The Premises of Condorcet's Jury Theorem Are Not Simultaneously Justified

Franz Dietrich

March 2008

Abstract

Condorcet's famous jury theorem reaches an optimistic conclusion on the correctness of majority decisions, based on two controversial premises about voters: they are competent and vote independently, in a technical sense. I carefully analyse these premises and show that: (i) whether a premise is justified depends on the notion of probability considered; (ii) none of the notions renders both premises simultaneously justified. Under the perhaps most interesting notions, the independence assumption should be weakened.

1 Introduction

Roughly stated, the classic *Condorcet Jury Theorem*[1] (CJT) asserts that if a group (jury, population, etc.) takes a majority vote between two alternatives of which exactly one is objectively 'correct', and if the voters satisfy two technical conditions, *competence* and *independence*, then the probability that the majority picks the correct alternative increases to one (certainty) as the group size tends to infinity. Though mathematically elementary, this result is striking in its overly positive conclusion. If majority judgments are indeed most probably correct in large societies, majoritarian democracy receives strong support from an epistemic perspective.

This paper goes back to the very basics and aims to answer whether the theorem's two premises are justified. The answer will be seen to depend in a rather clear-cut way on the kind of probability (uncertainty[2]) considered; unfortunately, in each case exactly one of the premises is not justified. A central distinction will be whether only voting behaviour or also the decision problem voters face is subject to uncertainty. I suggest that this distinction marks the difference between two versions in which the basic CJT can be found in the

---

[1] See Condorcet's (1785) writings at the dawn of the French Revolution.

[2] I use the term 'uncertainty' in a general sense, that is, not only when referring to someone's subjective uncertainty but also when referring to an objective probability.

literature;[3] I accordingly label these versions the *fixed-problem CJT* and the *variable-problem CJT*, respectively. In the fixed-problem CJT, competence is the problematic assumption, whereas in the variable-problem CJT, independence is problematic. So the two versions of the CJT, which might have appeared to be just notational variants, are in fact fundamentally distinct.

Let me start by sketching a tempting but sloppy argument that seems to support the CJT's two premises, and hence its striking conclusion. Consider, for instance, a group of judges in a collegial court facing an 'acquit or convict' choice in a criminal law case; 'convict' is correct if and only if the defendant has committed the crime. First, the CJT's competence assumption requires (roughly) that each voter's probability of making a correct judgment exceeds 1/2. While on a particular criminal law case a judge may easily be mistaken – say, if there is highly misleading evidence[4] – surely such cases are rather the exception than the rule, and so within the large class of related court cases a voter's rate (frequency) of correct judgments exceeds 1/2. Hence the competence assumption holds. Second, the CJT's independence assumption requires (roughly) that it be probabilistically independent whether judge 1 is right, judge 2 is right, etc. While it is true that the problem's circumstances – such as evidence observed by all judges or the process of group deliberation – can make it likely that the judges cast the same vote (hence all are right or all wrong), probabilistic independence is secured if by 'probability' we mean 'probability *conditional* on the problem'; indeed, conditional on the same exact body of evidence, process of group deliberation, and so on, nothing is left that could create a probabilistic dependence between the voters (who do not look on each others' ballot sheets).

What has gone wrong in the argument? To justify the competence assumption, I have appealed to a *variable* decision problem, one that is picked at random from a class of relevant problems. But to justify the independence assumption, I have *fixed* (i.e. conditionalised on) the decision problem, with its particular body of evidence, process of group deliberation, and so on. One cannot have it both ways.

More generally, the source of the disagreement on which premises of the CJT are justified is that different authors or branches of the literature more or less implicitly rely on different notions of uncertainty, ranging from the *objective* uncertainty of a random process to the *subjective* uncertainty of a social planner or, in game-theoretic models, of the voters themselves; and ranging from uncertainty about votes *given* a specific decision problem to uncertainty about both votes *and* the decision problem. I believe that most arguments made in the literature for or against some premise are correct under the author's notion

---

[3]Technically, the two versions differ in whether voters' competence and independence are required to hold *unconditionally* or *conditional* on what alternative is correct.

[4]For instance, if in the court room the innocent defendant cries out "I am guilty" to protect the true murderer.

of uncertainty, and incorrect under other notions. For instance, group deliberation prior to voting is often viewed as undermining independence (Rawls 1971, Grofman, Owen, and Feld 1983, Ladha 1992, 1995, Dietrich and List 2004), or as *not* undermining independence provided voters are isolated once it comes to voting (Waldron in Estlund et al. 1989, Estlund 2007). The latter is correct if the decision problem (including specific circumstances) is fixed, the former if the problem is variable. Also, shared information (Lindley 1985, Dietrich and List 2004), opinion leaders, or other influences (Nitzan and Paroush 1984, 1985, Owen 1986, Boland 1989, Boland, Proschan, and Tong 1989, Estlund 1994) are often taken to induce correlations, which is again correct in the variable-problem setting. I shall consider group deliberation, common information, and opinion leaders as examples of what I more generally call the *circumstances*, defined as the collection of common causes/influences of votes, among which might also be the room temperature and singing birds.

In some important aspects, the analysis I offer resembles or generalises existing arguments,[5] and it shows their underlying notions of uncertainty/probability. My main goal is to clarify a discussion that seems to suffer from some confusions and mutual misunderstandings.

The large literature on the CJT contains a number of technical refinements; besides the papers just cited, see, for instance, Young (1988), Berend and Paroush (1998), List and Goodin (2001), and Bovens and Rabinowicz (2006). A recent game-theoretic literature investigates whether sincere or informative voting is a rational strategy (Austen-Smith and Banks 1996, Feddersen and Pesendorfer 1997, Conghlan 2000, Koriyama and Szentes 1995); I have some comment on this approach in Section 6.

