) to the preferences of the voters. Observe that this agrees with the output of the Kemeny rule, which we computed in Table 3.

The space X of all preference orders is homogeneous. Thus, the Kemeny (i.e. median) rule is the MLE for *any* anonymous exponential error model on X; this was first noted by Young (1986, 1988, 1995, 1997).[25] For example, consider an exponential error model on these preference orders, such that, if the true preference order is x, then the probability of a voter choosing preference order v is given by $p(v|x) = C\ (\frac{1}{2})^{d(v,x)}$ (where $C = 8/21$).[26] For the seven-voter profile above, the likelihoods are given by Table 6. The maximum

20 This means, for all i in I, there are constants $C_i > 0$ and $b_i > 0$ such that $E_i(r) = C_i/b_i^r$, for all $r > 0$.

21 See Pivato (2012: section 4).

22 This is equivalent to requiring $C_1 = C_2 = \ldots = C_N$ and $b_1 = b_2 = \ldots = b_N$.

23 Formally: for any x and y in X, there is an isometry of X which maps x into y. (An *isometry* is a bijection from X to itself which preserves distances.) See Pivato (2012: Corollary 3.2).

24 This means there are constants $C > 0$ and $b > 0$ such that, for all voters i in I, we have $E_i(r) = C/b^r$. In the temperature prediction example, we had $b = 2$.

25 See also Pivato (2012: Example 4.6).

26 We multiply by $C = 8/21$ to ensure that the probabilities of all possible preference orders sum up to 1, as required for a probability distribution.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

Table 5. The Kemeny rule as a median rule

Voter	Actual preferences	Distance to $a>b>c$	Distance to $b>c>a$	Distance to $c>a>b$	Distance to $c>b>a$	Distance to $b>a>c$	Distance to $a>c>b$
Alice	$a>b>c$	0	2	2	3	1	1
Bob	$b>c>a$	2	0	2	1	1	3
Chiara	$c>a>b$	2	2	0	1	3	1
Dave	$c>a>b$	2	2	0	1	3	1
Emily	$b>a>c$	1	1	3	2	0	2
Fred	$b>a>c$	1	1	3	2	0	2
Gwenn	$c>b>a$	3	1	1	0	2	2
Average distance		11/7	**9/7**	11/7	10/7	10/7	12/7

Table 6. The Kemeny rule as a maximum likelihood estimator

Voter	Actual preferences	Likelihood of $a>b>c$	Likelihood of $b>c>a$	Likelihood of $c>a>b$	Likelihood of $c>b>a$	Likelihood of $b>a>c$	Likelihood of $a>c>b$
Alice	$a>b>c$	C	$C/4$	$C/4$	$C/8$	$C/2$	$C/2$
Bob	$b>c>a$	$C/4$	C	$C/4$	$C/2$	$C/2$	$C/8$
Chiara	$c>a>b$	$C/4$	$C/4$	C	$C/2$	$C/8$	$C/2$
Dave	$c>a>b$	$C/4$	$C/4$	C	$C/2$	$C/8$	$C/2$
Emily	$b>a>c$	$C/2$	$C/2$	$C/8$	$C/4$	C	$C/4$
Fred	$b>a>c$	$C/2$	$C/2$	$C/8$	$C/4$	C	$C/4$
Gwenn	$c>b>a$	$C/8$	$C/2$	$C/2$	C	$C/4$	$C/4$
Group likelihood		$C^7/2048$	$C^7/512$	$C^7/2048$	$C^7/1024$	$C^7/1024$	$C^7/4096$

likelihood preference order is $b>c>a$, which attains a likelihood of $C^7/512$. Observe that this agrees with the output of median and the Kemeny rule.[27]

Here is another example. Suppose we must pick a committee with two members from a set of four candidates: *a, b, c*, and *d*. Let V and X both be the set of all admissible committees; that is, $V = X = \{ \{a,b\}, \{a,c\}, \{a,d\}, \{b,c\}, \{b,d\}, \{c,d\} \}$. For any two committees x and y in X, let $d(x,y)$ be the number of members on which they disagree (for example, $d(\{a,b\},\{b,c\}) = 1$). Consider the seven-voter profile described by Table 7. The median committee is $\{a,b\}$, which attains the minimal average distance (of 6/7) to the preferred committees of the voters.

