Economics & Philosophy (2025), 1–25
doi:10.1017/S0266267125100503

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Does this democracy work? Measuring reliability, Trojan horses and the fundamental problem of evaluation

Richard Beadon Williams

Department of Science, Technology and Society, School of Social Sciences and Technology, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany
Email: richard.williams@tum.de

(Received 03 January 2024; revised 09 March 2025; accepted 07 April 2025)

Abstract

Epistemic democrats indirectly evaluate democratic decisions by directly evaluating the inputs into the election. However, the fundamental problem of measurement in the philosophy of science shows that procedures are often as difficult to evaluate as outcomes. This paper brings this highly refined framework into political philosophy to show that epistemic democrats face an analogous 'fundamental problem of evaluation'. This cross-fertilization of political philosophy with the philosophy of science shows that the quality of democratic mechanisms and their inputs regarding their ability to track the truths of justice is as difficult to evaluate as the quality of the resulting decisions themselves.

Keywords: Epistemic democracy; Condorcet Jury theorem; diversity trumps ability; legitimacy; evaluation

1. Introduction

Did UK democracy perform well during the 2016 Brexit referendum? Did US democracy perform well during the 2016 Clinton/Trump or the 2020 Biden/Trump presidential elections? Do those democratic states perform better after Brexit and Trump's first term? How well did those democratic states perform a decade before? Epistemic democrats argue that voting or deliberating can empower democracy to outperform antidemocratic alternatives. However, for epistemic democracy to become a practical ideal for evaluating the reliability of real democratic states, epistemic democrats need the ability to evaluate the reliability of real democratic states at specific times and places in light of their theorems.

The core contribution of this paper is to bring the highly refined framework of the fundamental problem of measurement in the philosophy of science into political philosophy to show that epistemic democratic evaluations of real democratic states remain defective in light of what I call 'the fundamental problem of evaluation'. In the philosophy of science, the measurement problem shows that it is not always easy

for scientists to accurately quantify the value of various variables because they can face a vicious justificatory circle. For example, chemists must know that a thermometer is accurate before they can measure the temperature of some water, but they must know the temperature of the water to check that the thermometer is accurate. Similarly, the evaluation problem shows epistemic democrats also face a vicious justificatory circle when they aim to evaluate the reliability of real democratic states. Epistemic democrats must know the election was reliable before they can know that the decision was correct, but they must know that the decision was correct to check that the election was reliable. By using recent research in the philosophy of science, this paper provides a novel and innovative discussion of this long-standing but neglected problem for epistemic democracy.

This paper uses the measurement problem in the philosophy of science to analyse a variety of plausible epistemic democratic solutions to the evaluation problem – including those of Dryzek and List (2003), Goodin and Estlund (2004), Estlund (2008), Goodin and Spiekermann (2018), and Berger and Sales (2019) – to show that they are much harder to defend than typically thought. Epistemic democrats typically accept what I will call 'procedure-first' solutions. However, philosophers of science typically reject 'procedure-first' solutions and accept what I will call 'coherentist' solutions. Following the philosophy of science framework, epistemic democrats should prefer 'coherentist' solutions to their dominant 'procedure-first' solutions because the reliability of democratic mechanisms regarding their ability to track the truths of justice is no easier to evaluate than the correctness of the resulting democratic decisions themselves. Whatever the preferred solution might be, this cross-fertilization of political philosophy with the philosophy of science shows that epistemic democrats should significantly revise how they aim to evaluate the reliability of real democratic states. With polarized electorates doubting the ability of democracy itself to work well in divisive referendums and controversial elections, this theoretical problem is of growing practical concern.

2. Epistemic Democracy

Epistemic democracy is one of the most popular and plausible defences of democratic legitimacy (Cohen 1986; Coleman and Ferejohn 1986; Anderson 2006; Estlund 2008; Landemore 2012; Perote-Peña and Piggins 2015; Goodin and Spiekermann 2018; Cerovac 2020). Epistemic democrats successfully show that fairness is not enough for legitimacy. As Estlund explains 'if making the decision in a fair way (as in a coin flip) is insufficiently likely to produce the fair or just or morally required outcome, it may not be good enough' (Estlund 1997, 176). The epistemic turn in democratic theory turns away from prioritizing fairness and towards prioritizing competence. It is not enough for democracy to be a fair procedure to be legitimate. Democracy needs to be competent at making the correct decisions.

How can epistemic democrats show that real democratic states are competent at making the correct decisions when which decisions are correct is so deeply disputed? The central virtue of epistemic democracy is that it abstains from direct moral evaluations of democratic decisions. Epistemic democrats retreat to (seemingly) more modest evaluations of the democratic process itself. As

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

explored below, they argue that the epistemic qualities of democratic voting or democratic deliberation allow us to expect democracy to make the correct decisions more often than antidemocratic alternatives. Epistemic proceduralism limits antidemocratic alternatives to fair or otherwise acceptable procedures. Democracy can outperform a fair coin flip (Estlund 2008: 98–116). More demandingly, epistemic instrumentalism does not limit antidemocratic alternatives to fair procedures. Independently of how unfair expert rule – epistocracy – might be, democracy can outperform epistocracy anyway (Landemore 2012: 50–52; Goodin and Spiekermann 2018: 225–243).

