Episteme (2026), 1–20
doi:10.1017/epi.2026.10133

episteme

ARTICLE

A Constructive Condorcet Jury Theorem

Eleonora Cresto

Departamento de Humanidades, Universidad de San Andrés, Argentina
Email: ecresto@udesa.edu.ar

(Received 4 November 2025 UTC; revised 10 April 2026 UTC; accepted 4 July 2026 UTC)

Abstract

There is a well-established research program that relies on the Condorcet Jury Theorem (CJT) to justify democracy and majority voting from an epistemic perspective. This paper explores how the CJT can be extended beyond its standard epistemic interpretation. I propose a version of the CJT that does not presuppose the existence of an objective fact of the matter (or moral fact) on which group members must vote; I call it the "Constructive CJT." This constructive version of the CJT is designed to show that a close relative of majority voting (here modeled as majority polling followed by individual action) can help bring about outcomes that are desirable independently of the procedure itself. In particular, it helps agents build situations in which cooperative behavior is achieved, as long as certain conditions are met. This provides a middle ground between epistemic and procedural justifications; under certain conditions, majority voting contributes to generating the environment under which desirable outcomes arise.

Keywords: Condorcet Jury Theorem; epistemic democracy; majority polling; coordination; Stag Hunt; collective action; epistemic constructivism

1. Introduction

There is a well-established research program that uses the Condorcet Jury Theorem (CJT) to defend democracy and majority rule on epistemic grounds. Under mild assumptions of competence and independence, large electorates almost surely reach correct decisions. This classical picture presupposes an exogenous truth; it presupposes that votes track a fact whose correctness is prior to, and independent of, the collective procedure that aggregates them.

The present paper develops a Constructive Condorcet Jury Theorem (CCJT), a formal result that preserves the epistemic architecture of Condorcet's model while releasing it from truth realism. The key idea is that aggregation can itself be a truth-generating mechanism. Instead of assuming an external fact F to which votes must correspond, the CCJT shows how a majority poll can foster convergence toward an endogenously produced, normatively desirable state of affairs that emerges through the process of polling and subsequent coordination.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

In a nutshell, the argument proceeds as follows. First, I present a generalization of the Stag Hunt game to an n-person setting, to model situations where collective action depends on majority behavior. I then formalize this idea within a two-tier aggregation model of polling and acting (the CCJT). If individuals first vote on what they would do, with votes satisfying competence and independence assumptions, then, under sincerity, majority polling converges in probability to a twice-idealized majority outcome (what I call the "proto-truth") that either drives collective cooperation or, at the very least, improves payoffs. Under mild incentive conditions, polling outcomes become reliable public signals that coordinate expectations and foster cooperation. The resulting framework identifies a middle ground between epistemic and procedural accounts of collective rationality, thus yielding a form of epistemic constructivism: voting does not *reflect* a truth but participates in its construction. If successful, the proposal may have consequences for institutional design, as it will suggest that certain voting or aggregation procedures may be epistemically valuable, precisely because they help bring about antecedently desirable results.

The argument is developed within the CJT framework rather than through some other formal apparatus, because the CJT is the reference point within the research program that uses formal aggregation results to evaluate the epistemic credentials of majority voting. Showing that a constructive version holds therefore contributes directly to that program on its own terms.

In Section 2 I present the standard CJT, discuss its role in epistemic justifications of democracy, and point to the limitations of epistemicism. Section 3 generalizes the Stag Hunt game and introduces the idea that players can vote before acting; voting in this setting, however, is non-binding (agents can vote one way and act differently later). In Section 4, I present the basic structure of the two-tier aggregation model of voting and acting and explain how to understand the classical properties of independence and competence in the present constructivist setting. In Section 5 I show that results analogous to the classic CJT hold here as well. Section 6 clarifies the meaning of various crucial concepts, while Section 7 explains why the resulting fact (obtained by collective action) is antecedently desirable. Section 8 situates the CCJT in relation to existing approaches in the literature, and Section 9 concludes.

2. The Condorcet Jury Theorem

The CJT tells us that a group deciding on a "yes" or "no" issue by majority voting will be more reliable than any of its members and will be virtually infallible when the number of members tends to infinity, provided a couple of conditions on individual reliability and independence are in place.[1]

More formally, assume a set of n voters; for simplicity, suppose n is odd. Suppose each voter i casts a "yes" or "no" vote on some fact F. Let V_i be a random variable with values $\{0, 1\}$, according to whether the vote is (1) or is not (0) correct. Correctness depends on the value of another random variable X; suppose X has again values $\{0, 1\}$.

[1]The original formulation is found in Condorcet (1785).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

Assume that the following two conditions hold:

Competence condition (individual reliability): For every agent *i*:

$$P(V_i = 1|X = 1) = P(V_i = 0|X = 0) = r, \quad \text{for } 1 > r > 0.5$$

In other words, individuals are not infallible, but they get things right more often than they would if they were to decide the vote by tossing a fair coin; we assume the same *r* for every voter.

Independence condition:

The vote cast by each member of the group is independent of the votes of the others. Then, $S_n = \sum V_i$ is another random variable, with binomial distribution.

Normally, the independence condition is understood conditionally: votes are independent conditional on the true state of the world (which, ideally, acts as a common cause that explains the votes), or conditional on both the true state and all other common causes.[2]

From these premises we obtain:

The Condorcet Jury Theorem:

1. Nonasymptotic part:
   a Group reliability, as calculated by majority voting, is always larger than individual reliability: $P(S_n \geq (n+1)/2|X) > r$
   b Group reliability is strictly increasing with the number of members: for any odd n, $n^* = n + 2$, we have

$$P(S_{n^*} \geq (n*+1)/2|\mathrm{X}) > \mathrm{P}(\mathrm{S_n} \geq (\mathrm{n}+1)/2|\mathrm{X})$$

2. Asymptotic part: When n tends to infinity, the probability that the group gets things right through majority voting tends to 1.

It should be noted that group reliability grows pretty fast, so large groups are virtually infallible. Grofman et al. (1983) show that for the CJT to go through, it is sufficient to assume that *average* individual reliability is above 0.5.

We could question whether the assumptions of the theorem are ever met, or whether they are mutually consistent. Dietrich (2008), for example, contends that the competence and independence conditions are not well defined, and that once we make them precise, they are mutually incompatible. On the other hand, Austen-Smith and Banks (1996) argue that there are implicit conditions, such as sincere voting, which, once made explicit, can also cause some trouble. In any case, there are several attempts to vindicate its use for epistemic democracy (Estlund 1994; Estlund et al. 1989; Goodin and Spiekermann 2018; Landemore 2013) whereas other authors support close variants of the CJT that are sensitive to some of its problems (Dietrich and Spiekermann 2013a, 2013b; List and Goodin 2001).

Here, I want to focus on what we may call *the epistemic presupposition*: the theorem (as well as the aforementioned variants) presupposes the existence of some objective fact of the matter, or perhaps moral fact *F*, whose truth, or desirability, does not depend on the method used to aggregate different opinions on whether *F* holds or should hold.