Now suppose there is some committee which is truly the best choice. Consider an exponential error model such that, if x is truly the best committee, then the probability of a

27 This MLE interpretation of the Kemeny rule depends on the premise that there is an objectively 'correct' preference order. This premise is not suitable in all contexts. In many contexts, preference orders are best seen as the subjective motivational attitudes of the voters; hence these preferences cannot be 'correct' or 'incorrect'. However, there are some situations where an objectively correct preference order might exist. E.g. suppose that all the voters agree on the ultimate goal, and A is a set of possible strategies to achieve this goal, and the voters want to rank these strategies in terms of how effective they will be in reaching the goal. Or suppose A is a set of contestants (e.g. for prizes, scholarships, jobs, or places in a prestigious school), and the voters are trying to rank these contestants according to some (presumably) objective notion of 'quality' (e.g. ability).

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

Table 7. The median rule for committee selection

Voter	Preferred committee	Distance to *{a,b}*	Distance to *{a,c}*	Distance to *{a,d}*	Distance to *{b,c}*	Distance to *{b,d}*	Distance to *{c,d}*
Alice	*{a,b}*	0	1	1	1	1	2
Bob	*{a,c}*	1	0	1	1	2	1
Chiara	*{a,d}*	1	1	0	2	1	1
Dave	*{a,d}*	1	1	0	2	1	1
Emily	*{b,d}*	1	2	1	1	0	1
Fred	*{b,c}*	1	1	2	0	1	1
Gwenn	*{b,c}*	1	1	2	0	1	1
Average distance		**6/7**	1	1	1	1	8/7

Table 8. The MLE committee selection

Voter	Preferred committee	Likelihood of *{a,b}*	Likelihood of *{a,c}*	Likelihood of *{a,d}*	Likelihood of *{b,c}*	Likelihood of *{b,d}*	Likelihood of *{c,d}*
Alice	*{a,b}*	C	$C/3$	$C/3$	$C/3$	$C/3$	$C/9$
Bob	*{a,c}*	$C/3$	C	$C/3$	$C/3$	$C/9$	$C/3$
Chiara	*{a,d}*	$C/3$	$C/3$	C	$C/9$	$C/3$	$C/3$
Dave	*{a,d}*	$C/3$	$C/3$	C	$C/9$	$C/3$	$C/3$
Emily	*{b,d}*	$C/3$	$C/9$	$C/3$	$C/3$	C	$C/3$
Fred	*{b,c}*	$C/3$	$C/3$	$C/9$	C	$C/3$	$C/3$
Gwenn	*{b,c}*	$C/3$	$C/3$	$C/9$	C	$C/3$	$C/3$
Group likelihood		$C^7/729$	$C^7/2187$	$C^7/2187$	$C^7/2187$	$C^7/2187$	$C^7/6561$

voter choosing the committee v is given by $p(v|x) = C\,(\frac{1}{3})^{\mathrm{d}(v,x)}$ (where $C = 9/46$).[28] For the seven-voter profile above, the likelihoods are shown in Table 8. The maximum likelihood committee is $\{a,b\}$, which attains a likelihood of $C^7/729$. Note that this agrees with the outcome of the median rule, computed in Table 7.

This example illustrates a more general pattern. Let A be a set of possible candidates for some committee. (In Tables 6 and 7, we had $A = \{a,b,c,d\}$.) Let X be the set of all committees comprised of exactly n of these candidates (for some fixed $n \leq |A|$). For any two committees x and y in X, let $d(x,y)$ be the number of candidates on which the committees x and y disagree. Then the structure described by X and d is a homogeneous space. Thus, the median rule is the MLE for any anonymous exponential error model on X.[29]

An MLE rationalization of the median rule is only valid if X is *homogeneous* in the sense described above. But most metric spaces are *not* homogeneous. On inhomogeneous

28 We multiply by $C = 9/46$ to ensure that the probabilities of all possible committees sum up to 1, as is required for a probability distribution.