Joshua Cohen provides one of the original and still useful frameworks for conceptualizing epistemic democracy (Cohen 1986: 34). Firstly, for any decision to *be* correct, epistemic democrats typically assume a procedure-independent standard of correctness.[1] The default assumption is that democracy typically makes a binary choice between the correct answer and an incorrect answer.[2] In theory, epistemic democrats could associate the standard of correctness with whatever bundle of values they wish.[3] However, in practice, epistemic democrats typically associate the standard of correctness with justice (Estlund 2008: 169; Landemore 2012: 45; Goodin and Spiekermann 2018: 5). The correct answer is whichever decision promotes justice better. Second, for people to *care* about which decisions are correct, epistemic democrats typically assume sincerity. Rather than cynically advocating for whatever they expect to advance their narrow interests better, people in politics sincerely advocate for whatever they judge will promote justice better. Third, for people to *know* which decisions are correct, epistemic democrats assume some democratic mechanism – typically elections, referendums or deliberations – will discover which decisions are correct from the standpoint of justice. Epistemic democrats directly evaluate the quality of the inputs – the votes and voices – put into the democratic process to indirectly evaluate which democratic decisions are correct.

Both the Condorcet Jury theorem and the 'Diversity Trumps Ability' theorem provide two paradigmatic models of epistemic democracy. In an aggregative direction, the classic Condorcet Jury theorem contains three major assumptions (Goodin and Spiekermann 2018: 17–36). First, the sincerity assumption assumes all people vote to express their judgements about justice rather than to advance their narrow interests. Second, the competence assumption assumes all people perform better than random. In a binary choice between the correct decision and an incorrect decision, a fair coin flip will choose the correct decision as often as the incorrect decision. In contrast, a Condorcetian competent person chooses the correct decision more often than not. Third, the independence assumption assumes all people vote to express *their* judgements about justice rather than the judgements of someone else. This protects against large-scale error. If every correct person votes

[1]Of course, the procedure-independent standard of correctness does not assume moral Platonism or moral realism (see Landemore 2012: 210–219; Goodin and Spiekermann 2018: 38–42).

[2]This assumption can be relaxed in various ways to empower epistemic models of democracy to answer multiple-choice questions as well (Goodin and Spiekermann 2018: 26–31).

[3]Some epistemic democrats retreat from any standard of correctness (Peter 2013; Schwartzberg 2015). This interesting but atypical retreat from substantive epistemic benefits exceeds the scope of this paper.

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

to express the judgements of an incorrect person, nobody votes correctly. Conversely, when sincere and competent people vote independently, it is a mathematical truism that the probability that the majority chooses correctly increases exponentially as the number of people increases. So, Condorcetian majorities typically make the correct decision.

In a deliberative direction, the 'Diversity Trumps Ability' theorem assumes that sufficiently diverse problem-solving heuristics are dispersed among the members of a cognitively diverse group. In Hélène Landemore's words, 'it is often better to have a group of cognitively diverse people than a group of very smart people who think alike' (Landemore 2012: 103). It assumes people are competent enough to deliberate effectively and diverse enough for the group to contain effective heuristics – whatever the problem might be. When cognitively diverse people deliberate with each other, they spread those effective but scattered heuristics and temporarily induce competence throughout the group.

3. The Fundamental Problem of Evaluation

Political philosophers have already questioned whether epistemic democracy is a practical ideal. For example, Jason Brennan questions whether epistemic democracy is a practical *prescriptive* ideal (Brennan 2014). In fact, he argues that there is strong social scientific evidence to believe that democratic citizens will probably never become as competent as epistemic democracy typically demands of them. In contrast, I question whether epistemic democracy is a practical ideal in a much more modest sense. I question whether epistemic democracy is a practical *evaluative* ideal: can epistemic democrats evaluate what is *good* and what is *defective* about real democratic states in light of their theorems? This evaluative question is worth asking. Before epistemic democrats can prescribe specific reforms (and before epistocrats can then judge that such reforms are impractical), epistemic democrats should first establish whether they can use their theorems to evaluate how reliable real democratic states are – and how competent real democratic citizens are – to begin with.

Epistemic democrats typically aim to mirror the strategies that scientists use to evaluate scientific procedures to evaluate political procedures. As Estlund says, 'when some scientific procedure is held to have epistemic value, the argument must normally proceed in what I have called the formal epistemic manner. Arguments must be offered to show, whatever the truth is, this process has certain tendencies to ascertain it' (Estlund 2008: 170). However, I will show that because of their reliance on a scientific approach, epistemic democrats also invite the problems with evaluating scientific procedures into the process of evaluating political procedures. The politics/science analogy is a double-edged sword, with a significantly neglected second edge. It does not only cut one way.

The core contribution of this paper is to bring the highly refined conceptual framework of the fundamental problem of measurement in the philosophy of science into political philosophy. Philosopher of science Hasok Chang extensively analyses the history of measuring temperature in chemistry (Chang 2004). Most laypeople think that chemists simply need to put a mercury thermometer into a cup

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

of water and read what the thermometer says in order to measure the water's temperature. However, the history of chemistry shows that knowing the temperature of the water is not so easy because of the fundamental problem of measurement. In order to know the true temperature of the water, the chemists must first know that the thermometer is truly accurate. However, to check the true accuracy of the thermometer, the chemists must know the true temperature of the water in advance. So, the measurement problem for the chemists shows that they cannot really know the temperature of the water because they are trapped in a vicious justificatory circle.