2   Voters, decision problems, circumstances

Throughout we consider a group of individuals (judges, citizens, experts, etc.), labelled $i = 1, 2, ..., n$, where n (≥ 2) is the group size. To be able to address the CJT, we allow the group size to vary. We think of the group of size n as containing the first n individuals of an infinite sequence of potential voters $i = 1, 2, 3, ...$ Although this paper discusses the CJT in its asymptotic version, the critique of competence and independence assumptions applies similarly to non-asymptotic CJT's.[6]

The group decides by majority voting between two alternatives, labelled 0 and 1, such as 'acquit' or 'convict' the defendant. By assumption, one of the alternatives is factually *correct* and the other one *incorrect*. Here, 'correctness'

---

[5]For instance, Ladha (1993, sect. 2) seems to make similar points to mine. Also, Dietrich and List (2004) make similar points; for instance, about misleading evidence.

[6]An non-asymptotic CJT states that, under certain conditions, a group is more likely to get it right (in majority) than a smaller group or a single individual. An asymptotic CJT, by contrast, is concerned with the limiting correctness probability as $n \to \infty$.

can mean different things, but importantly, it constitutes an (unknown) objective fact; for instance, in choosing between 'acquit' and 'convict', the correct alternative is given by whether or not the defendant has in fact committed the crime.[7] Each voter votes for the alternative he believes to be correct.[8]

The models discussed below differ in whether the group's decision problem is fixed. But what counts as part of the 'decision problem'? I define a **decision problem** as the task of *finding a certain correct alternative x (0 or 1) under certain circumstances c*. Thus a decision problem is characterised by two components:

- The **correct alternative** or **state** x. It is either 0 or 1.
- The **circumstances** c individuals face. Some are of an evidential kind, others of a non-evidential kind.
  - **Evidential** circumstances are generally observable facts that support the correctness of alternative 0 or 1, including the specific nature of alternatives 0 and 1 (Is it 'acquit Mr. Smith' vs. 'convict him to 7 years prison'? Or 'acquit him' vs. 'convict him to 3 years prison'?), and several observable events such as, again in a court case, finger-prints, a witness report, the defendant's facial expression during the trial, relevant statistical data, the process of group deliberation, etc.
  - **Non-evidential** circumstances are events that carry no information on which alternative is correct but may affect different voters in their voting behaviour, such as room temperature while voting, whether birds are singing (which might induce optimistic belief in the defendant's innocence). One might regard non-evidential circumstances as factors that affect whether voters observe evidential circumstances and how they interpret them.[9]

---

[7]In general, the assumption that one alternative is *objectively correct* is natural in (at least) two cases. First, the decision problem might be to say 'yes' or 'no' to some factual proposition (hypothesis) H, such as "CO_2 emissions cause climate change"; the objectively correct answer is then simply given by whether H is factually true or false. Second, the group might choose between two *actions* (e.g., two day trips), where all the individuals share the same preferences (e.g., to make the cheaper trip, with possible disagreements on which trip is cheaper); the objective correctness of an alternative then comes from the shared preferences. Despite the obvious difference between the two – the goal is now to satisfy (shared) desires, not to form true beliefs – one might recast the action-choice problem as a belief-formation problem, namely as the problem of knowing whether the first action satisfies the individuals' preferences more than the second action.

[8]The question of whether such sincere voting is strategically optimal is discussed in Section 6.

[9]Non-evidential circumstances can affect a voter's beliefs (subjective probabilities) *either* through enabling him to observe some evidence on which he then conditionalises (e.g., singing birds make voters see and conditionalise on the innocent smile of the defendant), *or* in a non-Bayesian way, i.e. without voters observing evidence (e.g., singing birds might cause voters to simply raise their prior probability of innocence). The latter might be thought of as a change of prior rather than a move to a conditional probability; it might be called a dynamic inconsistency of beliefs.

4

A subtle question is that of what exactly should (not) be called a part of the circumstances (and hence of the description of the decision problem). The more is included, the less randomness is left in voting behaviour *conditional on* the decision problem. Voter-specific information (such as whether voter 3 had good sleep, or whether he saw the defendant's smile) is *not* part of the circumstances; otherwise we would risk eliminating any randomness in voting behaviour conditional on the problem. In Section 3.3 I suggest conceptualising the circumstances as the *common causes/factors* of votes.

3 The fixed-problem CJT: objective uncertainty about voters given a specific decision problem

Some authors – perhaps only a minority, and perhaps mainly when arguing in favour of the independence condition[10] – think of the decision problem as being fixed. This notion of uncertainty is not only needed to defend independence, but it also is the best way of making sense of a popular version of the CJT, to be called the *fixed-problem CJT*.

3.1 The fixed-problem CJT

I now formally state the fixed-problem model and CJT. Votes are represented by random variables $V_1, V_2, \ldots$ that take values in the set $\{0, 1\}$, where V_i takes the value 0 (1) if individual i votes/judges that alternative 0 (1) is correct.[11] In this section, the probability function[12] is denoted Pr and represents objective uncertainty *given* some fixed problem, as described by a fixed correct alternative x (0 or 1) and fixed circumstances (see Section 2). One may interpret Pr as arising from an underlying probability function P (studied in Section 4) by

---

[10]For instance, the popular comparison of the votes of voters 1, 2, ... with the outcomes of independent coin tosses assumes a fixed decision problem, because the shape of the coin is fixed rather than random.

[11]I assume throughout that an individual's vote V_i does not depend on the group size n, thereby neglecting that circumstances (in particular, group deliberation) may be group-size-dependent. This idealisation is not essential for the arguments for/against the two premises (to avoid it, one would need to make i's vote group-size-dependent, i.e. use random variables V_i^n representing i's vote in a group of size n). An interpretational subtlety is that individual i's vote V_i more precisely represents how i *would* vote *if* i is among the voters, i.e. if $i \leq n$.