29 We can also MLE-rationalize the median rule on the space of all committees with an *odd* number of members, or on the space of all committees having either n or $n+1$ members out of a set of $2n+1$ candidates; see Pivato (2012: Example 4.8) for details.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

spaces, the median rule can still be MAP-rationalized, but the prior probability distribution will generally *not* be uniform. For instance, in the temperature-prediction example, it is unrealistic to suppose that the temperature could be *any* real number. It certainly must be greater than −100 degrees Celsius (at which temperature carbon dioxide would freeze and snow out of the atmosphere). Furthermore, it is most likely less than 100 degrees Celsius on the surface of the earth (except perhaps near intense geothermal activity). So perhaps we should let X be the interval of all real numbers between −100 and 100. But this interval is *not* homogeneous, so if we did this, then the median rule would no longer be MLE-rationalizable. If the number of voters was relatively small, and their predictions were far from the boundaries of the interval (as was the case in our example), then the deviation from a uniform prior probability distribution would be very small, so the median would still agree with the MLE almost all the time. However, as the number of voters becomes large, the rationalizing prior probability distribution will become increasingly concentrated near the exact center of the interval (which is 0 degrees, in this case). In this case, the median temperature estimate will often deviate from the maximum likelihood estimator.[30]

For another example, consider the set of all committees containing at least two members and at most four members, drawn from a set of ten possible candidates. Unlike the committee space we considered in Tables 6 and 7, this committee space is *not* homogeneous. The median rule on this space is MAP-rationalizable (with a non-uniform prior probability distribution), but it is *not* MLE-rationalizable.[31]

The Borda rule is another metric voting rule with an interesting statistical rationalization. To illustrate this, let $A = \{a,b,c\}$ be a set of three alternatives, and consider the seven-voter profile of preferences we used in Tables 3, 5 and 6 to illustrate the Kemeny rule. Table 9 shows the same profile, with the voters' preference orders represented as *ranking vectors* of $\{a,b,c\}$; that is, three-dimensional vectors (r_a, r_b, r_c), such that, for all n in {1,2,3}, there is exactly *one* element x in $\{a,b,c\}$ such that $r_x = n$. (Lower ranks are better.) For every voter i in I, every signal y in X, and every alternative x in X, let $S(i,y,x) := -|x-y|^2$; then the resulting scoring rule is a metric voting rule, because the score function $S(i,y,x)$ only depends on the Euclidean distance from x to y.[32] In fact, this scoring rule is the *Borda rule*, expressed in terms of ranking vectors. In the example in Table 9, the Borda rule chooses the ranking vector (3,1,2), which achieves the maximal score (namely −24). This corresponds to the preference order $b > c > a$.

Now suppose that one ranking of the set $\{a,b,c\}$ is the 'true' ranking. Consider an error model such that, if x is truly the best ranking, then the probability of a voter choosing the ranking v is given by $p(v|x) = C \exp(-|v-x|^2)$ (where C is a constant).[33] For the seven-voter profile in Table 9, the likelihoods are given by Table 10. The correspondence with the Table 9 is obvious. The last row of Table 10 shows the group likelihoods of the six possible preference orders in X, which are obtained by multiplying the likelihoods

30 Pivato (2012: section 4.1).

31 Ibid., Example 4.9(a).

32 If $x = (x_a x, {}_b x, {}_c)$ and $y = (y_a y, {}_b y, {}_c)$ are three-dimensional vectors, then $|x-y|^2 = (x_a - y_a)^2 + (x_b - y_b)^2 + (x_c - y_c)^2$ is the squared Euclidean distance from x to y.