Following philosopher of science Christine Elgin, philosopher of science Julian Reiss explores the fundamental problem of measurement in its most general terms (Elgin 1996, 1997; Reiss 2008). It is beneficial to quote Reiss in full:

> In order to know the value of a variable, we need to know that the measurement procedure associated with it is veridical (that is, that the procedure gives the correct result). But in order to be able to check whether the procedure is veridical, we need to know the variable's value. Since we have no independent access to either the value of the variable or the accuracy of the procedure, we can never know whether the measurement procedure is veridical or what the value of the variable is. (Reiss 2008: 64)

The fundamental problem of measurement applies when scientists lack independent access to the value of the variable and to the veracity of the measurement procedure.

Similarly, political philosophers face an analogous fundamental problem of evaluation. In order to know the correctness of a political decision from the standpoint of justice, political philosophers need to know that the political mechanisms are reliable. But to be able to check whether political mechanisms are reliable, political philosophers need to know the correctness of the political decision. Since political philosophers have no independent access to either the correctness of the political decision or the quality of political mechanisms, they can never know whether the political mechanisms are reliable or what the correctness of the political decision is. The evaluation problem applies to epistemic democracy. To know the correctness of democratic decisions from the standpoint of justice, epistemic democrats must first know that the democratic mechanisms that produced them were reliable. However, to check the quality of democratic mechanisms, epistemic democrats must know the correctness of the decisions in advance. So, the evaluation problem for epistemic democrats shows that they cannot really know how well real democratic states perform in light of their theorems because they are also trapped in a vicious justificatory circle.

Epistemic democrats might be concerned that the evaluation problem is too demanding. It would threaten too much of social science if social scientists had to evaluate *every* input into *every* theoretical model in light of *every* predicted output. If social scientists can frequently evaluate inputs independently of outputs, so can epistemic democrats. However, the problem is not whether epistemic democrats must evaluate *every* input into the democratic process in light of *every* resulting decision. The problem is whether epistemic democrats can evaluate *any* input into

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

the democratic process from the standpoint of justice independently of *any* resulting decision. Social scientists typically do not evaluate every input into every theoretical model in light of every predicted output because similar inputs have already gone through rigorous analysis in similar models. However, as explored next, it is very difficult for epistemic democrats to evaluate *any* inputs into *any* democratic process from the standpoint of justice. This is because whether *any* resulting decision promoted justice better than the alternative is typically deeply disputed.

4. Trojan Horse Justice

In order to see the significance of the fundamental problem of evaluation, it is helpful to recall philosopher of science Nancy Cartwright's analysis of public policy. A public policy does not work well purely because of its internal qualities, regardless of its relationship to external factors. A public policy works well largely because of its relationship with context-specific variables. Cartwright calls these context-specific variables 'support factors' (Cartwright 2012: 979). SImilarly, Cartwright calls intervening variables that harmfully interfere with how well a public policy works 'derailers' (Cartwright 2021: 13110). Whether a public policy works well or not largely depends on the presence of support factors and the absence of derailers. For example, Cartwright analyses the successful Progresa programme in Mexico, offering conditional cash transfers to mothers to improve child welfare. Progresa worked well in Mexico largely because of Mexico's clinical infrastructure. This one support factor out of many enabled the effective nutritional monitoring of children and significantly contributed towards improving child welfare. However, Progresa was then copied in 30 countries, whether the countries had comparable clinical infrastructure or not. Indeed, many of the later programmes failed because the lack of comparable clinical infrastructure meant a lack of effective nutritional monitoring of children present in the original Progresa programme.

Similarly, elections, referendums and deliberations are not reliable purely because of their internal qualities, regardless of their relationship to external factors. Effective voting behaviour and deliberation in one place and time may not remain effective in another place or time. As explored next, reliable democratic mechanisms risk derailers that harm how they work. In practice, the reliability of democracy is constantly competing against a bundle of derailers. Among the many threats to the reliability of democracy, I will use 'Trojan horses' as a paradigmatic type of derailer. Suppose a few cynical people cultivate Trojan horse conceptions of justice – deceptive conceptions of justice designed to appear public-spirited when they are actually self-interested. In order to advance narrow interests or partisan ideologies, Trojan horses mislead sincere people into mistaking incorrect decisions for correct decisions. They mask self-interested agendas with seemingly public-spirited arguments that say the agendas will promote justice.[4]

With very small margins for error, epistemic democrats must evaluate whether a bundle of derailers tip the balance and convert reliable democratic states into unreliable ones. Epistemic democrats could argue that the competence of

[4]Despite Jon Elster's expectation that the civilizing effect of hypocrisy will deter strategic self-interest in deliberation, the effect may only be to change the rhetorical strategies of cynical people (Elster 2000).

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

democratic citizens far exceeds their assumptions (Goodin and Spiekermann 2018: 9, 49, 52–53). However, epistemic democrats should not defend excessively strong assumptions. Otherwise, they risk imposing excessively demanding requirements on democratic citizens that they are less likely to meet. So, epistemic democrats should only argue that democratic citizens can meet the minimum thresholds that their assumptions require. Condorcetian democrats should only assume competent people perform *slightly* better than random. So, the margin of error between a reliable Condorcetian democracy and an unreliable one is actually very small. For example, Trojan horses need only deceive a few Condorcetian competent people to perform slightly worse than random to tip the balance and convert a reliable Condorcetian democracy into an unreliable one. Following the 'Diversity Trumps Ability' theorem, epistemic democrats typically assume that spreading a diverse range of problem-solving heuristics discovers the best solution. So, Trojan horses only need a few deceptive heuristics to deceive a few people into mistaking a worse solution for the better solution to tip the balance. Consequently, epistemic democrats need the ability to evaluate real democratic states at specific times and places to know if their theorems ever translate into practice.