[12]Formally, Pr is the probability function of an underlying probability space on which all random variables $V_1, V_2, \ldots$ are defined. Like in the frameworks of later sections, these technicalities are left implicit and need not bother the reader. (Formally, a probability space consists of a set of worlds Ω on which all random variables are defined and a σ-algebra of events $E \subseteq \Omega$ on which a probability function is defined. I use standard notation; e.g. $Pr(V_i = x)$ more precisely stands for $Pr(\{\omega \in \Omega : V_i(\omega) = x\})$.)

conditionalising on the particular problem.[13]

The first premise of the CJT requires votes to be probabilistically independent from each other:

Independence (Ind). The votes $V_1, V_2, ...$ of individuals 1, 2, ... are independent.

Now to the second premise. An individual i's *competence (on the given problem)* is defined as the probability $p_i := Pr(V_i = x)$ that he votes for x, the correct alternative. I stress that p_i represents i's competence not within a general class of problems (e.g., all criminal court problems) but on the specific problem at hand; more on this in Section 3.2. In its strongest version, the competence assumption states as follows:

Competence on the problem (Com). Competence on the problem $p_i = Pr(V_i = x)$ exceeds $1/2$ and is the same across individuals i.

The unrealistic requirement of equally competent individuals is more demanding than necessary for the CJT; the following weaker requirement still suffices:

Competence-on-average on the problem ($\overline{\text{Com}}$). Average competence on the problem $\bar{p} := \lim_{n\to\infty}(p_1 + ... + p_n)/n$ (exists[14] and) exceeds $1/2$.

I now state the classic CJT in one of its versions, which I interpret as the 'fixed-problem' version. I use the weaker competence assumption ($\overline{\text{Com}}$), but of course the result stays true for the (more classical) assumption (Com).

The fixed-problem CJT.[15] If (Ind) and ($\overline{\text{Com}}$) hold, the probability of a correct majority outcome, $Pr(\#\{i \leq n : V_i = x\} > n/2)$, tends to one as the group size n tends to infinity.

As argued in the next two subsections, this theorem's independence premise can be defended, but its competence premise cannot be *known to hold*: knowing whether ($\overline{\text{Com}}$) holds *for this specific problem* might be even harder than

---

[13]P captures objective uncertainty also about the problem, and $Pr = P(.|PROBLEM = problem)$, where the (highly multi-dimensional) random variable $PROBLEM$ represents the randomly generated problem and *problem* is the particular problem considered here. Following our conceptualisation of problems as state-circumstances pairs, we may view $PROBLEM$ as a pair (X, C) of a random state variable X and the random circumstances variable C; and so, $Pr = P(.|X = x, C = c)$.

[14]That is, the (finite) group's average competence, $(p_1 + ... + p_n)/n$, converges as $n \to \infty$ (rather than, say, oscillating), a very natural assumption in practice.

[15]See footnote 23 for a re-interpretation of uncertainty (probability Pr) that makes this theorem a variable-problem CJT.

knowing the true state x in the first stage. To know whether $(\overline{\text{Com}})$ holds, one would have to know whether the specific problem involves misleading evidence, which one can hardly know without knowing the true state x.

First, though, an important remark. In this section's fixed-problem model it would not make sense to distinguish between 'competence given that 0 is correct' and 'competence given that 1 is correct'. To see why, recall that the state x is fixed (though unknown to an observer) and that probability represents objective chance (rather than an observer's subjective belief). So, by conditionalising on (say) alternative 0 being correct one conditionalises *either* on a zero-probability event (if x is 1) *or* on a sure event (if x is 0); in the former case the conditional probability is undefined, in the latter it coincides with the unconditional probability, i.e., with competence $p_i = Pr(V_i = x)$ simpliciter. Using a pair of conditional competence parameters will become meaningful in the variable-problem model (see Section 4) or for subjective rather than objective uncertainty (see Sections 5 and 6), though in the latter case voter i's (conditional or unconditional) correctness probabilities represent not i's competence but an observer's beliefs about whether i votes correctly.

For completely analogous reasons, condition (Ind) requires unconditional independence of the votes, not independence conditional on the state. Again, this will change once we introduce uncertainty about the problem.

3.2   Competence: not *known* to hold

The problem is not that competence usually fails, but that one does not know when it holds. Let me explain. Whether $(\overline{\text{Com}})$ holds depends on whether the problem's circumstances make it easy to find out the truth x. Average competence $\bar{p}$ is likely to be below $1/2$ if the problem, more precisely its circumstances, are *misleading*, that is, if either

- *evidential* circumstances are misleading, for instance if an innocent defendant pretends to have committed the crime or nervously breaks out in tears; or
- *non-evidential* circumstances have a fatal effect on the voters' abilities, for instance if the optimistic singing of birds causes the judges to believe in the innocence of a guilty defendant (see footnote 9 on the effects of non-evidential circumstances).

On the other hand, most problems do not have misleading circumstances, and average competence $\bar{p}$ on the problem exceeds $1/2$. In general, one might interpret $\bar{p}$ as a measure for *easiness* of the problem, and $1 - \bar{p}$ as a measure of *difficulty* or *misleadingness* (since $1 - \bar{p}$ is the average probability of voting incorrectly).

Importantly, though, an observer – the potential applier of the CJT, interested in whether the majority outcome is correct – can usually not *know* how easy or misleading the problem is, hence whether the voters are competent, indi-

vidually or on average. Assessing whether $(\overline{\mathrm{Com}})$ holds for the specific problem might even be harder than assessing the correct alternative x in the first place.