[2]See Dietrich and Spiekermann (2013a), and, more generally, Bradley et al. (2014).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

But what if there is no such *F*? To see what is at stake, recall how a voting system may be justified (finer-grained distinctions are of course possible, but for the moment these two will do):

- In a purely proceduralist or formalist account, a voting process establishes its own correctness, so to speak: whatever the result, it is deemed correct simply because it was obtained by following a given procedure. The procedure itself may be valued for its formal properties. In particular, majority voting is monotonic and satisfies anonymity and neutrality, all of which can be regarded as desirable features.
- An epistemic account, by contrast, tells us that the result is legitimate because the procedure tracks the truth.[3]

Here is where the CJT enters the picture: from an epistemic perspective, majority voting is vindicated because in the limit, there is a perfect match with a fact. For example, a guilty verdict is correct if the defendant actually committed the crime; in the limit, a jury that satisfies the premises will always be right. But not all voting situations are like this. A Parliament may vote on whether to legalize same-sex marriage, or on the advisability of leaving the EU; a plebiscite may require people's opinions on abortion rights. At this point, an epistemicist may raise the stakes and claim that there are also axiological facts (as in: a certain economic policy may be the ideally correct policy because it instantiates the common good). Of course, we may well be skeptical about the existence of truly desirable facts in the objective structure of the world. In this case, the relevance of the theorem seems to be drastically reduced.

Goodin and Spiekermann are fully aware of this problem and devote part of their (2018) work to addressing it. Appealing to moral realism, moral conventionalism, or "moral separability" (facts and values are separable, so there is always at least some factual component on which we can vote) will likely be met with skepticism by many. So they finally turn to what they call "moral majoritarianism." Different people may hold different values within a community. Even if people's preferences and desires are not the kind of thing that can be called true or false, there can still be *true statements about what people want and prefer*. It might serve some good social purpose to *discover the truth* about what most people want or prefer; aggregating independent judgments about this might be a good way to achieve it.

Goodin and Spiekermann thus perform a curious twist regarding the possible nature of *F*: *F* may just be a fact about what people prefer. But it is far from clear that we can say this and continue in the epistemicist camp. First, note that, if the relevant truth is a statement about what most people prefer or value, then it is trivially true that majority voting will give us that; the "truth" we discover through majority voting requires no special theorem. Most importantly, such a defense of majority voting is circular, as it assumes the very point in question, namely, why the majority's preferences should have any special normative weight. Identifying the most common preference is easy;

[3]Rawls (1971) famously distinguished between perfect, imperfect, and pure procedural justice. For the purposes of this paper, however, the key division is simply between procedures that presuppose an external standard and those that do not.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

explaining why it matters requires further argument, one that will no longer be epistemic in nature. For that matter, there is also a truth about what the majority of voters consider the *worst* option, or about what the majority of any subset X of the electorate prefers. If *those* were the truths we wanted to capture (if we wanted to track pluralities of detested options, or majorities within some qualified subset), then the traditional majority rule would no longer serve us.

In short, unless we have other reasons to let the social outcome coincide with what the majority likes, simply noting that the majority likes it gets us nowhere. To show that there is genuine social value in honoring the majority's wishes (beyond the mere fact that they are the majority's), we need some further argument, and it will no longer be a purely epistemic one.

On a charitable reading, Goodin and Spiekermann aim to track a truth that is, in some sense, constructed through the voting process. In this paper, I explore a way of recovering this intuition. What follows can be viewed as an attempt to take their project seriously and to make the most charitable sense of the "moral majoritarian" view. To this end, I develop a version of the CJT that applies even when there is no pre-existing fact of the world F, but where F is instead brought about by the very mechanisms of aggregation. If we can further show that the resulting state of the world is antecedently desirable, then we will have demonstrated that the standard dichotomy between truth-tracking and purely procedural accounts of democracy is not exhaustive.

3. An example. Voting* and hunting

Consider the following generalization of the classic Stag Hunt game.[4] Suppose we have a group with n people, each of whom can either seek to hunt Hare or Stag. Suppose a Stag is actually chased if and only if a majority of hunters goes for Stag. Here is the payoff structure for three players:

Stag:

	Stag	Hare
Stag	$10, 10, 10$	$8, 6, 8$
Hare	$6, 8, 8$	$5, 5, 0$

Hare:

	Stag	Hare
Stag	$8, 8, 6$	$0, 5, 5$
Hare	$5, 0, 5$	$5, 5, 5$

Call this example GENERALIZED STAG HUNT. I am offering precise numbers for the sake of concreteness, but payoffs need not be exactly like this for my argument to work, of course. The rationale is that getting Stag is better than getting Hare, but going for Stag is also tiresome, so the more people involved in the task, the better. Although in this case

[4]For other ways to extend the 2-person Stag Hunt game see for example Vanderschraaf (2019), or Silva (2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

the trophy is shared among more participants, one can conjecture that the decrease in individual effort offsets the reduction in meat intake. Going for Stag is also risky because you can end up with empty hands. Note that, if actually chased, the Stag is not shared with those who go for Hare, so the interaction does not allow for free riders. However, we can also assume that the fewer the people who go for Hare, the easier it is – at least weakly – to get it (they have fewer competitors).[5]

For a less abstract example with the same payoff structure, consider speaking up in the face of an injustice. Call this example COURAGE. Imagine a group of people who have suffered some form of oppression or harassment (e.g., sexual harassment). If enough people in the group speak up, then those who speak up will be substantially better off. If those who raise their voices are only a minority, they will be punished by the system, while those who remain silent are unaffected; when a majority speaks up, even the silent benefit indirectly (just a little) from the raised awareness.

In what follows, I will focus on the Stag and Hare case, which is more schematic and easier to discuss.

Suppose players *vote* first on what to do. By "voting" I mean something slightly non-standard; this is to be expected, given that we want a non-standard application of the CJT. Let me call it "voting*." The meaning of "agent *i* votes* for Stag [Hare]" is: "agent *i* announces that, were she in a Stag Hunt scenario, she'd rather go for Stag/Hare." Crucially, voting* is not binding; it is a sort of "surveying" or polling of each member's intentions to behave one way or another. Intuitively, if at a later stage the players learn that there is a majority voting* for Stag, individuals may feel safer to go for Stag (assuming certain conditions are met); if not, they may not feel equally safe. In any case, *agents need not know whether they will have access to the result of the poll, and they need not know whether they will be pressed to act (go for Stag or Hare) right after the polling*. In light of this, it is logically possible for individuals to change their minds; they may vote* for Stag now and go for Hare later, or vice versa. People may vote* for one thing and later act otherwise, for various reasons, one of which (but not the only one) is learning the outcome of the vote* itself. This does not mean that they had voted* insincerely. The moment they voted*, they might have been convinced that Stag (or Hare) was the right choice.[6]

Note that, whether a majority ends up chasing Stag, or whether a majority finally speaks up, is the result of a collective action, not of the actions of one single individual or representative (unlike standard voting scenarios). Group acting in our examples amounts to the aggregation of individual actions, and when the moment comes, each agent can do whatever they choose, regardless of the result of the voting*.