I do not want to assume too high a burden of proof on epistemic democrats. Of course, the evaluation problem is a significant problem for *both* epistemic democrats and epistocrats. So, in the democracy/epistocracy debate, epistemic democrats need only show that the evidence supports epistemic democratic assumptions better than the epistocratic rejections of them. Alternatively, epistemic democrats need only show that the evidence supports that democracy is more reliable than epistocracy. Nevertheless, whatever the precise burden of proof might be, it remains very difficult for epistemic democrats to evaluate whether a bundle of derailers does tip the balance and convert reliable democratic states into unreliable ones. In particular, it is very difficult to evaluate if vote aggregation or diverse deliberation remain effective at empowering democratic states to outperform antidemocratic alternatives despite Trojan horses, or whether such mechanisms become ineffective or even counterproductive and start to mistake too many incorrect decisions from the standpoint of justice for correct ones.

5. Procedure-First Solutions

5.1 The Arbitrary Solution

Both philosophers of science and epistemic democrats can deploy what I will call 'procedure first' solutions. One type of solution to the fundamental problem of measurement is to say that scientists can know that the measurement procedure most closely associated with the variable is veridical after all. However, scientists lack independent access to the procedure's accuracy. So, scientists can decide that the measurement procedure is veridical by fiat. Following Elgin, Reiss calls this the arbitrary solution (Reiss 2008: 65). The measurement procedure is made to measure the variable's value accurately by definition. The variable's value is whatever the measurement procedure says. So, the measurement procedure becomes infallible. Returning to our original example, chemists can decide that the thermometer is veridical by fiat (Chang 2004: 148–152, 222). In so doing, the thermometer is made

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

to measure the temperature of the water accurately by definition, and the temperature of the water is whatever the thermometer says.

Unfortunately, the arbitrary solution has significant defects. First, scientists should evaluate the accuracy of measurement procedures in light of empirical research rather than a priori definitions. They should not know the accuracy of measurement procedures by definition, independently of any empirical research. Second, empirical research shows that measurement procedures are not infallible. The history of science shows that scientists typically discover more accurate measurement procedures over time. So, in principle, scientists should let empirical research determine the accuracy of measurement procedures and, in practice, empirical research frequently does show that new measurement procedures are more accurate than older ones. Consequently, a measurement procedure should not *define* the variable's value. Following Elgin, Chang and Reiss, the arbitrary solution should be rejected. Scientists cannot know the veracity of the measurement procedure associated with the variable in advance. As explored next, similar (but not the same) problems spill into 'procedure-first' solutions for epistemic democrats.

5.2 The Formal Solution

Epistemic democrats can defend a similar 'procedure-first' solution. One type of solution to the fundamental problem of evaluation is to say that epistemic democrats can know when their preferred democratic mechanism is reliable after all. For example, a purely procedural populist decides that elections are veridical by fiat. The election determines which decision is correct by definition. However, epistemic democrats obviously reject purely procedural populism (Cohen 1986: 28–29). They preserve the pre-theoretical intuition that democracy does not always choose the correct answer.[5] Epistemic democrats accept that *a procedure-independent standard of correctness* determines which decisions are correct, and that their preferred democratic mechanism tends to let democratic citizens *know* which decisions are correct. This is precisely what makes them *epistemic* democrats.

In a more plausible direction, epistemic democrats argue that they can gain independent access to the quality of the mechanism. In other words, access to the quality of the mechanism does not depend on evaluating its outputs. Rather than evaluating its *outputs*, access to the quality of the mechanism can depend on evaluating its *inputs* instead. Estlund says formal epistemic accounts evaluate democratic decisions in light of the epistemic qualities of the elections, referendums, or deliberations that produce them (Estlund 2008: 169–171). So, following Estlund, I will call this the formal solution. More generally, people in politics often evaluate the democratic process independently of the outcome. Even if people agree with the result, they can still think that the election or referendum contained significant problems. Alternatively, people can disagree with the result but still think that the election was not particularly dishonest or unfair.

The formal solution also has significant defects. Epistemic democrats typically defend conditional knowledge: if enough people meet certain conditions, elections

[5]One paradox for purely procedural populism is when a referendum and an election produce different answers to the same question. Unfortunately, this exceeds the scope of this paper.

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

are reliable. As Robert Goodin and Kai Spiekermann say, 'our analysis has been a conditional one. Assuming certain conditions (about competence, independence and sincerity) are satisfied, the pooling of votes by majority rule has epistemically beneficial properties' (Goodin and Spiekermann 2018: 321). Landemore says, 'there are good theoretical reasons to believe that when it comes to epistemic reliability, under some reasonable assumptions, the rule of the many is likely to outperform any version of the rule of the few' (Landemore 2012: 232). In Estlund's words, 'we expect communication (under the right conditions) to tend to make the individuals and the group better than random (the individuals less so than the group)' (Estlund 2008: 234). A significant virtue of epistemic democracy is its rigorous defence of this conditional knowledge.[6] However, a significant defect of epistemic democracy is whether this conditional knowledge ever translates into practice. Perhaps real democratic states never perform well because the antecedents of epistemic democratic conditionals are never met in practice. So, to know the actual quality of democratic mechanisms, epistemic democrats should also provide antecedent knowledge: enough people actually do meet their antecedents. Epistemic democrats need the ability to evaluate the quality of the actual inputs – the actual votes and voices – put into actual elections, referendums and deliberations to know if their theorems ever translate into practice. Otherwise, epistemic democracy can never become a practical ideal for evaluating the reliability of real democratic states.