The simple reason is that easiness and misleadingness are defined relative to the (unknown) state x, i.e., relative to what alternative is correct. Circumstances are misleading if they 'suggest' the opposite of the truth x. What the observer can often see is that circumstances strongly 'suggest' some state, say 'suggest' alternative 1's correctness, in which case the observer can guess that $Pr(V_i = 1)$ is close 1, at least on average over the group; but this only tells him that average competence $\bar{p}$ is either close 1 (if $x = 1$) or close 0 (if $x = 0$). To know which of these two cases applies, the observer would need to know the state x, which he knows no more than the voters themselves.

3.3 Independence: holds provided we have conditionalised on all common causes

Unlike the competence assumption, independence (Ind) is arguably a safe assumption *provided* that the circumstances of the problem (on which probability is conditional) cover sufficiently many facts. Why this proviso? And what facts exactly must be conditionalised upon? Suppose for instance that room temperature is *not* taken to be part of the (fixed) circumstances. Suppose further that high temperature reduces judgmental ability. Then votes can be positively correlated, because each of them is positively correlated with the event of low room temperature. In short, given that person 1 votes correctly, room temperature is probably low, so that person 2 probably votes correctly too.

This reasoning would not go through if the common cause 'room temperature' were fixed, because then voter 1's vote would not have provided new information on room temperature, hence not on voter 2's vote.

As the example suggests, independence is a reasonable assumption provided that all common causes/factors of votes are held fixed. Why must common (not private) causes be fixed? In general, we can think of a voter's vote V_i as being fully determined by the combination of a (large) set of causes, which can be subdivided into private and common causes:

- *Private causes/factors* are facts that can affect only i's vote V_i, none of the other votes, such as: evidence that only i can observe[16], whether i indeed observes it, whether i had good sleep last night, whether i was listening properly while the witness was reporting, and so on.[17]

---

[16]Often there is none, so that all private causes are non-evidential.

[17]Private causes should not be confused with private information: a voter's lack of sleep may indeed be *observed* by others; what makes it a *private* cause is that votes of others are not affected by it. Moreover, note that whether a given (private or common) cause of i's vote makes a difference may depend on other causes. For instance, suppose V_i has just two causes: (1) what the evidence consists in, and (2) whether i observes the evidence. Then the first cause makes no difference if the second cause takes the value 'not observed'.

- *Common causes/factors* are facts that can affect more than one voter's vote, such as publicly observable evidence, room temperature, and even the entire process of group deliberation prior to voting.

My suggestion is to identify the problem's circumstances with the common causes (a possibly rich set of facts).[18] Then, since the problem (including its circumstances) are fixed, independence seems secured. This is vindicated by Reichenbach's (1956) famous *Common Cause Principle* and more recently the theory of Bayesian networks (e.g. Pearl 2000).[19]

But individual i's private causes/factors are *random* variables (except from background causes[20]), and this is precisely what makes i's vote V_i random. In fact, i's vote could be viewed as a function $V_i = f_i(C_i)$, where C_i is the vector of i's private causes, hence a vector of random variables. Nevertheless, perhaps not much objective uncertainty is left: V_i might be 1 with high probability, or 0 with high probability (meaning very high or very low competence on the specific problem). This is so in particular in the (plausible) case that all causes to V_i with evidential content (such as the witness report) are *common* rather than private. In this case, all randomness in V_i comes from evidentially irrelevant private factors (such how well voter i has slept last night) which might play no big role in determining the vote. If, moreover, the voter is a Bayesian rational, then his beliefs (hence his voting-behaviour) are entirely immune to non-evidential facts (hunger or lack of sleep); his beliefs could change only through (conditionalising on) evidence. For such a voter, V_i is deterministic, i.e., takes some value (0 or 1) with objective certainty (again, given the fixed circumstances). The possibility of having 'conditionalised away' most randomness in voting behaviour (and hence in the majority outcome) potentially makes the fixed-problem model less attractive; yet fixing the problem is what guarantees us independence.

---

[18]This would render the notion of 'circumstances' clear-cut, and hence also the notion of a 'problem' (i.e. a state-circumstances pair).

[19]According to the Common Cause Principle, variables in the world that do not causally affect each other (such as the votes $V_1, V_2, ...$, provided the voters do not look on each other's ballot sheets) are probabilistically independent conditional on their common causes. For instance, two medical symptoms of a patient, coughing and feeling tired, might be positively correlated, but conditional on the patient having a flu they are independent (assuming that flu is the only common cause). Intuitively, the common causes *screen off* the variables from each other, in a sense that can be made precise by the technical notion of *d-separation* in a causal Bayesian network; using the latter, one can also *prove* the Common Cause Principle from a more basic property, the *Parental Markov Condition* (whereby any variable in the causal network is probabilistically independent from its non-descendants *given* its parents; see Pearl 2000). Incidentally, causation can be indirect: the flu might cause tiredness indirectly by first causing bad sleep, which then causes tiredness.

[20]Private causes also include background facts such as the voter's genes, school education, and so on. These are 'background' facts insofar as (an interesting notion of) objective uncertainty, as represented by Pr, treats them as fixed. Indeed, it would be weird to imagine a random experiment that randomly re-selects the voter's school education while fixing common causes such as the temperature of the court room. The fixed private background causes, of course, affect the voter's competence.