33 To be precise, $C = 1/(1 + 2(e^{-2} + e^{-4}) + e^{-8})$. We multiply by C to ensure the the probabilities of all rankings sum to 1, as required for a probability distribution.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

Table 9. The Borda rule as a metric voting rule (metric error model on ranking vectors)

Voter	Actual ranking	Score of (1,2,3)	Score of (3,1,2)	Score of (2,3,1)	Score of (3,2,1)	Score of (2,1,3)	Score of (1,3,2)
Alice	(1,2,3)	0	−6	−6	−8	−2	−2
Bob	(3,1,2)	−6	0	−6	−2	−2	−8
Chiara	(2,3,1)	−6	−6	0	−2	−8	−2
Dave	(2,3,1)	−6	−6	0	−2	−8	−2
Emily	(2,1,3)	−2	−2	−8	−6	0	−6
Fred	(2,1,3)	−2	−2	−8	−6	0	−6
Gwenn	(3,2,1)	−8	−2	−2	0	−6	−6
Total score		−30	**−24**	−30	−26	−26	−32

Table 10. The Borda rule as an MLE (Gaussian error model on ranking vectors)

Voter	Actual ranking	Likelihood of (1,2,3)	Likelihood of (3,1,2)	Likelihood of (2,3,1)	Likelihood of (3,2,1)	Likelihood of (2,1,3)	Likelihood of (1,3,2)
Alice	(1,2,3)	C	$C \exp(-6)$	$C \exp(-6)$	$C \exp(-8)$	$C \exp(-2)$	$C \exp(-2)$
Bob	(3,1,2)	$C \exp(-6)$	C	$C \exp(-6)$	$C \exp(-2)$	$C \exp(-2)$	$C \exp(-8)$
Chiara	(2,3,1)	$C \exp(-6)$	$C \exp(-6)$	C	$C \exp(-2)$	$C \exp(-8)$	$C \exp(-2)$
Dave	(2,3,1)	$C \exp(-6)$	$C \exp(-6)$	C	$C \exp(-2)$	$C \exp(-8)$	$C \exp(-2)$
Emily	(2,1,3)	$C \exp(-2)$	$C \exp(-2)$	$C \exp(-8)$	$C \exp(-6)$	C	$C \exp(-6)$
Fred	(2,1,3)	$C \exp(-2)$	$C \exp(-2)$	$C \exp(-8)$	$C \exp(-6)$	C	$C \exp(-6)$
Gwenn	(3,2,1)	$C \exp(-8)$	$C \exp(-2)$	$C \exp(-2)$	C	$C \exp(-6)$	$C \exp(-6)$
Group likelihood		$C^7 \exp(-30)$	C^7 **exp(-24)**	$C^7 \exp(-30)$	$C^7 \exp(-26)$	$C^7 \exp(-26)$	$C^7 \exp(-32)$

from the seven previous rows.[34] The maximum likelihood ranking is again (3,1,2), in agreement with the Borda ranking.[35]

This example illustrates a much more general pattern, which can be seen as a generalization of the earlier temperature prediction example with Gaussian errors (Figure 3). Let M be a positive integer, and let X be any homogeneous subset of M-dimensional Euclidean space (i.e. $\mathbb{R}^M$), which we endow with the standard Euclidean notion of distance. (In the temperature estimation example in Figure 3, we had $M = 1$ and $X = \mathbb{R}$; in Table 10, we have $M = 3$ and $X = \{(1,2,3), (3,1,2), \ldots, (3,2,1)\}$.) Consider an anonymous *Gaussian* error model on X.[36] The MLE for this error model is the metric voting rule where $S(i,y,x) = -|x-y|^2$ for every voter i in I, every signal y in X, and every alternative x in X.[37] In this case, maximizing the total score is equivalent to minimizing the mean squared distance to the votes sent by the voters. It is easy to check[38] that this scoring

34 In these computations, we use the key fact that $\exp(r)\exp(s) = \exp(r+s)$, for any real numbers r and s.

35 If $r > s$, then $\exp(r) > \exp(s)$. Thus, $\exp(-24) > \exp(-26) > \exp(-30) > \exp(-32)$, because $-24 > -26 > -30 > -32$.