Notice that there are two different things to aggregate here: voting* and hunting behavior; correspondingly, there are two majorities at stake: voting* for Stag or Hare,

[5]More precisely, the underlying thought is that competition over hares is a threshold effect rather than a strictly monotone one. For arbitrary (odd) group size n, the stipulation reads: the payoff of going for Hare is constant at 5 across all states in which Hare-hunters form a majority (so that the Stag escapes), and constant at 6 – for simplicity – across all states in which they do not. Together with the assumptions that the per-capita Stag payoff, conditional on the Stag being chased, increases with the number of participants and exceeds 6 at full participation, this guarantees that the efficiency claims of Section 7 hold exactly for groups of any size, not merely for $n = 3$.

[6]For further nuances on the distinction between voting and voting*, see Section 6.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

and going for Stag or Hare. As voting* is not binding, voting* and acting are not logically necessitated. However, they may well be correlated, if certain properties hold. As we shall see, if we assume the analogues of Competence and Independence for voting*, we obtain a convergence result analogous to that of the CJT: a majority voting* for Stag, in the limit, will "match" a majority going for Stag – even though the latter has not occurred yet. In what follows, we explore this basic intuition.

4. The basic setting

As before, let the group be composed of n members, where n is odd. Let V_i be a random variable with values $\{0, 1\}$. The intended meaning of $V_i = 1$ $[V_i = 0]$ is "agent i votes* for Stag [Hare]"; as already discussed, this should be interpreted as "agent i announces that, if she were in a Stag Hunt setting, she would prefer to go for Stag/Hare." In addition, let $V = \sum V_i$. A majority vote* for Stag amounts to the claim that $V \geqslant (n+1)/2$. To simplify notation, abbreviate this as V_{Stag}, and write $V < (n+1)/2$ as V_{Hare}.

Consider now a new random variable G_i, also with values $\{0, 1\}$, such that $G_i = 1$ $[G_i = 0]$ stands for "agent i actually goes for Stag [Hare]"; i.e., agent i actually seeks to hunt Stag [Hare]. We abbreviate $\sum G_i = G \geqslant (n+1)/2$ as G_{Stag}, and $G < (n+1)/2$ as G_{Hare}. G_{Stag} $[G_{\text{Hare}}]$ then asserts that there is [is not] a number $k \geqslant (n+1)/2$ of individuals who actually seek to hunt Stag (contrast this with: individuals who simply vote* for Stag/Hare).

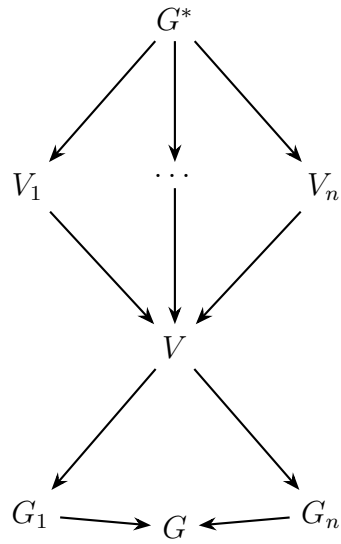
Suppose that, at the time of voting*, individuals have not acted yet; hence it is still undetermined whether G_{Stag} or G_{Hare} holds. Still, individual voters* may be more or less attuned to what the majority thinks should be done; agents may have recollections of past experiences in similar scenarios, or they may be knowledgeable of how their fellow group members are psychologically built, or they may just be good at predicting human behavior. We will return shortly to what such "attunement" amounts to, in the present context, as there are some important caveats. But first, it will be helpful to represent the causal structure of the intended model using a Bayesian network.[7]

Clearly, we should begin by positing a directed acyclic graph (DAG) that proceeds from individual votes* $V_1, \ldots, V_n$ to an aggregated opinion V. If known, V acts as a causal influence on $G_1, \ldots, G_n$; in turn, $G_1, \ldots, G_n$ jointly determine the collective action G. However, if we want to incorporate the aforementioned "attunement" into the diagram, this cannot be the whole story.

Note first that we cannot have both that the votes* are an indirect cause of the collective action G (via the DAG of the Bayesian network), and that they are all independent conditional on the future state of the world G (i.e., we cannot have that, for all i, j, $V_i \perp V_j | G$), as the d-separation criterion reveals. G is a common effect of the votes* – in graph-theoretic terms, a descendant of the collider V on the path $V_i \rightarrow V \leftarrow V_j$ – and conditioning on a collider's descendant, like conditioning on the collider itself, unblocks the path, introducing dependence between the V_i's. In the light of this, in what follows, I will distinguish between a "proto-truth," or "normative truth" G^*, where $G^* \in \{G_{\text{Stag}}, G_{\text{Hare}}\}$, and the collective action finally performed by the group, or G, such that the causal structure now looks like this:

[7]See Bradley et al. (2014); the locus classicus is of course Pearl (2000).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133



The intended interpretation is that the top layer captures an epistemic stage (agents form opinions based on an underlying proto-truth G^*), whereas the bottom layer captures a behavioral stage (agents observe the majority V of the initial poll and respond with behavior G_i). This structure makes it possible to have a two-tier aggregation theorem, to which I will come back soon.

What is G^* exactly? By construction, G^* is a target outcome that a collective decision-making procedure might aim to track. But G^* cannot just be the actual majority vote, as in this case any statement of the form "the majority should aim for G^*" becomes tautological. We will stipulate that G^* denotes a *twice-idealized majority*, namely, the limit of the majority outcome when (i) the population size n tends to infinity and (ii) the vote–update procedure is iterated k times, when k again tends to infinity. In other words, the twice-idealized majority G^* is a fixed point for infinite votes* and infinite rounds. Under this idealized process, the second limit assumes that the outcome of each round becomes publicly available information and feeds into the next round, so the infinite sequence of votes* represents a process of collective updating. In this ideal mechanism, the (infinite) group observes each intermediate majority, adjusts its beliefs or expectations, and eventually converges toward a stable point (the proto-truth) that reflects the end state of this hypothetical learning process. Crucially, the proto-truth G^*, as an idealized construct, is independent of any actual majority outcome produced by the real, finite group voting* only once. The role of the second limit is to build publicity into the very target of the procedure: G^* is the opinion on which the group would settle were the poll repeatedly announced, a disposition that ratifies itself upon becoming public.

Note that each agent i does not vote* to *purposefully try to conform to* G^*; her vote* reflects what she thinks she would do when the time comes. However, recall that an agent finds it convenient to go for Stag if and only if she thinks a majority will go for Stag, and, presumably, this will be so if a majority votes* for Stag. So, had she formed an opinion or belief about G^*, that would obviously influence her voting*. In later sections we will return to the reasons agents may have to place their vote one way or other.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

We are now in a position to offer a Constructive CJT. We turn to this task in the next Section.

5. A constructive CJT

The central idea is as follows. If agents respond independently and with probabilistic accuracy to the majority signal V, and the first-stage majority V is itself a good estimator of the truth G^*, then G tracks G^* again, via another CJT-like mechanism, so the majority behavior G will converge in probability to the (normative) proto-truth G^*. We will then have a compositional epistemic model of two-tier aggregation, through two different CJTs. This makes V a kind of epistemic catalyst, i.e., a mediating signal that fosters reliable action even when the truth G^* is never directly observed. Let us see this more carefully.