4 The variable-problem CJT: objective uncertainty about voters and the decision problem

Perhaps most (not all) authors' thinking about the CJT is better represented by envisaging a broader random process than that examined in Section 3: one that generates not just people's voting behaviour when faced with a given problem, but also the problem itself. The problem might thus be viewed as randomly drawn from a reference class of relevant problems, such as all criminal court problems or all medical decision problems. One reason for introducing objective uncertainty of the decision problem might be that we wish to evaluate how majority rule performs *in general*, say for all decision problems a committee faces. Indeed, in order to justify majoritarianism as an institution or as part of the constitution of a decision-making body, one has to consider the whole class of decision problems to which majority rule will be applied. Another reason for treating the decision problem as random might be that objective uncertainty/probability then comes closer to someone's subjective uncertainty/beliefs. Indeed, an observer will not know the problem's true state or its entire circumstances (in the broad sense introduced above).

4.1 The variable-problem CJT

The variable-problem framework and CJT require one to consider random variables $X, V_1, V_2, ...$, taking on values in the set $\{0, 1\}$, where the *state* variable X represents the correct alternative and V_i represents individual i's vote. The probability function is now denoted P (not Pr) and represents objective uncertainty under a random process generating people's votes *and* their decision problem. Section 3's probability function Pr can be interpreted as a conditional probability function derived from P by conditionalising on a particular problem; see footnote 13. To make the state genuinely random, let the probability that alternative 1 is correct, $P(X = 1)$, be neither 0 nor 1.

Unlike in the fixed-problem framework, the independence assumption now requires the votes to be independent *conditional* on the state (i.e. on the correct alternative) – which is meaningful because the state is now random:

Independence (IND). The votes $V_1, V_2, ...$ of individuals 1, 2, ... are independent conditional on $X = 0$, and also independent conditional on $X = 1$.

Again, in contrast to the fixed-problem framework, conditional competence is now a meaningful concept. For any alternative x (either 0 or 1), an individual i's *competence given that alternative x is correct* is the conditional probability $p_i^x := P(V_i = x | X = x)$ of voting for x given that x is correct. Individual i's *(unconditional) competence* is the unconditional probability $p_i = P(V_i = X)$ of voting for the correct alternative; it is a combination of the two conditional

competence parameters:

$$p_i = P(X = 0)p_i^0 + P(X = 1)p_i^1.$$

The theorem's competence assumption can again be stated in a stronger way (that requires equally competent individuals) and in a weaker way (that requires competence on average). The two competence conditions state as follows.

Competence (COM). For each alternative $x \in \{0, 1\}$, conditional competence $p_i^x = P(V_i = x | X = x)$ exceeds $1/2$ and is the same across individuals i.

Competence-on-average ($\overline{\text{COM}}$) For each alternative $x \in \{0, 1\}$, average conditional competence $\bar{p}^x := \lim_{n\to\infty}(p_1^x + ... + p_n^x)/n$ (exists[21] and) exceeds $1/2$.

I now state the CJT in what I call the 'variable-problem' version. It holds for the weaker competence assumption ($\overline{\text{COM}}$), hence a fortiori for the stronger one (COM).

The variable-problem CJT.[22,23] If (IND) and ($\overline{\text{COM}}$) hold, the probability of a correct majority outcome, $P(\#\{i \leq n : V_i = X\} > n/2)$, and also for each alternative $x \in \{0, 1\}$ the conditional probability of a correct majority outcome, $P(\#\{i \leq n : V_i = x\} > n/2 | X = x)$, tend to one as the group size n tends to infinity.

Unlike in the fixed-problem CJT, this time it is the independence assumption that is problematic, as explained in the two following subsections. The criticism of the variable-problem theorem will have a slightly different (and in a certain

---

[21] That is, the group's average conditional competence $(p_1^x + ... + p_n^x)/n$ converges as $n \to \infty$, a plausible assumption.

[22] For the sake of completeness, I mention another frequently used version of the classic CJT, a *variable*-problem CJT that faces essentially the same analysis of its (problematic) independence and (unproblematic) competence assumptions as the present variable-problem CJT. The competence assumption is now that *unconditional* competence $p_i = P(V_i = X)$ be larger than $1/2$ and the same across individuals (or that average unconditional competence be larger than $1/2$). The independence assumption is that the events $\{V_1 = X\}$, $\{V_2 = X\}$, ... of correct votes by individuals 1, 2, ... be *unconditionally* independent. It follows that the probability of a correct majority outcome, $P(\#\{i \leq n : V_i = X\} > n/2)$, tends to 1 as $n \to \infty$.

[23] From a purely formal angle, one may re-interpret Section 3's fixed-problem CJT as a variable-problem CJT by changing the meaning of uncertainty. Indeed, let Section 3's probability function Pr represent the conditional probability $P(.|X = x)$ (rather than $P(X = x, C = c)$ as in footnote 13); that is, rather than fixing the full problem, we fix only the state, not the circumstances. Then the conditions (Ind), (Com), and ($\overline{\text{Com}}$) contain 'half' of the conditions (IND), (COM), and ($\overline{\text{COM}}$), respectively, and the theorem contains 'half' of the variable-problem CJT. As a result, (Ind) rather than ($\overline{\text{Com}}$) becomes the theorem's problematic condition.

respect more severe) status than that of the fixed-problem theorem: one premise is not just *unknown* to hold but even known *not* to hold.

The theorem also implies that, if (IND) and $(\overline{\mathrm{COM}})$ hold, the (Bayesian) posterior probability of an alternative x (0 or 1) being correct given that there is a majority for x converges to certainty:[24]

$$P(X = x | \#\{i \leq n : V_i = x\} > n/2) \to 1 \text{ as } n \to \infty.$$

4.2   Competence: holds usually

Informally, most individuals are competent because the problems with misleading circumstances form the minority of all problems. To make the argument for competence more precise,[25] let us think of the problem as drawn from a given set Π (reference class) of problems, e.g., a set of 'convict or acquit' problems. Only to simplify the exposition, let Π be finite and each problem have equal probability to be picked. Then a person's competence can be identified with the proportion of problems in Π on which he judges the state correctly (recall that a problem can be seen as a pair (x, c) of a state x, either 0 or 1, and circumstances c, as explained earlier[26]). Competence is obviously sensitive to the reference class Π: someone may be more competent within one reference class than within another. Needless to say, one may easily construct an artificial reference class within which a given person is arbitrarily incompetent: simply include only problems on which the person gets it wrong.