36 Formally, this means there are some constants $\sigma > 0$ and $C > 0$ such that $E_i(r) = \exp(-r^2/2\sigma^2)/C$ for every voter i in I and every $r > 0$. In the example in Table 10, we have $\sigma = 1$ and $C = 1/(1 + 2(e^{-2} + e^{-4}) + e^{-8})$.

37 If $x = (x_1, x_2, \ldots, x_M)$ and $y = (y_1, y_2, \ldots, y_M)$ are two vectors in $\mathbb{R}^M$, then $|x-y|^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_M - y_M)^2$.

38 Pivato (2012: section 3.2).

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

rule simply chooses the element(s) of X which are closest to the *average* of the signals v_1, v_2,...,v_N, when treated as M-dimensional vectors in $\mathbb{R}^M$.[39]

In particular, let A be a set of M social alternatives. A *ranking vector* of A is an M-dimensional vector r (with co-ordinates indexed by the elements of A) such that, for all m in $\{1,2,\ldots,M\}$, there is exactly *one* element a in A such that $r_a = m$. (Lower ranks are better.) We can regard r as a vector in the M-dimensional Euclidean space $\mathbb{R}^M$. Let X be the set of all such ranking vectors; then X is a subset of $\mathbb{R}^M$. Then X is homogeneous, and the averaging rule (i.e. the MLE for any Gaussian error model) is the *Borda rule*, as we saw in the example from Tables 9 and 10 (where $A = \{a,b,c\}$, so that $M = 3$). This argument provides another MLE-rationalization for the Borda rule, which is quite different from the rationalization given by Young (1986). Variations of this setup yield MLE-rationalizations of other 'ranking' rules, including the plurality rule and anti-plurality rule portrayed in Table 2.

RATIONALIZING VOTING RULES AS EXPECTED UTILITY MAXIMIZERS

Now let us turn to the problem of expected utility maximization. Many voting rules invite each voter to assign a numerical 'score' to each social alternative. Then the rule chooses the alternative with the highest average score. Rules differ on what range of scores are admissible. For example:

- *Classical utilitarianism* allows the score of an alternative to be any real number.
- *Range voting* allows the score of an alternative to be any real number between 0 and 1.
- *Approval voting* requires the score of each alternative to be either 0 or 1.

In all of these rules, a voter can give the *same* score to two or more alternatives (unlike the Borda rule, (anti)plurality rule, and other 'ranking' rules). I will refer to rules with this property as *quasiutilitarian*.

Any quasiutilitarian voting rule can be rationalized as an expected utility maximizer. Suppose that the score a voter assigns to an alternative reflects her (possibly inaccurate) estimate of the 'social utility' of that alternative. Although the voters' estimates may be incorrect, we can assume that they are not totally perverse: under reasonable assumptions about the voters' (mis)perceptions, one can show that, if alternative a receives an average score of r, and alternative b receives an average score of s, and $r > s$, then the *conditionally expected utility* of a, given this information, is higher than the conditionally expected utility of b. (This does *not* mean that r itself is a good estimate of the expected utility of a – but this isn't necessary.) I will say an error model with this property is *regular*. It is easy to see that a quasiutilitarian rule is an expected utility maximizer for *any* regular error model. In particular, this yields EUM-rationalizations of classical utilitarianism, range voting, approval voting, and a 'relaxed' version of the Borda rule (where a voter can give the same rank to two or more alternatives).[40]

39 E.g. the average of the vectors (1,2,3), (2,3,1), and (2,1,3) is the vector (5/3, 2, 7/3).

40 Pivato (2012: section 5).

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

For example, consider a hiring committee which must choose one of four candidates for a position. The four candidates are Walid, Xavier, Yvonne and Zara (so $X = \{w,x,y,z\}$). The hiring committee is the same seven voters who have appeared in previous examples. Suppose each candidate is either competent to do the job, or incompetent – there are no finer gradations of skill. Hiring a competent candidate yields a utility of 1; hiring an incompetent candidate yields a utility of 0. The hiring committee simply wants to maximize its expected utility – or equivalently, maximize the probability that it hires a competent candidate. Each voter makes a private estimate of the competency of each candidate: either 1 if she thinks the candidate is competent, or 0 if she thinks he is incompetent. Suppose this estimate is correct with probability ⅔, and incorrect with probability ⅓. We also suppose the *true* competencies of the candidates are independent, identically distributed random variables. To be concrete, suppose each candidate is competent with probability ¼, and incompetent with probability ¾.