Consider first the following definitions [*Mutatis mutandis* for "Hare"]:

- Individual Competence: $P(V_i = 1|G^*_{Stag})$.
- Group Competence: $P(V_{Stag}|G^*_{Stag})$.
- Collective Trustworthiness: $P(G_{Stag}|V_{Stag})$
- Individual Hope: $P(G_i = 1|V_{Stag})$
- Independence of votes*: $V_i \perp V_j|G^*$ (The V_is are independent conditional on the proto-truth)
- Independence of actions: $G_i \perp G_j|V$ (agents act independently after learning the results of the poll)

Our two-stage structure, as captured in the diagram of the previous section, is compatible with assuming Independence of votes*, Independence of actions, and both Individual Competence and Individual Hope larger than 0.5. In the next section, we will go back to whether such assumptions are indeed reasonable; for the moment, I rest content with pointing out that they are *possible*, given our overall causal model.

Assume also that it is common knowledge that people vote* sincerely. To vote* sincerely means that, the moment they cast their vote, people genuinely express what they intend to do now (what they would do, were they to act now), even if they change their minds later. This is a reasonable assumption if, as suggested in Section 3, we require that agents do not know whether they will ever know the result of the poll, or that they do not know whether they will need to act right away, or both. In the next Section, I will provide further elaboration on the idea of sincere voting*, and in Section 7, I will explore the consequences of relaxing this assumption altogether.

If the aforementioned assumptions hold, both $V = \sum V_i$ and $G = \sum G_i$ are random variables with a binomial distribution, and the following propositions are straightforward:

Proposition 1. Assume Individual Competence > 0.5 and Independence of votes*; assume moreover that voting* is sincere. Then, Group Competence is higher than Individual Competence, increases steadily as n grows (for odd n), and converges to 1 when n tends to infinity.

In other words, the probability of achieving majority voting* for Stag [Hare], conditional on the twice-idealized majoritarian opinion G^* becomes 1 in the limit; i.e., the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

probability that a given majority coincides with the proto-truth converges to 1 as n goes to infinity.

Note that, in this model, G^* plays the same formal and functional role as F in the classical CJT: it is the proposition toward which agents may converge. However, G^* is not part of the real world, so to speak, but a normative fact that can become true as a result of the very voting process that aims to track it. On the other hand, G is the descriptive result of a collective action.

Proposition 2. Assume Individual Hope > 0.5 and Independence of actions. Then, Collective Trustworthiness is higher than Individual Hope, increases steadily as n grows (for odd n), and converges to 1 when n tends to infinity, assuming a common value of Individual Hope across agents (for the asymptotic part, average Hope above 0.5 suffices, cf. Grofman et al. 1983).

Proposition 3. Constructive Condorcet (CCJT). Assume the hypotheses of Propositions 1 and 2, and let the prior over the proto-truth be non-dogmatic, i.e., $0 < P(G^*_{Stag}) < 1$. Then, as n tends to infinity, $P(G_{Stag}|G^*_{Stag})$ converges to 1; moreover, $P(G^*_{Stag}|G_{Stag})$ and $P(V_{Stag}|G_{Stag})$ converge to 1 as well [mutatis mutandis for Hare].

The composition of the two tiers is not automatic (probabilistic tracking is not transitive in general), but the causal structure of Section 4 supplies exactly what transitivity requires, namely, the announced result V screens off G from G^*. The converse conditionals then follow by Bayes' theorem, the non-dogmatic prior keeping the relevant denominators bounded away from zero.[8] Notice that what is being tracked primarily by the voting* situation is not G itself, but G^*; nonetheless, votes* will eventually also track G; a majority voting* for Stag [Hare] will eventually match a majority going for Stag [Hare] (i.e., a future event).

We have thus obtained a constructive version of the CJT. Notice, however, that this alone is not enough to secure the middle ground between formalism and epistemicism that we are striving for; we return to this point, and complete our task, in Section 7. First, though, a few comments and clarifications are in order, which may help address concerns about whether we are heading in the right direction.

6. Discussion and clarifications

*6.1. On voting and voting**

Unlike the standard CJT, the current analysis does not focus on standard majority voting. Let me state the differences more clearly.

A voting rule is normally a means of establishing the preferences of a group as a function of the preferences of its members, coupled with the decision to *act* on the preferences of the group. When there are only two options, an odd number of voters, and no ties are allowed for individual votes, majority voting has as input a vector with the most preferred option of each voter, and yields as the choice of the group the option with the most votes. Thus, the real difference between majority rule and majority voting*

[8]For screening-off as a sufficient condition for the transitivity of probabilistic support, see Shogenji (2003).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

as described in this paper is that for the latter we explicitly sever the link between group preference and group acting. And this is because, by construction, the examples discussed are such that:

> (i) no one can really act on behalf of the group (we are dealing with cases of collective action), and
> (ii) voting* does not bind individuals to act one way or another, as there is no way to enforce behavior.

In other words, a voting* rule, as we are using the concept here, is a rule for opinion aggregation. Rules for opinion pooling are more general than standard voting rules: every voting rule is at the same time a pooling rule, but not vice versa.

On the other hand, note that a poll, by itself, does not instantiate majority voting*, nor any other voting* rule, for that matter. Majority vote* is actually *a way to interpret* the results of the poll: it is the aggregation rule that sanctions that, when there are only two options and no ties are allowed for individual votes*, the results of the poll are to be interpreted as telling us that the social (or group) preference coincides with the option that has the most votes*. In a similar guise, say, special majority voting* with a quota of $2/3$ is the rule that interprets the results of a poll as telling us that the social preference is the one with at least $2/3$ of the votes*.

It is a crucial feature of our model that the future fact be the result of a collective action. Consider again COURAGE, and assume once more that majority voting* gives us the opinion of the group. If there were a representative of the group in charge of speaking up on behalf of her peers, the CCJT would be inapplicable: the probability that the group does F given that a majority votes* for F is 1, regardless of the number of members. Group reliability would no longer follow a binomial distribution.

6.2. Attunement to the proto-truth

Recall that we defined the proto-truth as the majority a specific group would achieve in the long run, if infinitely enlarged with further voters and iterated over infinitely many rounds, whatever that majority turns out to be. Note that this notion is inevitably indexed to a given group. One possible source of uneasiness, however, is that this seems to require an identity criterion for a group that is not already there. A closely related perplexity is what "attunement" can mean in this context. How can players be attuned to the voting behavior of people they have not met (and can never meet, because the group is infinitely large)?