However, for a *natural*[27] reference class Π – e.g., that of all 'convict or acquit' problems but not that of all 'convict or acquit' problems with misleading circumstances – most voters' competences within Π should exceed $1/2$. The situation of incompetence on average not only appears rather extreme, but also *unstable*: as soon as a person notices that he gets it wrong more often than

---

[24]Forming posterior probabilities of the state would not have made sense in the fixed-problem model of Section 3; there, the state is certain to take a given value, hence stays certain to take this value after conditionalising on any event (such as on a majority outcome).

[25]To simplify the exposition, I phrase the argument in terms of people's *unconditional* competence rather than in terms of their conditional competence parameters p_i^0, p_i^1 to which (Com) and $(\overline{\mathrm{Com}})$ refer. Our argument is easily adapted to conditional competence (by considering not the entire class of problems Π but the subclass of those problems whose true state is 0, or of those problems whose true state is 1).

[26]Strictly speaking, a given problem (x, c) may not fully determine the person's vote, since private factors may play a role (this is why voting behaviour was not treated as deterministic in Section 3's fixed-problem model). So voter i's competence relative to reference class Π is not the proportion of problems $(x, c) \in \Pi$ on which he gets it right, but his average competence within Π, given by $\frac{1}{|\Pi|} \sum_{(x,c) \in \Pi} p_i^{(x,c)}$ where $p_i^{(x,c)}$ is i's competence on problem (x, c).

[27]By 'natural' I mean that randomly drawing a problem from this class represents a realistic or interesting kind of objective uncertainty. For instance, a realistic court is not confronted *only* with trials with misleading circumstances (and even if were so, the instability argument below would kick in).

12

right, he can regain competence by simply inverting any of his judgments, i.e., systematically voting for what appears wrong to him.

4.3 Independence: usually violated in favour of positive correlation

As argued in Section 3.3, any common cause/factor to the votes – such as non-private evidence or room temperature – can induce correlation if this cause is not held fixed. What secured us the independence assumption in the fixed-problem model was precisely that we conditionalised on the problem, hence on the circumstances which I have interpreted as containing the common causes. As circumstances are not fixed in the present model, independence is typically violated. As a drastic example, suppose we learn that alternative 1 is correct but that 99 of the first 100 voters incorrectly vote 0. Then we can deduce that most probably misleading circumstances are around – misleading evidence, for instance – which in turn tells us that the remaining voter most probably votes incorrectly too. But this violates (IND), since conditional on 1 being correct, the votes of 99 voters should tell us nothing about how someone else votes. This example also illustrates that it is *positive* correlation that is typically induced by common causes.

5 A social planner's subjective uncertainty

Objective uncertainty is a property of a random mechanism in the world. It is usually not known to human observers (otherwise statistics would not exist as a discipline). Should a social planner – say, in charge of deciding between two alternatives, or in charge of deciding whether majority rule is institutionalised or written into the constitution – *believe* in what the majority says? This, of course, depends on what he knows about the abilities of voters and the difficulty of the decision problem. Typically, he is uncertain both about voters (including perhaps about their identity) and about the decision problems. So subjective uncertainty looks more like uncertainty in a variable-problem framework than uncertainty in a fixed-problem framework.[28] In Section 4's variable-problem setting, let us now reinterpret P as an observer's subjective probability function. Modulo re-interpretations, much of the analysis of Section 4 still applies, that is: while the competence assumption ($\overline{\text{COM}}$) is usually unproblematic, independence (IND) does not apply to the observer's beliefs. So the variable-problem CJT does not apply to the observer's beliefs; he need not be close-to-certain that a large electorate gets it right.

---

[28] The case that the observer is certain about the problem is not only unusual but also uninteresting: the observer then need not care about the majority outcome as he already knows the correct alternative.

The most notable reinterpretation needed is that a voter i's correctness probability $p_i = P(V_i = X)$ (and its two conditional variants p_i^0 and p_i^1) are not interpretable anymore as i's 'competence': p_i measures how strongly the observer believes that i gets it right. If the observer takes i to be a genius, p_i is close 1 even if i is objectively incapable. The conditions (COM) and $(\overline{\text{COM}})$ might be called *trust conditions* rather than competence conditions.

Ladha (1993) presents an interesting CJT in which probability is indeed best interpreted as an observer's subjective uncertainty. Ladha replaces the (problematic) independence condition by the plausible assumption that votes are *exchangeable*[29], a condition that not only allows for strong correlations but is also well-motivated if the observer knows nothing that allows him to distinguish between voters. Together with other technical conditions, Ladha shows that majority outcomes are more probably correct than individual judgments, but (except in extreme cases) do not converge to certainty as the group size increases.

6 Game-theoretic models: voters' own subjective uncertainty

Rather recently, an interesting game-theoretic literature has developed around the CJT (following Austen-Smith and Banks 1996; see also the citations in the introduction). Compared to the 'classic' CJT approach, the focus is changed in at least four ways:

- The focus is not anymore on the objective or a social planner's subjective probability that majorities (democracies) find correct decisions; rather the relevant notion of uncertainty is uncertainty of the voters themselves, seen as players involved in a strategic game created by the voting situation. Accordingly, a voter's probability of voting correctly does not anymore measure his competence but the belief of the other players about whether he votes correctly.
- Sincere voting is not anymore taken for granted, a clear progress. In many models, it turns out that sincere voting is not rational: in a Nash equilibrium, not all voters vote sincerely. I should, however, point out that this relies on assuming that voters' preferences attach no intrinsic value to being sincere, a disputable assumption in some contexts.[30]
- Voters are perfect Bayesian rationals. Each voter forms beliefs (subjective probabilities) about the state by performing Bayesian updating on private information. So non-evidential belief changes, e.g., through circumstances

---

[29]That is, the joint distribution of the random variables $V_1, V_2, \ldots$ stays the same if two variables V_i and V_j $(i \neq j)$ are exchanged, i.e. V_i is replaced by V_j and V_j by V_i.