Unfortunately, it is very difficult to know if and when the antecedents are met. As explored next, epistemic democrats lack independent access to the quality of the actual inputs from the standpoint of justice. It seems epistemic democrats can only gain access to whether the actual inputs are *good enough* from the standpoint of justice in light of the outputs – the resulting decisions.

Estlund already accepts that Condorcetian democrats cannot easily evaluate the Condorcetian competence of actual people unless Condorcetian democrats already know which decisions are correct. It is very difficult for them to evaluate if actual people perform better than random and choose the correct decision more often than not. In Estlund's words, 'the problem raised by the .5 threshold [Condorcetian competence], though, is not that this is higher than the actual average competence. It is rather that *we don't know whether it is or not* . . . the problem is that even if this is so [and people are Condorcetian competent], it seems impossible to establish publicly without independent access to the truth' [emphasis in original] (Estlund 1993: 93). Epistemic democrats – whether they defend the Condorcet Jury theorem, the 'Diversity Trumps Ability' theorem, or something else – can and do revise their assumptions in a variety of ways (Austen-Smith and Banks 1996; Dietrich and Spiekermann 2013; Goodin and Spiekermann 2018: 24–36; Benson 2021). Nevertheless, whatever the preferred revisions might be, the evaluation problem remains. As explored next, epistemic democrats cannot easily evaluate if *any* actual inputs meet *any* revised assumptions – and consequently the actual quality of the democratic process from the standpoint of justice – *unless* they already know which democratic decisions are correct.

[6]For a critical discussion of the 'Diversity Trumps Ability' theorem, see Thompson (2014). Unfortunately, this controversy exceeds the scope of this paper.

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

Epistemic democrats cannot easily evaluate if and when the actual inputs are *good enough*. They must evaluate whether the inputs that are good enough in theory are, in fact, good enough in practice. For example, many voters are largely sincere, competent and independent, but many are not. An easy way for Condorcetian democrats to evaluate if *enough* voters are largely sincere, competent and independent is to evaluate the resulting decision. Condorcetian democrats can reasonably infer that enough voters were largely sincere, competent and independent if the resulting decision was indeed correct. However, as explored next, Condorcetian democrats lack an easy way to evaluate if enough voters were largely sincere, competent and independent because they lack independent access to the correctness of the resulting decision.

It is useful to explore William Berger and Adam Sales's economic strategy for evaluating voter competence in some detail. They give an economic analysis of elections to give empirical evidence for the epistemic benefits of democracy. In Berger and Sales's words, 'we offer a statistical model to provide evidence that voters' beliefs about economic performance actually result in them voting for the better candidate in regard to the economy's performance' (Berger and Sales 2019: 24). More specifically, they apply a regression model to US National Election Survey data to show that, in US presidential elections since 1980, voters choose the growth candidate more often than not. As Berger and Sales explain, 'were voters to assess economic progress merely at random, the model indicates that the popular presidential vote would have flipped at least five elections' (Berger and Sales 2019: 29). They then use this narrow economic hypothesis that voters typically choose the growth candidate as evidence for the broader epistemic democratic hypothesis that voters are competent.

Berger and Sales say that the growth candidate is a good proxy for the correct candidate because there is a virtual consensus among voters that they want economic growth (Berger and Sales 2019: 24). So, they use *sociotropic economic* voting as a proxy for competent voting (Berger and Sales 2019: 27). In other words, competent voters aim to promote economic growth rather than some non-economic social good or some personal economic good. Berger and Sales argue that voters know if they are 'tightening their belts' (Berger and Sales 2019: 30). This is largely because voters have easy epistemic access to changes in real disposable income (RDI), and changes in RDI also correlate with changes in gross domestic product and inversely correlate with changes in unemployment. Consequently, the reason why sociotropic economic voters typically choose the growth candidate is that changes in RDI either directly or indirectly shape how voters see the economy (Berger and Sales 2019: 28–29). In Berger and Sales's words, 'voters who believe that the economy has grown or will grow are more likely to vote for the incumbent party' (Berger and Sales 2019: 31). How voters see the economy typically steers them to vote for the correct candidate for economic growth.

Unfortunately, Berger and Sales's rigorous empirical analysis of US presidential elections does not avoid the evaluation problem. Even if Berger and Sales do successfully show that voters typically choose the growth candidate, voters still might not be competent. The *growth* candidate is not identical to the *correct* candidate from the standpoint of justice. First, the *growth* candidate is not always the *correct* candidate. In specific elections, other considerations swamp

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

considerations about growth. Second, when the growth candidate *is* the correct candidate, it is not always *because* they are the growth candidate. They are the correct candidate for other reasons. So, when the growth candidate *is not* the correct candidate, competent sociotropic economic voting may not count as competent voting. In specific elections, competent voters know to avoid sociotropic economic voting as other considerations swamp growth in this election. Similarly, when the growth candidate *is* the correct candidate *but for other reasons*, competent sociotropic economic voting still might not count as competent voting. In specific elections, sociotropic economic voters and competent voters may vote the same way but for different reasons, making sociotropic economic voters count as correct but incompetent voters. With independent access to the growth candidate, Berger and Sales can evaluate if and when sociotropic economic voters are competent sociotropic economic voters. However, competent sociotropic economic voting is not identical to competent voting. Without independent access to the correct candidate, they lack an easy way to infer if and when competent sociotropic economic voting counts as competent voting.