On a closer look, however, neither the identity criterion for the enlarged group nor the attunement condition is as strange as they might seem at first sight. We need to assume that a particular group has an identity criterion robust enough to generate a determinate limiting outcome, but the group identity need not be given by physical co-presence or causal interaction; rather, it can be defined by an abstract distribution of types (propensities, dispositions, probabilities of voting). As a result, the proto-truth is not an ontological property of the world in the way the actual outcome is, but a modal or dispositional property (e.g., "if this particular group were infinitely enlarged, and voted in infinitely many rounds, this is what would eventually win."). In turn, agents can have good informal heuristics for anticipating what the extended group will converge to, in the sense that they might be good at reading cues, norms, or history, and capturing regularities that are already latent in the finite group as each of them knows. In this

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

sense, attunement can be understood as the possession of an implicit model of group dynamics, or perhaps as a tracking relation between an individual and the group's dispositional property.[9]

6.3. Sincere voting*

We should note here that sincerity and strategic reasoning need not be opposed. A sincere vote is one that faithfully expresses the voter's genuine expectations about the outcome (so, indirectly, each vote* functions as a private estimate of the future collective action G, even if voters* need not aim at making explicit estimates). If the voter is rational, those expectations will typically be shaped by strategic considerations, i.e., by what they take others to be likely to do, and by how the payoffs align.

In any case, what motivates an agent to cast a sincere vote* for Stag or for Hare? Evidently, for any sincere voter* i, we will have $V_i = 1$ iff i's payoff from going for Stag is higher than the payoff from going for Hare. Whether this is or is not the case depends on i's assessment of what others might do. Let "$c_i(G_j = 1)$" be i's credence that j goes for Stag (note that c_i is *not* a credence about what j will vote* for, but only about what j will *do*). Thus, for example, for the concrete matrix proposed in Section 3, for every player i, if $c_i(G_j = 1) = q_1$ and $c_i(G_k = 1) = q_2$, then $V_i = 1$ iff $8q_1 + 8q_2 - 7q_1q_2 - 5 > 0$; in particular, if $q_1 = q_2$, a brief calculation shows that $V_i = 1$ iff $q_1 > \frac{8-\sqrt{29}}{7} \approx 0.374$.

6.4. Individual Hope

Individual Hope, as measured by both $P(G_i = 1|V_{Stag})$ and $P(G_i = 0|V_{Hare})$, represents the probability that the outcome of the voting round will shape player i's actions. Recall that to make the Constructive Condorcet setting possible, we need both quantities to be above 0.5, although they need not be the same number. This is not a credence, but an objective probability. As before, player i will *go* for Stag iff her going for Stag has higher payoff, and this depends on what she believes other agents will do. This time, however, player i has additional information, namely, the result of the voting* scenario. To what extent do agents view the aggregation of votes* *as a signal* of what others collectively will do, i.e., as a signal of the true value of G? This is the key question that Individual Hope tries to capture.

As we have seen, if Individual Hope > 0.5, both $P(G_{Stag}|V_{Stag})$ and $P(G_{Hare}|V_{Hare})$ will tend to 1 as n tends to infinity. Explaining the reasonableness of Individual Hope greater than 0.5 takes a detour through the players' individual credences regarding what others will do. The reason is that, if rational players come to know that $V_{Stag}[V_{Hare}]$ is true, and they take V to be a signal to some extent, they will change their initial c_i accordingly. In fact, this is what taking V as a signal means: it prompts them to update their prior c_i. In other words, for each i and j, if c' is the new revised credence, $c'_i(G_j = 1|V_{Stag})[c'_i(G_j = 0|V_{Hare})]$ will tend to be higher than $c_i(G_j = 1)[c_i(G_j = 0)]$, and $c'_i(G_j = 0|V_{Stag})[c'_i(G_j = 1|V_{Hare})]$ will tend to be lower than $c_i(G_j = 0)[c_i(G_j = 1)]$. Note that, in the limit (but not necessarily before), all agents i will coincide in having

[9]In any case, the proto-truth depends heavily on how we individuate the *current* group, because it is a construct of how we project the group's identity into the indefinite future. An agent is then aligned with the proto-truth insofar as he or she shares, anticipates, or reinforces the practices and expectations that sustain that group identity.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

credence 1 or 0 that all of their fellow players go for Stag/Hare, and hence all *i* will in turn perform the same action.[10] We return to this in the next Section.

Individual Hope marks the extent to which an agent is predisposed to follow V as a signal of what people will do. Having Individual Hope >0.5 is a reasonable assumption if we also assume sincere voting*, as in that case the aggregated votes* are a good indicator of what a majority of people were considering doing. Individual Hope > 0.5 in fact *means* that there is a probability greater than 0.5 that the c'_is will now change, if needed.

7. Majority voting*, uniformity, and cooperative behavior

So far, we have obtained a constructive application of the CJT: Under certain conditions, majority voting* tracks a future event. But, by itself, this falls short of establishing the possibility of a middle ground between proceduralism and epistemicism. In a purely proceduralist account, an outcome is justified merely in virtue of being the outcome of a rule, which is in turn justified by its formal properties. In a purely epistemicist account, an outcome is justified in virtue of something being (antecedently) true in the world. In our case, so far, we have shown how the voting* scene can help build a particular social outcome (Stag or Hare). But a question remains: is tracking future truths in the manner just described desirable?

In fact, we already have enough clues to answer this question, at least partially. In the long run, the group will tend to be uniform. In the limit, the social outcome will be either all Stag or all Hare. This means that there will be a potential improvement in payoffs: it reduces the chance of Stag outcomes when the majority goes for Hare (so we progressively reduce the minority that gets 0 utils), and it also reduces the chance of Hare outcomes when the majority goes for Stag (so it progressively reduces the minority that gets 6 when they can instead get 10). In the limit, even if we cannot guarantee that they will all receive the maximum payoff, we can at least rule out that some agents obtain the minimum. Indeed, with these payoffs no agent's lot ever worsens in the transition to uniformity; incumbent Hare-hunters remain at 5, while everyone else strictly gains. The uniform outcome the mechanism promotes is thus Pareto-superior to the status quo, and, given the stipulation of Section 3, this holds for odd groups of any size, not merely for $n = 3$.[11] Even if nothing else could be said in favor of majority voting*, this already amounts to a nice justification. Majority voting* is efficiency-conducive.

What if we drop the sincerity assumption? I will argue that, as long as a few other mild assumptions hold, we can still obtain that majority voting* helps build a desirable outcome.

In this spirit, let us make no particular stipulations on whether players know/do not know whether the result of the voting* will be made public; likewise, let us make no

[10]Note that we do not need the objective probabilities $P(G_i = 1|V_{Stag})$ and $P(G_i = 0|V_{Hare})$ to be individualized or identified by anyone before the actual actions are aggregated. We can say that agents behave *as if* there were such a P. On the other hand, we could mediate the change of c_is through each agent's conjecture that *others take it to be a signal as well*, but we do not in fact need such a detour in the present setting. This will change if we change the sincerity assumption; see the next section.

[11]This is of course reminiscent of Skyrms (2004, 2010), who shows how public signals and minimal structural features can tip coordination games toward the cooperative equilibrium. Likewise, Bicchieri (2006) argues that social norms emerge when individuals prefer to conform, conditional on expecting others to do the same; before them, Schelling (1960) had famously shown that coordination often depends on the emergence of focal points.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

particular stipulations on whether they know/do not know whether they will in fact need to act right after the voting*.