[30]For instance, the judges in a legal court might attach a higher value to voting their sincere opinion (say, following a professional vow) than to a correct verdict as the voting outcome.

such as room temperature (see footnote 9) or through private causes such as bad sleep, are excluded.[31]

- Voting is construed purely as a process of information pooling. Disagreements between voters are seen as coming from distinct private information.[32] This excludes the case that disagreements are due to different interpretations of the same information;[33] disagreements of the latter kind may persist even after people deliberate and the disagreement becomes common knowledge between the voters. The standard game-theoretic models exclude that voters agree to disagree prior to voting; if they do, sincere voting is rational again.

What these points indicate is that the game-theoretic approach has brought considerable new insights (by looking for 'explaining' voting behaviour), but has done so under rather restrictive assumptions which the original CJT enquiry was not making.

To come back to the central topic of this paper, what role (if any) do assumptions of independence and competence play in game-theoretic models? First of all, since voting behaviour is not an input to the models but an output, the two assumptions must be reformulated in terms of assumptions on private information/signals. So the question is whether private signals are independent and whether they are likely to indicate the truth[34]. Since probability reflects subjective uncertainty (of players), arguments similar to those given in Section 5 tell us that we should be cautious mainly about independence assumptions. Austen-Smith and Bank's (1996) independence assumption is part of what drives their striking finding that sincere voting is (usually) not rational.

---

[31]But they could perhaps be modelled using Bayesian games *with different prior beliefs*. To see why different priors are needed to capture non-evidential belief formation within players, consider this example. Suppose each player's beliefs about whether the defendant is guilty are formed not based on evidence but *solely* on how well the player has slept. In a game-theoretic model, a player's type is then given by how well he has slept, and the beliefs held by type θ of a player are represented by the player's conditional beliefs given that his type (sleep) is θ. So, in a player i's beliefs prior to becoming a type (prior to sleeping), his type (sleep) is correlated with whether the defendant is guilty (which is only a metaphor: the player never held such absurd beliefs). But in other players' beliefs, no correlation should exist between i's type (sleep) and whether the defendant is guilty. So the model needs to assign different players different prior beliefs.

[32]Some model refinements also include differences in preferences, or costs of acquiring information.

[33]By 'different interpretations' I do not refer to cases with a hidden form of different information. One way to capture difference in interpretation is through different prior beliefs. See footnote 31.

[34]As noted above, the probability that a player receives a truth-indicating signal is not interpretable as the player's competence but as the other players' uncertainty about this player's information; accordingly, the term 'competence' is rarely used in the game-theoretic literature.

7   Concluding remarks

In the literature, many jury theorems are derived that, though mathematically interesting, suffer from problematic premises. Whether a premise is justified depends on the notion of uncertainty; possibly the most interesting notions include uncertainty about the decision problem (with its specific circumstances), be it objective uncertainty (as in Section 4) or a social planner's subjective uncertainty (as in Section 5). For such uncertainty, independence assumptions become problematic.

The doubts that the game-theoretic approach has cast on the hypothesis of sincere voting should be taken seriously, though I have also indicated that the game-theoretic approach may itself have to be modified, possibly rehabilitating the rationality of sincere voting.

Future research should concentrate on jury theorems with justifiable premises. A good indicator for whether premises are justified is whether the conclusion is prima facie plausible. In my view, the conclusion of *non-asymptotic* jury theorems (namely that groups are more competent than smaller groups or single individuals) is plausible. And the conclusion of *asymptotic* jury theorems (namely that majority correctness tends to one as the group size increases) is usually only plausible if the model allows one to re-interpret a 'correct' decision as one that is 'justified' based on all available information (spread over the voters); under this re-interpretation, asymptotic jury theorems explore conditions for majority voting to successfully aggregate information in the limit.[35]

A   Proof of the two theorems

Though well-known, let me give simple proofs of the two theorems above.

Proof of the fixed-problem CJT. Assume (Ind) and $(\overline{\text{Com}})$. Suppose $x = 1$ (the proof is analogous if $x = 0$). As one easily checks, the random variables $V_i - p_i$, $i = 1, 2, ...$, have zero expectation. As they are also independent, the law of large numbers applies: $Pr\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(V_i - p_i) \to 0\right) = 1$. Using that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}p_i \to \bar{p}$, it follows that $Pr\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}V_i \to \bar{p}\right) = 1$. So (since convergence with probability one implies stochastic convergence), for each $\epsilon > 0$ we have $Pr\left(\left|\bar{p} - \frac{1}{n}\sum_{i=1}^{n}V_i\right| < \epsilon\right) \to 1$. Choosing ϵ sufficiently small (namely

---

[35]For instance, 'acquit' is the justified decision (the 'full-information' decision) whenever evidence for guilt is insufficient, *even if* the defendant is truly guilty. In particular, Miller (1986), Ladha (1995), and part of the game-theoretic literature seem to interpret correctness along these lines.