Perhaps epistemic democrats can argue that deliberations put good inputs into elections. As John Dryzek and Christian List say, 'the role of deliberation is to bring about situations in which the antecedents of these 'if-then' results are satisfied' (Dryzek and List 2003: 28). Nevertheless, they merely push the evaluation problem back a step. Epistemic democrats should accept that they cannot easily evaluate the actual inputs into deliberations precisely enough (and consequently the actual inputs into elections) unless they already know which decisions are correct from the standpoint of justice. In reliable deliberations, diverse heuristics discover Trojan horses. However, in unreliable deliberations, diverse heuristics do not discover what I will call 'resistant' Trojan horses.[7] If ineffective heuristics dominate deliberations, Trojan horses resist the best heuristics, and they survive the deliberative process. Trojan horses are not always obvious and epistemic democrats lack an easy way to evaluate if and when people are cognitively diverse enough and deliberate effectively enough to know which conceptions of justice are Trojan horses.

Similarly, epistemic democrats cannot easily evaluate if and when the bad inputs overpower the good inputs. In practice, elections typically contain good inputs and bad inputs. So, epistemic democrats must evaluate whether the bad inputs overpower the good ones. For example, not everyone is either insincere, incompetent or dependent, but many are. An easy way for Condorcetian democrats to evaluate if too many voters were either too insincere, incompetent or dependent in practice is to evaluate the resulting decision. Condorcetian democrats can reasonably infer that too many voters were too insincere, incompetent or dependent if the resulting decision was indeed incorrect from the standpoint of justice. However, Condorcetian democrats lack independent access to the correctness of the resulting decision. So, Condorcetian democrats lack an easy way to evaluate if too many voters were too insincere, incompetent or dependent.

[7]Political extremism is potentially highly resistant since the content of extremist beliefs can undergo significant change while their basic epistemic structures remain largely unchanged (Toole 2021).

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

In counterproductive elections, cynical people are competent enough to popularize Trojan horses and sincere people are not competent enough to marginalize them. Following Robert Goodin and Kai Spiekermann's plausible analysis of the 2016 Brexit referendum, the referendum might have chosen Brexit incorrectly from the standpoint of justice because of campaign lies (Goodin and Spiekermann 2018: 325–330). Deceptive heuristics can spread what I will call 'infectious' Trojan horses. A few cynical people advancing narrow interests or partisan ideologies could popularize deceptive heuristics that misleadingly highlight the apparent good of incorrect decisions and downplay, distort or distract from the extensive bads of those decisions. As Vote Leave misleadingly said, why not spend the European Union membership fee on the National Health Service (Goodin and Spiekermann 2018: 326)? Similarly, a few cynical people may spread deceptive heuristics that misleadingly highlight the apparent bad of correct decisions and downplay the extensive good of those decisions. For example, Vote Leave misleadingly inflated the cost of the UK's EU membership and neglected the rebate, EU investment into the UK and the economic benefits of frictionless trade among EU member states. When deceptive heuristics dominate deliberations, they can popularize infectious Trojan horses and spread them throughout the deliberative process. Trojan horses can harmfully interfere with Condorcetian competence. Resistant Trojan horses survive the ineffective heuristics of sincere people and counterproductive heuristics potentially spread infectious Trojan horses. So, Trojan horses potentially deceive enough people into mistaking incorrect decisions for the correct decisions and convert reliable democratic states into unreliable ones.

Contrary to Goodin and Spiekermann's plausible analysis, it nevertheless remains a live possibility that the Brexit referendum result was largely independent of such lies. Perhaps those lies were actually less powerful than they seem or maybe the 'pushbacks' were actually more powerful than they seem. Unless Goodin and Spiekermann know which decision was correct in advance, they cannot evaluate if those lies did *enough* to overpower the epistemic benefits of competent voting or diverse deliberation. The voters might have chosen correctly despite the lies. Alternatively, it is even possible that the voters chose correctly *because of* the lies. Two wrongs may make a right if *cynical lies* deceive *sincerely mistaken people* into voting for the correct decision. The wrong of the lies possibly counteracted the bad of voter incompetence in order for the correct decision to win after all. Of course, if most experts agree that particular campaign promises were highly misleading and if polling shows that many voters believed them, we have a good reason to doubt that the referendum was reliable. However, at best, we only have a good *pro tanto* reason rather than a good *all-things-considered* reason. Many voters believing known lies does reduce the reliability of the referendum, all else being equal. Nevertheless, whether such lies did *enough* to convert a reliable referendum that would have chosen correctly into an unreliable referendum that chose incorrectly remains an open question. Without independent access to which decision was correct, it is not easy to evaluate whether the campaign lies tipped the balance or not. So, the formal solution remains defective. Whatever the preferred analysis might be, it seems epistemic democrats simply cannot evaluate the quality of democratic mechanisms regarding their ability to track the truths of justice *precisely enough* by only evaluating the actual inputs and without evaluating the actual outputs.