Consider the profile of voter opinions described by Table 11. (For example, in this profile, Alice believes that Walid is competent, but she believes that the other three candidates are incompetent.) In the second-last line of Table 11, we have summed up the scores for each candidate. In effect, this is each candidate's score in the *approval voting* procedure. Observe that Xavier is the winner of the approval voting procedure, attaining a score of 4. In the last line of Table 11, we have also computed the conditionally expected utility for each candidate, given the opinions of the voters and based on the scenario we described above. Equivalently, this is the conditional probability that each candidate is competent. (The details of this computation are omitted.) Xavier is competent with probability 2/5, which is much higher than the competency probability of any other candidate. Thus, hiring Xavier will maximize expected utility. This illustrates a general pattern: in hiring problems of this kind, the approval voting procedure will always chose the candidate who yields the maximal expected utility.[41]

If we suppose that the competency of each candidate can be *any* value between 0 and 1, then we can obtain a very similar EUM interpretation of *range voting* instead of approval voting. We can likewise interpret the classic utilitarian rule as an expected utility maximizer. Unlike the Rationalization Theorem, these results do *not* depend upon a specific choice of error model; they are true for *any* 'regular' error model, as defined above. Thus, under quite plausible conditions, these rules will maximize the expected level of social utility. This provides a very strong argument for using these rules in social decision problems where the group agrees on an underlying utility function, but is uncertain about the best course of action.

CONCLUSION

As these results illustrate, a large variety of voting rules can be 'rationalized' as MAP, MLE or EUM rules for some prior probability distribution and some more-or-less plausible model of the voters' error patterns. However, there are some foundational problems with this entire program. First of all, this program begins with a familiar voting rule, and then 'rationalizes' it with some probabilistic scenario, after the fact. But this is the *reverse*

41 Pivato (2012: Example 5.2).

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

Table 11. The hiring problem (approval voting as an expected utility maximizer)

		Candidates			
		Walid	**Xavier**	**Yvonne**	**Zara**
Voters	Alice	1	0	0	0
	Bob	1	0	0	1
	Chiara	1	1	0	0
	Dave	0	1	0	1
	Emily	0	1	0	0
	Fred	0	1	1	1
	Gwenn	0	0	1	0
Total score		3	**4**	2	3
Conditionally expected utility		1/7	**2/5**	1/25	1/7

of the correct procedure. One should *begin* by specifying a prior probability distribution and an error model for the voters which captures the underlying epistemic problem as realistically as possible, and *then* compute the MLE/MAP/EUM for this model; this may or may not end up being a familiar voting rule.

Second of all, these results (as well as the aforementioned work by Young, Conitzer, Xia, and others) assume that the mistakes made by different voters are *independent* random variables. But this is totally unrealistic, because voters presumably come from similar cultural and educational backgrounds, draw upon the same body of public (mis)information and are in constant communication with one another. Furthermore, a psychological desire to conform or avoid conflict may lead to 'herding' or 'groupthink'. Lorenz et al. (2011) have shown empirically that such 'social influences' can undermine the reliability of epistemic social choice mechanisms. Thus, any plausible error model must involve correlated errors. But it is not clear how to realistically model such correlations, and a realistic model might be mathematically intractable. Dietrich and Spiekerman (2011) have extended the Condorcet Jury Theorem to a model where voters draw upon common information sources. It would be fruitful to extend their approach to other epistemic social choice problems.