Indeed, if players have strong reasons to suspect that the result of the voting* will be made public and that they will have to act right after knowing the result of the voting*, then they will have incentives to vote in such a way as to produce an effect on others at the final stage. In this scenario, their initial c_i becomes irrelevant. In fact, it is reasonable for all players to vote so as to signal that they will do $G_i = 1$ in due course, with the hope that this will in turn move others to go for Stag when the time comes. As players will only learn the result of the aggregation of votes*, this is tantamount to saying that it is in the players' interest to actively seek that V_{Stag} be the case, and hence it is in the players' interest to vote* for Stag. Hence, a vote* for Stag is no longer a mark of their trust in whether others will do Stag when the time comes, but it is actually an attempt to force (the construction of) a collective signal.

There is an obvious complication, of course, which is that, if players suspect that they are all trying to force the construction of a collective signal, it is not clear whether we can still have Individual Hope > 0.5, i.e., it is not clear whether we can still assume an objective probability $P(G_i = 1|V_{Stag}) > 0.5$.

This line of thought may motivate a skeptical conclusion, according to which the result of the voting* is irrelevant; acting will just require trust in others, as in a regular Stag Hunt game.

The skeptical conclusion can be resisted, however. All we need for this to happen is that at least some players conjecture that others might behave *as if V were a signal*.[12]

According to this analysis, securing Individual Hope greater than 0.5 depends crucially on the *credence* each agent has *about whether their fellow players have Individual Hope* > 0.5. Call this credence p_{ij}:

$$p_{ij} := c_i(P(G_j = 1|V_{Stag}) > 0.5) \quad (\textit{for } j \neq i)$$

where c_i is the credence function defined in Section 6.3. Define also i's aggregate credence about Individual Hope as

$$\bar{p}_i = \frac{1}{n-1}\sum_{j \neq i} p_{ij}.$$

It is now *the aggregation of the* p_{ij}*s* that needs to be above a certain threshold for agents to choose the cooperative option, insofar as the new relevant (objective) probability for i is now $P(G_i = 1|V_{Stag}, \bar{p}_i = x) = x$. In this scenario, agent i's Individual Hope just is $\bar{p}_i$: her objective probability of going for Stag given V_{Stag} equals her average credence that others treat V as a signal. And, thanks to Grofman et al. (1983), the CCJT mechanism then only needs the average across agents of the $\bar{p}_i$ to exceed 0.5.

Two small points:

1. First, note that a sufficiently high credence that Individual Hope is greater than 0.5 *creates* objective Individual Hope greater than 0.5.
2. Second, many authors contended that, at the time of making a decision, we cannot have credences on what we will ourselves do. This is what Isaac Levi a long time

[12]For example, agents may conjecture that at least some of their fellow players reason as follows: "Given that the aggregated vote* V will actually fail as a signal, I might as well vote sincerely."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

ago termed "deliberation crowds out prediction" (Levi 1997). Schick (2003) takes the "deliberation crowds out prediction" phenomenon as a prime example of a blindspot for the first person. Regardless of whether we find this compelling, it is important to point out that we do not *need* to assume agents can have credences on what they will themselves do.

To sum up, if all the players know that V_{Stag} is true, and they conjecture that the premises of *the second* CJT are fulfilled (i.e., independence of individual actions given the signal, and Individual Hope higher than 0.5), then they will also conjecture that $P(G_{Stag}|V_{Stag})$ tends to 1. Then rational agents will fix their c_i's accordingly. Again, in the limit (but not before), for all i, their c_is will coincide in giving the same 1 or 0 to every other agent, and all i will coincide in doing the same action. This time, however, as V is most certainly V_{Stag}, in the limit they will all cooperate. Note, however, that under these incentives V carries no information about the distribution of intentions (every player votes* for Stag out of interest) so it operates as a focal public event rather than as a Condorcetian aggregate; the convergence of actions is driven by (conjectured) Individual Hope, i.e., by second-order trust.

Recall that, in standard assurance games, players cooperate if they trust each other. In our setting, by contrast, acting has been preceded by voting*, and instead of simply assuming that agents have sufficient trust in the fact that their fellow players will go for Stag, we assumed that, on average, players trust that others will treat V as a signal. There are in turn various well-known ways of making this conjecture reasonable, as long as we allow for some flexibility in our framework. For example, we can introduce small changes in the incentive structure, such as a mild penalty $\kappa > 0$ for acting in a way that deviates from the original vote, or perhaps a reputational cost.[13] I will not develop this point further; it suffices to observe that the assumption required is, in principle, weaker than those typically made in assurance games.

Let us take stock. Under the assumption of sincere voting*, it is rational for agents to interpret V as a signal; as a result, in the limit payoffs are improved, even if no maximum payoff is guaranteed. Even if we weaken the sincerity assumption, we can still obtain a payoff improvement in the limit (actually, *maximum* payoff improvement), as long as we replace sincerity by a different requirement, to wit, that players may conjecture that others behave as if V were in fact a signal. This requirement is typically less stringent than merely conjecturing that others will behave cooperatively, although this depends on the flexibility of the framework.

The resulting account is epistemic (as opposed to merely procedural) in the sense that the procedure tracks something whose value is not constituted by the tracking itself. In the classical CJT, that something is a factic truth, while here it is a normatively desirable outcome. Pareto-superiority over the status quo serves as the criterion that anchors this value independently of the procedure. It is, in this sense, the constructivist analog of objective truth.

[13]See for instance Skyrms (2010). Note that introducing a small consistency penalty slightly modifies the assumption that votes* are entirely non-binding (so it is no longer purely "cheap talk"), but does not trivialize the setting. Agents may still vote one way and act another, and it can still be rational to do so. For instance, a voter who supported Stag may later play Hare if a Hare majority emerges and the payoff advantage of Hare exceeds the small penalty κ. The penalty's role is only to make the polling outcome a credible public signal, not to force obedience outright.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

8. Relation to existing approaches

The result established in the preceding sections invites comparison with several families of mechanisms that address related problems. This section examines those connections to identify what it shares with each family, where it diverges, and what it contributes that the alternatives do not. As we shall see, related approaches differ both in their formal machinery and, more importantly, in the role they assign to aggregation.

1. *Opinion Pooling: The Hegselmann–Krause Model and Variants*

   The Hegselmann–Krause (HK) model (Hegselmann and Krause 2002) describes a process in which agents iteratively revise their opinions by averaging with those of neighbors whose views fall within a given confidence radius. Under suitable conditions, the group converges to one or more consensus clusters. HK-type models have generated a substantial literature in formal epistemology and social simulation, and they share with the CCJT the ambition of explaining convergence without presupposing an antecedent external fact.

   There are significant differences between the two approaches, however. First, HK is a model of belief revision, rather than a model of decision under strategic uncertainty. The primary focus is on the temporal evolution of beliefs, including the conditions under which consensus, polarization, or clustering emerges. The CCJT, by contrast, is embedded in a game-theoretic structure; Individual Competence and Individual Hope are defined relative to payoffs, and the convergence result bears on equilibrium selection. This means the CCJT can address questions that HK models cannot. In particular, it can address the question of why rational agents with strategic concerns would treat a poll as a reliable signal of future collective behavior. In other words, the CCJT isolates a setting in which an appropriately structured aggregation procedure can function as a coordination device. Thus, under suitable conditions, aggregated signals can guide agents toward an equilibrium that is not otherwise accessible.