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

6. Outcomes-First Solutions

6.1 The Absolute Solution

Both philosophers of science and epistemic democrats can deploy what I will call 'outcomes-first' solutions. A second type of solution to the fundamental problem of measurement is to say that scientists can know the variable's value after all. Scientists can gain independent access to the variable's value at fixed points. Once scientists know that the measurement procedure measures the variable's value accurately at fixed points, they can infer that the measurement procedure is generally accurate. Following Elgin, Reiss calls this the absolute solution (Reiss 2008: 65). In practice, chemists can know the temperature of the water after all (Chang 2004: 8–56). Chemists can gain independent access to the temperature of the water at fixed points. Chemists know the temperature of the water at boiling point is 100° C. Once chemists know the thermometer measures the temperature of the water accurately at boiling point, they can infer that the thermometer is generally accurate.

Unfortunately, the absolute solution has significant defects. Firstly, scientists cannot gain independent access to the variable's value at fixed points. Chemists do not know that the temperature of the water at boiling point is 100°C. In practice, impurities in the water are not always easy to detect and fully extract. So, impurities always risk making the boiling point above 100°C. Consequently, scientists cannot know that the measurement procedure measures the variable's value accurately at fixed points. In practice, impurities in the water can always mask the inaccuracy of the thermometer.

Secondly, even if the measurement procedure does measure the variable's value accurately at fixed points, scientists still should not infer that the measurement procedure is generally accurate. A generally inaccurate measurement procedure could still accurately measure the variable's value at fixed points. In practice, impurities in the mercury can make it expand too quickly, slowly or inconsistently for the thermometer to measure anything except boiling point accurately. Following Elgin, Chang and Reiss, the absolute solution should be rejected. Scientists cannot know the variable's value in advance. As explored next, similar problems spill into 'outcomes-first' solutions for epistemic democrats.

6.2 The Primary Bads Solution

Epistemic democrats can defend a similar 'outcomes-first' solution. A second type of solution to the fundamental problem of evaluation is to say that epistemic democrats can know the correctness of democratic decisions from the standpoint of justice after all. Estlund says substantive epistemic accounts evaluate democratic decisions in light of a substantive procedure-independent standard of correctness (Estlund 2008: 169–171). More generally, people in politics frequently evaluate the result independently of the process. People agree or disagree with the result regardless of how honest and fair the election or referendum was, or what problems the democratic process contained.

In a more modest direction, Estlund argues that epistemic democrats can gain independent access to the correctness of *particular* decisions from the standpoint of justice. In Estlund's words, 'there is, I argue, independent access to some of the

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

content of justice, namely, the primary bads' (Estlund 2008: 170). Estlund defines primary bads as the worst events that everyone should wish to avoid, including war, famine, economic collapse, political collapse, epidemic and genocide. Unable to evaluate the quality of democratic mechanisms regarding their ability to track the truths of justice without evaluating the actual outputs, Estlund defends a substantive standard of primary bads that epistemic democrats can use to evaluate particular decisions. Once epistemic democrats know that elections avoid primary bads, they can infer that elections are generally reliable. Estlund says, 'good performance with respect to primary bads is taken as support for thinking the same procedure would perform well on other matters' (Estlund 2008: 170–171). So, following Estlund, I will call this the primary bads solution.

Unfortunately, the primary bads solution also has significant defects. As explored next, whether democratic decisions avoid primary bads is typically deeply disputed. So, epistemic democrats cannot easily gain independent access to the correctness of democratic decisions from the standpoint of justice for particular decisions. First, epistemic democrats do not know which *types of events* count as primary bads. On the one hand, epistemic democrats could accept that very few types of events count as primary bads. I will call this a minimalist conception of primary bads. Perhaps a minimalist conception of primary bads only contains war, genocide and similar types of violence. Other types of events are *bad*, but they should not count as *primary bads*. Epidemics are not morally on par with war or genocide. On the other hand, epistemic democrats could accept that many types of events count as primary bads. I will call this a maximalist conception of primary bads. For example, a maximalist conception of primary bads may contain suicides, overdoses and similar types of deaths (Case and Deaton 2020). Suicides and overdoses are morally on par with epidemics, especially if they contain high death rates.

Second, epistemic democrats do not know which *specific events* count as primary bads. On the one hand, epistemic democrats could accept that very few specific events count as primary bads. I will call this a minimalist application of primary bads. Perhaps a minimalist application judges that only the world wars and similar wars count as primary bads. Other specific wars are *bad*, but they should not count as *primary bads*. For example, the 2001 USA-Afghanistan War and the 2003 USA–Iraq War are not morally on par with the world wars with incomparable death rates. On the other hand, epistemic democrats could accept that many specific events count as primary bads. I will call this a maximalist application of primary bads. Even if they are not morally on par with the world wars, the Afghanistan War, Iraq War and other specific wars still might have lacked the justification of self-defence, humanitarian intervention or some comparable justification for war and may count as primary bads (Estlund 2008: 163).

The problem with maximalist standards – using maximalist conceptions and applications of primary bads – is that they risk mistaking reliable democratic states for unreliable ones. They mistake lesser bads that reliable democratic states need not avoid for primary bads that reliable democratic states must avoid. Conversely, the problem with minimalist standards – using minimalist conceptions and applications of primary bads – is that they risk mistaking unreliable democratic states for reliable ones. They mistake primary bads that reliable democratic states must avoid for lesser bads that reliable democratic states need not avoid. Absent of a general

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

consensus, a critical mass of democratic citizens with more minimalist standards will judge that the accepted standard includes too much and is too demanding to use. Alternatively, a critical mass of democratic citizens with more maximalist standards will judge that the accepted standard excludes too much and is too undemanding to use. Either way, epistemic democrats risk either overestimating or underestimating the reliability of democratic states for a critical mass of citizens. Without independent access to the correctness of democratic decisions regarding primary bads, epistemic democrats lack an easy way to evaluate the reliability of real democratic states.