$\epsilon = \bar{p} - 1/2$), it follows that $Pr\left(\frac{1}{n}\sum_{i=1}^{n}V_i > 1/2\right) \to 1$, i.e. that

$$Pr\left(\sum_{i=1}^{n}V_i > n/2\right) = Pr\left(\#\{i \leq n : V_i = 1\} > n/2\right) \to 1. \quad \blacksquare$$

Proof of the variable-problem CJT. The two premises (IND) and $(\overline{\text{COM}})$ guarantee that, for each state x in $\{0,1\}$, the conditional probability function $P(.|X = x)$ satisfies the premises (Ind) and $(\overline{\text{Com}})$ of the fixed-problem CJT. So, by the latter, $P\left(\#\{i \leq n : V_i = x\} > n/2|X = x\right) \to 1$. So, letting A_n be the event that $\#\{i \leq n : V_i = x\} > n/2$, we have

$$P(A_n) = \sum_{x=0}^{1}P(A_n|X = x)P(X = x) \to \sum_{x=0}^{1}1 \cdot P(X = x) = 1. \quad \blacksquare$$

Acknowledgment

References

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

Ben-Yashar, R. and J. Paroush. 2000. A Nonasymptotic Condorcet Jury Theorem. *Social Choice and Welfare* 17: 189-99.

Berend, D. and J. Paroush. 1998. "When Is Condorcet's Jury Theorem Valid?" *Social Choice and Welfare* 15: 481-8.

Boland, P. J. 1989. "Majority Systems and the Condorcet Jury Theorem." *Statistician* 38: 181-9.

Boland, P. J., F. Proschan, and Y. L. Tong. 1989. "Modelling Dependence in Simple and Indirect Majority Systems." *Journal of Applied Probability* 26: 81-8.

Bovens, L. and W. Rabinowicz. 2006. "Democratic Answers to Complex Questions: An Epistemic Perspective." *Synthese* 150(1): 131-53.

Condorcet, M. J. A. C. 1785. "Essai sur l'application de l'analyse à la probabilité rendues à la pluralité des voix." *Les Archives de la Révolution Française*. Oxford: Pergamon Press.

Conghlan, P. 2000. "In Defense of Unanimous Jury Verdicts: Mistrials, Communication, and Strategic Voting." *American Political Science Review* 94(2): 375-93.

Dietrich, F. and C. List. 2004. "A Model of Jury Decisions Where All Judges Have the Same Evidence." *Synthese* 142: 175-202.

Estlund, D. 1994. "Opinion Leaders, Independence, and Condorcet's Jury Theorem." *Theory and Decision* 36(2): 131-62.

Estlund, D. 2008. *Democratic Authority: A Philosophical Framework.* Princeton: Princeton University Press.

Estlund, D., J. Waldron, B. Grofman, and S. Feld. 1989. "Democratic Theory and the Public Interest: Condorcet and Rousseau Revisited." *American Political Science Review* 83(4): 1317-40.

Feddersen, T. and W. Pesendorfer. 1997. "Voting Behavior and Information Aggregation in Elections with Private Information." *Econometrica* 65(5): 1029-58.

Grofman, B., G. Owen, and S. L. Feld. 1983. "Thirteen Theorems in Search of the Truth." *Theory and Decision* 15: 261-78.

Koriyama, Y. and B. Szentes. 1995. "A Resurrection of the Condorcet Jury Theorem." *Working paper.* University of Chicago

Ladha, K. K. 1992. "The Condorcet Jury Theorem, Free Speech, and Correlated Votes." *American Journal of Political Science* 36: 617-34.

Ladha, K. K. 1993. "Condorcet's Jury Theorem in Light of De Finetti's Theorem." *Social Choice and Welfare* 10: 69-85.

Ladha, K. K. 1995. "Information Pooling Through Majority-rule Voting: Condorcet's Jury Theorem with Correlated Votes." *Journal of Economic Behavior and Organization* 26: 353-72.

Lindley, D. 1985. "Reconciliation of Discrete Probability Distributions." In J. M. Bernando et al. (eds.), *Bayesian Statistics*, vol. 2, pp. 375-90. Amsterdam: North-Holland.

List, C. Forthcoming. "The Epistemology of Special Majority Voting." *Social Choice and Welfare.*

List, C. and R. E. Goodin. 2001. "Epistemic Democracy: Generalizing the Condorcet Jury Theorem." *Journal of Political Philosophy* 9: 277-306.

List, C. and P. Pettit. 2004. "An Epistemic Free Riding Problem?" In P. Catton and G. Macdonald (eds.), *Karl Popper: Critical Appraisals*, pp. 128-58. London: Routledge.

Miller, N. R. 1986. "Information, Elections, and Democracy: Some Extensions and Interpretations of the Condorcet Jury Theorem." In B. Grofman and G. Owen (eds.), *Information Pooling and Group Decision Making.* Greenwich, CT: JAI Press.

Nitzan, S. and J. Paroush. 1984. "The Significance of Independent

Decisions in Uncertain Dichotomous Choice Situations." *Theory and Decision* 17: 47-60.

Nitzan, S. and J. Paroush. 1985. *Collective Decision Making.* Cambridge: Cambridge University Press.

Owen, G. 1986. "'Fair' Indirect Majority Rules." In B. Grofman and G. Owen (eds.), *Information Pooling and Group Decision Making.* Greenwich, CT: JAI Press.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Rawls, J. 1971. *A Theory of Justice.* Cambridge MA: Harvard University Press.

Reichenbach, H. 1956. *The Direction of Time.* Berkeley, CA: University of California Press.

Young, H. P. 1988. "Condorcet's Theory of Voting." *American Political Science Review* 82: 1231-44.

Academic biography

Franz Dietrich is an Assistant Professor of Economics at Maastricht University, and a Nuffield New Career Fellow at the London School of Economics. His research focusses on decision theory, social choice theory, and analytic philosophy.