   Second, HK convergence is iterated (it requires many rounds of pairwise interaction), whereas the CCJT operates through a single aggregation step followed by individual action, a structure that better fits one-shot institutional settings such as pre-election polls, organizational surveys, or straw votes in deliberative assemblies.

   Finally, the HK convergence result is not derived from competence and independence conditions, and it cannot be evaluated by the standards that epistemic democrats apply to collective procedures, which is why it cannot serve as a bridge between the epistemic democracy literature and the coordination literature.

2. *Norm Dynamics*

   On a different line, a large literature in philosophy and economics explains the emergence of social norms and conventions through repeated interaction and evolutionary processes (Schelling 1960; Skyrms 2004). These models typically show how, over time, populations may converge on stable patterns of behavior, often in a path-dependent manner. Skyrms (2004, 2010), for example, shows through evolutionary game-theoretic analysis that public signals and minimal structural conditions can tip coordination games toward cooperative equilibria. Related to this, Bicchieri (2006) argues that social norms emerge and are sustained by conditional preferences; agents prefer to conform when they expect others to conform. Both accounts are well suited to explaining how, once a cooperative

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

equilibrium has been reached, it is maintained. The CCJT, on the other hand, addresses the logically prior problem of how a particular coordination outcome comes to be selected in the first place, given that agents face genuine uncertainty about what others will do. In other words, the polling mechanism operates at the pre-norm stage, providing a public signal that can break the symmetry between equilibria. In Skyrms's terms, the CCJT models the construction of the kind of public signal that his evolutionary analysis takes as a starting point; in Bicchieri's terms, it explains how agents come to form the conditional expectations that make norm-following rational. The CCJT and norm dynamics accounts are therefore complementary.

3. *Cheap Talk*

The cheap talk literature (Farrell and Rabin 1996) studies settings in which agents communicate before acting without any mechanism binding their subsequent behavior to their messages. Pre-play communication in Stag Hunt games is a well-studied case; for example, it is well-known that non-binding announcements of intent can support cooperative equilibria that would otherwise be difficult to sustain (Blume and Ortmann 2007; Charness 2000). Voting* as defined in this paper can be understood as a structured form of cheap talk: agents announce their intentions, no enforcement follows, and the announcement is aggregated before being made public. This also connects naturally with the voting-as-signaling literature (Aytimur et al. 2014), which treats voting as a way for individuals to signal private intentions. In the present setting, however, the signal is aggregated before being made public, and votes* are meant to track the proto-truth G*.

The CCJT differs from standard cheap talk models in various ways. Cheap talk models typically focus on which messages are incentive-compatible and under what conditions a sender can credibly transmit private information to a receiver. In the CCJT setting, by contrast, there is no sender-receiver asymmetry, and agents do not possess private information about an external state of the world. Instead, each agent has a private estimate of the proto-truth G* (the twice-idealized majority) and voting* aggregates these distributed estimates into a public signal. The CCJT framework then specifies conditions under which that aggregation is reliable; those conditions have no direct analog in standard cheap talk analysis. On the other hand, that agents do in fact treat aggregated voting outcomes as signals is supported by experimental evidence. Anderson et al. (2022) show that in sequential voting settings people systematically update their beliefs in response to earlier votes. Individual Hope thus can be seen as a formal account of a behavioral regularity that has been documented empirically.

Recall that the baseline model assumes sincere reporting of intended actions. This assumption may be questioned in settings where agents have incentives to misrepresent their preferences. The present framework should therefore be understood as identifying conditions under which aggregation can indeed function as a coordination device. Extensions that incorporate modest incentives for sincerity, such as small costs associated with deviating from stated intentions, suggest that the mechanism can remain effective under weaker assumptions, though a full treatment of strategic considerations is left for future work.

A further distinction concerns scope. The CCJT has a narrower scope than the cheap talk literature, but the two approaches address different questions; cheap talk models ask when communication helps across a wide range of strategic settings, while the CCJT asks why majority aggregation has epistemic value in the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

specific class of cases that matter for democratic theory. Related to this, Agranov (2024) finds experimental evidence to the effect that cheap-talk communication in Stag Hunt games is highly effective when both players benefit symmetrically from the Pareto-dominant equilibrium, but not so much under payoff asymmetry, even when the Pareto-dominant equilibrium still yields higher total surplus. This is consistent with what the CCJT would lead one to expect. The mechanism's success conditions in the CCJT are formulated for settings with two Pareto-rankable equilibria and symmetric payoffs, which is where Agranov's results locate the effectiveness of pre-play communication. A natural next step would be to extend the current framework to payoff-asymmetric settings and ask whether the conditions on Individual Hope survive there.

4. *Related Condorcet Approaches*

Finally, Muldoon's (2016) cross-perspectival CJT uses a Condorcet-inspired framework to evaluate the social robustness of beliefs. He does so by aggregating assessments from agents with different epistemic perspectives. The mechanism is designed to filter perspective-dependent biases and produce beliefs that are reliable across viewpoints. Note that this is in fact a complementary project. Muldoon's mechanism assesses the credibility of existing beliefs, while the CCJT explains the construction of a new social fact through coordinated action. The two results operate at different stages of the epistemic process and address different questions.

9. Conclusions

The main purpose of this paper has been to establish the existence of a middle ground between epistemicism and formalism. I have shown that majority voting* (majority polling) can track a future truth that is partly built as a result of the voting*. For this to be a defense of a middle ground, it was also important to show that such a future truth was antecedently desirable. Thus, I have also shown that, for settings structurally similar to a generalized Stag game, majority voting* has the potential to improve payoffs and perhaps also build cooperative results. The details of the setting are not important, as long as there are two possible equilibria, one optimal and the other sub-optimal. More precisely, if we assume sincerity, the group will converge on doing the same thing (either all Hare or all Stag). In addition, if we relax sincerity but assume that at least some players trust that others will treat V as a signal, we get payoff maximization; this last assumption is in turn justified if we also assume small penalties associated with any deviation from the original vote*.

A subsidiary result is that the CJT is more versatile than what we might have anticipated. It gives us the tool to establish a formal bridge between the well-known intuition that polling helps coordination, and the epistemic thesis according to which majority decisions track a kind of truth. In the model presented here, the truth is endogenous, as it is built via the voting* process itself (the "future truth" of cooperation). A constructivist application of the CJT, or CCJT, helps us show that finite majorities approximate the twice-idealized majority with probability converging to 1.

The comparisons developed in Section 8 suggest that the CCJT occupies a specific niche in the broader landscape of mechanisms for constructing social facts. The CCJT is meant to model the moment at which a group facing strategic uncertainty uses a structured aggregation procedure to construct the conditions for its own cooperation, and it does so in terms that allow the procedure to be evaluated by the criteria of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

epistemic democracy. This moment is common in pre-election polling, organizational surveys, straw votes in deliberative assemblies, and informal consultations before collective action, among many other situations. This suggests that the niche is practically important, even if the conditions the theorem requires are idealized.