One proposed solution to the problem of error correlation is to require the voters to formulate their opinions in isolation, without any communication or deliberation. But aside from its infeasibility in most practical situations, this solution is undesirable because it squanders one of the main assets of social epistemology: group cognition. Ideally, through deliberation and epistemic specialization, the voters could form a collective cognitive entity, which would be more intelligent than any of its components. But the cognitive machinery of this collective entity would involve forms of information exchange and processing much more complex than mere voting.

There is a third problem. We have assumed throughout that all the voters have the same utility function. But in practice, this is rarely true. Some voters may have obvious conflicts of interest (e.g. a doctor may have a financial relationship with certain pharmaceutical companies, or a government economist may anticipate future employment opportunities in the private sector). Others may have some psychological stake in the outcome (e.g. a doctor might prefer the treatment which is most suited to her own special expertise). And most voters' utility functions will be influenced by their own moral values, aesthetic

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

tastes, personal preferences and feelings of solidarity towards their economic class or cultural group. If a voter's private preferences deviate from the preferences of the group, then she may try to manipulate the group's decision by misrepresenting her own beliefs. Even in the unlikely event that all the voters have the exactly the same preferences, they may *still* have incentives to exaggerate their views or suppress countervailing evidence, in order to counteract what they believe to be the misperceptions of the other ('sincere but misguided') voters; see Austen-Smith and Banks (1996).

Finally, the 'statistical' approach to epistemic democracy assumes that it is possible to specify, with reasonable accuracy, the prior probability distribution and the error model of the voters. But this is not realistic. Consider the debate over climate change. Supposing we wanted to obtain a collective judgement using the statistical techniques described in this paper. What is a reasonable prior? And how could we even begin to specify an error model for the opinions of the climate scientists?

For some problems, a precise specification may not be necessary. For example, the Condorcet Jury Theorem and the aforementioned EUM-rationalizations of quasiutilitarian rules are each valid for a large class of scenarios. But the MAP-rationalization described by the Rationalization Theorem can often be quite sensitive to the underlying scenario. Perhaps, then, we should not seek a voting rule which is the 'optimal' statistical estimator for one particular scenario, but rather, seek a rule which is 'reasonably reliable' for a broad spectrum of scenarios.

However, to consider even a 'spectrum' of scenarios, we must first specify the underlying space of possible answers, and in some cases, there is no obviously correct way to do this. For example, consider a debate amongst scientists over the correct model of some phenomenon (e.g. climate). It is far from clear how to specify the 'space of possible scientific models'. Which structural parameters should be held fixed, and which should be allowed to vary? And how should these variables be coordinatized? There is no obvious answer to these questions, and the outcome of an MLE or MAP rule will be highly sensitive to how they are answered. In effect, the 'social consensus' would be an artifact of how we framed the debate.

Fortunately, the scientific community does not generally resolve dissensus by 'voting'. Instead, scientists deliberate, scrutinize each other's theories, and zero in on those lacunae in the empirical data which allow the dissensus to even exist. They then seek to fill these lacunae as efficiently as possible, through carefully designed experiments. This deliberation and empirical exploration continues until little or no dissensus remains, at which point the use of an epistemic social choice mechanism is no longer necessary.[42]

REFERENCES

Arrow, Kenneth J. 1963. *Individual Values and Social Choice*. New York: John Wiley & Sons.

Austen-Smith, D., and Banks, J. 1996. 'Information Aggregation, Rationality, and the Condorcet Jury Theorem.' *American Political Science Review*, 90: 34–45.

Bohman, J., and Rehg, W. (eds). 1997. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, MA: MIT Press.