I have not claimed that the assumptions introduced here are easily satisfied. Similar concerns have long been raised about the applicability of the standard CJT. Even though uniform individual competence and strict conditional independence are rarely if ever instantiated in actual deliberative settings, the theorem has remained a central reference point in the epistemic democracy literature; the gap between that idealization and actual democratic practice has not prevented the theorem from shaping the debate. The same charitable standard applies here. There is no clear reason to think that the assumptions required in this setting are harder to meet than those of the classical theorem. Moreover, the target is less demanding: the classical CJT asks a group to track a mind-independent truth with near infallibility, whereas the CCJT asks it to construct a desirable outcome under suitable conditions. Tracking a fact that exists independently of the procedure, and doing so with asymptotic certainty, is a more ambitious epistemic achievement than converging on a Pareto-superior equilibrium given sufficient coordination. Whether real-world groups satisfy these assumptions is ultimately an empirical matter, one that invites further investigation. The bar, at least, is not higher than the one the classical theorem already sets.

Acknowledgments. Previous versions of this paper were presented at various conferences and workshops, including the 8th Workshop of the UK-Latin America Network for Political Philosophy (University of San Andrés, Buenos Aires, April 2026), the Carlos Nino Colloquium 2021 (University Torcuato Di Tella, Buenos Aires, 2021, with comments by Ezequiel Monti), the III International Colloquium of Epistemology (University of Costa Rica, 2021), the Philosophy Department Colloquium at Carnegie Mellon University (online, 2021), CLMPST 2019 (Prague, 2019), and the VIII Conference on Social Epistemology (Federal University of Santa Maria, Brazil, 2018). I would like to thank the participants in all these events for very helpful discussions and suggestions. I am especially grateful to Francisco García Gibson for his careful reading of the payoff matrix and for prompting a correction. I am also grateful to Ariel Zylberman for pressing me on the relation between the present proposal and constructivist approaches to normativity, and to Martín Farrell for drawing my attention to possible connections between the present proposal and Carlos Nino's attempt to articulate a middle ground between purely procedural and strongly epistemic accounts of democracy. I leave fuller comparisons for another occasion.

Competing interests. The author confirms that there are no conflicts of interest relating to this work.

Use of AI. The author used AI-assisted tools during the revision process to improve the clarity and readability of the English. The intellectual content and arguments are entirely the author's own.

References

Agranov M. (2024). 'Communication in Stag Hunt Games: When Does It Really Help?' *Economics Letters* **244**, 111991.

Anderson L.R., Holt C.A., Sieberg K.K., and Freeborn B.A. (2022). 'An Experimental Study of Strategic Voting and Accuracy of Verdicts with Sequential and Simultaneous Voting.' *Games* **13**(2), 26.

Austen-Smith D., and Banks J.S. (1996). 'Information, Aggregation, Rationality and the Condorcet Jury Theorem.' *American Political Science Review* **90**(1), 34–45.

Aytimur R.E., Boukouras A., and Schwager R. (2014). 'Voting as a Signaling Device'. *Economic Theory* **55**(3), 753–77.

Bicchieri C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133

Blume A., and Ortmann A. (2007). 'The Effects of Costless Pre-Play Communication: Experimental Evidence from Games with Pareto-Ranked Equilibria.' *Journal of Economic Theory* **132**(1), 274–90.
Bradley R., Dietrich F. and List C. (2014). 'Aggregating Causal Judgments'. *Philosophy of Science* **81**(4), 491–515.
Charness G. (2000). 'Self-Serving Cheap Talk: A Test of Aumann's Conjecture.' *Games and Economic Behavior* **33**(2), 177–94.
Condorcet M. de (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.
Dietrich F. (2008). 'The Premises of Condorcet's Jury Theorem Are Not Simultaneously Justified.' *Episteme* **5**(1), 56–73.
Dietrich F., and Spiekermann K. (2013a). 'Epistemic Democracy with Defensible Premises.' *Economics and Philosophy* **29**(1), 87–120.
Dietrich F., and Spiekermann K. (2013b). 'Independent Opinions? On the Causal Foundations of Belief Formation and Jury Theorems.' *Mind* **122**(487), 655–85.
Estlund D.M. (1994). 'Opinion Leaders, Independence, and Condorcet's Jury Theorem.' *Theory and Decision* **36**(2), 131–62.
Estlund D., Waldron J., Grofman B., and Feld S.L. (1989). 'Democratic Theory and the Public Interest: Condorcet and Rousseau Revisited.' *American Political Science Review* **83**(4), 1317–40.
Farrell J., and Rabin M. (1996). 'Cheap Talk.' *Journal of Economic Perspectives* **10**(3), 103–18.
Goodin R.E., and Spiekermann K. (2018). *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.
Grofman B., Owen G., and Feld S.L. (1983). 'Thirteen Theorems in Search of the Truth.' *Theory and Decision* **15**(3), 261–78.
Hegselmann R. and Krause U. (2002). 'Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation.' *Journal of Artificial Societies and Social Simulation* **5**(3).
Landemore H. (2013). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton, NJ: Princeton University Press.
Levi I. (1997). 'Prediction, Deliberation, and Correlated Equilibrium.' In *The Covenant of Reason: Rationality and the Commitments of Thought, Ch. 5*, pp. 102–16. Cambridge: Cambridge University Press.
List C., and Goodin R.E. (2001). 'Epistemic Democracy: Generalizing the Condorcet Jury Theorem.' *Journal of Political Philosophy* **9**(3), 277–306.
Muldoon R. (2016). *Social Contract Theory for a Diverse World*. New York: Routledge.
Pearl J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.
Rawls J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Schelling T.C. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Schick F. (2003). *Ambiguity and Logic*. Cambridge: Cambridge University Press.
Shogenji T. (2003). 'A Condition for Transitivity in Probabilistic Support.' *British Journal for the Philosophy of Science* **54**(4), 613–16.
Silva R. (2024). 'Coordination in Stag Hunt Games.' *Journal of Behavioral and Experimental Economics* **113**, 102290.
Skyrms B. (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.
Skyrms B. (2010). *Signals: Evolution, Learning, and Information*. Oxford: Oxford University Press.
Vanderschraaf P. (2019). *Strategic Justice: Convention and Problems of Balancing Divergent Interests*. Oxford: Oxford University Press.

Eleonora Cresto (PhD, Columbia University, 2006) is Associate Professor at Universidad de San Andrés, Argentina, as well as Principal Researcher at Argentina's National Scientific and Technical Research Council (CONICET). Her research focuses on formal and informal epistemology, philosophy of science, decision and game theory, social choice, and philosophy of logic. She is an editor of *Erkenntnis* and Secretary-General of the *International Union of History and Philosophy of Science and Technology*.

Cite this article: Cresto E. (2026). "A Constructive Condorcet Jury Theorem." *Episteme* 1–20. https://doi.org/10.1017/epi.2026.10133

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 05:08:59, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/epi.2026.10133