42 I would like to thank Christian List and the participants of the 2011 EPISTEME Conference at Carnegie-Mellon University for many helpful comments.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press

Brams, S. J., and Fishburn, P. C. 1983. *Approval Voting*. Boston, MA: Birkhauser.
Cohen, J. 1986. 'An Epistemic Conception of Democracy.' *Ethics*, 97(1): 26–38.
Condorcet, M. d. 1785. *Essai sur l'application de l'analyse à la probabilite des décisions rendues à la pluralité des voix*. Paris.
Conitzer, V., and Sandholm, T. 2005. 'Common Voting Rules as Maximum Likelihood Estimators.' In *21st Annual Conference on Uncertainty in Artificial Intelligence* (UAI-05), pp. 145–52. Edinburgh: UAI.
Conitzer, V., Rognlie, M., and Xia, L. 2009. 'Preference Functions that Score Rankings and Maximum Likelihood Estimation.' In *21st International Joint Conference on Artificial Intelligence* (IJCAI-09), pp. 109–15. Pasadena, CA: IJCAI.
Dietrich, F., and Spiekerman, K. 2011. Epistemic Democracy with Defensible Premises. (preprint).
Estlund, D. 1997. 'Beyond Fairness and Deliberation: The Epistemic Dimension of Democratic Authority.' In Bohman and Rehg (1997: 173–204).
Galton, Francis. 1907. 'Vox Populi.' *Nature*, 75: 450–1.
Hummel, P. 2010. 'Jury Theorems with Multiple Alternatives.' *Social Choice and Welfare*, 34(1): 65–103.
Kaniovski, S. 2010. 'Aggregation of Correlated Votes and Condorcet's Jury Theorem.' *Theory and Decision*, 69(3): 453–68.
Kemeny, J. G. 1959. 'Math without Numbers.' *Daedalus*, 88: 571–91.
List, C., and Goodin, R. E. 2001. 'Epistemic Democracy: Generalizing the Condorcet Jury Theorem.' *Journal of Political Philosophy*, 9(3): 277–306.
Lorenz, Jan, Rauhut, Heiko, Schweitzer, Frank, and Helbing, Dirk. 2011. 'How Social Influence Can Undermine the Wisdom of Crowd Effect.' *Proceedings of the National Academy of Science*, 16 May, doi: 10.1073/pnas.1008636108.
Myerson, R. 1995. 'Axiomatic Derivation of Scoring Rules without the Ordering Assumption.' *Social Choice and Welfare*, 12(1): 59–74.
Pivato, M. 2011. 'Variable Population Voting Rules' (preprint).
Pivato, M. 2012. 'Voting Rules as Statistical Estimators.' *Social Choice and Welfare*.
Smith, Warren D. 2001. 'Range Voting': http://citeseerx.ist.psu.edu/viewdoc/summary? doi = 10.1.1.32.9150.
Surowiecki, James. 2004. *The Wisdom of Crowds*. New York: Doubleday.
Xia, L., and Conitzer, V. 2011. 'A Maximum Likelihood Approach towards Aggregating Partial Orders.' In *23rd International Joint Conference on Artificial Intelligence* (IJCAI-11). Pasadena, CA: IJCAI.
Xia, L., Conitzer, V., and Lang, J. 2010. 'Aggregating Preferences in Multi-Issue Domains by Using Maximum Likelihood Estimators'. In *9th International Joint Conference on Autonomous Agents and Multi Agent Systems* (AAMAS-10), pp. 399–406. Toronto: AAMAS.
Young, H. P. 1986. 'Optimal Ranking and Choice from Pairwise Decisions.' In B. Grofman and G. Owen (eds), *Information Pooling and Group Decision Making*., pp. 113–22. Greenwich, CT: JAI Press.
1988. 'Condorcet's Theory of Voting.' *American Political Science Review*, 82(4): 1231–44.
1995. 'Optimal Voting Rules.' *Journal of Economic Perspectives*, 9(1): 51–64.
1997. 'Group Choice and Individual Judgments.' In D. C. Mueller (ed.), *Perspectives on Public Choice: A Handbook*, pp. 181–200. Cambridge: Cambridge University Press.

MARCUS PIVATO received his Ph.D. in Mathematics from the University of Toronto in 2001, and is now a professor of Mathematics at Trent University, in Peterborough, Ontario, Canada. His main research interests are social choice and decision theory, and their applications to social welfare and social epistemology. He is also the author of the textbook *Linear partial differential equations and Fourier theory*, which was published by Cambridge University Press in 2010.

https://doi.org/10.1017/epi.2012.4 Published online by Cambridge University Press