A 'do both' strategy might be to assign *two* scores to democracy – one based on a maximalist standard and another on a minimalist standard – to see whether democracy is competent on either measure. However, this strategy has a regress problem. Whichever maximalist standard might be chosen, more maximalist-minded citizens will judge that it still excludes too much and remains too undemanding to use. Alternatively, more minimalist-minded citizens will judge that it still includes too much and remains too demanding to use. So, a 'do both' strategy causes a regress. Epistemic democrats should 'do both' again and assign *another two* scores to democracy – one based on a maximal-maximalist standard and another on a minimal-maximalist standard – to see whether democracy is competent on either measure. The same problem applies to whichever minimalist standard might be chosen. Consequently, epistemic democrats cannot easily assign scores to democracy based on either a maximalist standard or a minimalist standard. Any chosen standard is probably too deeply disputed as too demanding or as too undemanding to use to begin with.

Even if we generously grant that real democratic states do produce correct decisions regarding primary bads, epistemic democrats still should not assume that real democratic states are generally reliable. Epistemic democrats might merely presume that reliability regarding primary bads is good evidence of general reliability.[8] However, the burden of proof is on epistemic democrats to *show* that reliability regarding primary bads *is* good evidence of general reliability. Perhaps people are generally unwilling to vote and deliberate competently, but they are competent enough to avoid primary bads. After all, people should expect to lose much with primary bads and comparably less with lesser bads. So, it is plausible to presume that if anything will motivate democratic citizens and their elected representatives to acquire some level of competence, then avoiding war, famine, and economic collapse will. Consequently, epistemic democrats should not infer that democratic mechanisms are generally reliable from their observation that they are reliable regarding primary bads. Without independent access to the correctness of democratic decisions beyond primary bads, epistemic democrats lack an easy way to evaluate whether being reliable regarding primary bads translates into being generally reliable.

[8]A similar strategy is to show that democracy answers empirically verifiable questions correctly and then infer that democracy probably answers questions about justice correctly. However, even if experts agree on the facts, they still frequently differ in their values. So, epistocrats should not infer moral reliability from factual reliability. Even if all experts get the facts right, at best, only a few of them get the values right, and, at worst, they all get the values wrong. Similarly, it remains deeply controversial for epistemic democrats to infer moral reliability from factual reliability.

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

A very modest epistemic democrat might support democracy because democracy is reliable regarding primary bads, without inferring that democracy is, therefore, generally reliable. However, this very modest type of epistemic democrat is both too modest and not modest enough. First, she is too modest because she significantly weakens the typical epistemic justification for democracy. If democracy chooses correctly regarding primary bads but chooses incorrectly regarding everything else, why not prefer epistocracy? This potentially huge competence gap between democracy and epistocracy may swamp any non-epistemic reasons to support democracy and oppose epistocracy. So, epistemic democrats should argue that democracy is more than merely reliable regarding primary bads to close the potentially huge competence gap between democracy and epistocracy. Second, she is still not modest enough to avoid the evaluation problem. The very modest epistemic democrat must still evaluate whether democracy *is* reliable regarding primary bads. However, as explored above, whether democracy is reliable regarding primary bads remains much harder to evaluate than typically thought because of the regress problem.

Epistemic democrats cannot easily evaluate the quality of democratic mechanisms regarding their ability to track the truths of justice in light of the outputs – the resulting decisions. So, the primary bads solution remains defective. It seems epistemic democrats cannot know the correctness of particular decisions regarding primary bads in advance nor should they infer general reliability from a reliability regarding primary bads anyway.

7. Coherentist Solutions

7.1 The Equilibrium Solution

Both philosophers of science and epistemic democrats can deploy what I will call 'coherentist' solutions. A third type of solution to the fundamental problem of measurement is to concede that scientists cannot know the variable's value in advance or the procedure's accuracy in advance. Scientists lack independent access to the procedure's accuracy (the arbitrary solution fails) and the variable's value (the absolute solution fails). So, scientists should concede that their access to the variable's value is dependent on fallible scientific knowledge about the procedure's accuracy. Similarly, their access to the procedure's accuracy is dependent on fallible scientific knowledge about the variable's value. As a result, scientists should pursue a reflective equilibrium between fallible scientific knowledge about the variable's value and fallible scientific knowledge about the procedure's accuracy. Following Elgin, Reiss calls this the equilibrium solution and prefers the equilibrium solution in light of the significant defects of the arbitrary and absolute solutions (Reiss 2008: 65).

In the history of chemistry, chemists pursued a reflective equilibrium between fallible scientific knowledge about the temperature of the water and fallible scientific knowledge about the accuracy of the thermometer. Chang calls this scientific process 'epistemic iteration' (Chang 2004: 220–234). In Chang's words, 'based on the initially affirmed system [of knowledge] we launch inquiries that result in the refinement and even correction of the original system. It is this self-correcting progress that justifies (retrospectively) successful courses of development in science'

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

(Chang 2004: 6). In general terms, the first step is to revise fallible scientific knowledge about the water's temperature with fallible scientific knowledge about the thermometer's accuracy. The second step is to revise fallible scientific knowledge about the thermometer's accuracy with fallible scientific knowledge about the water's temperature. The third step is to repeat the process until an equilibrium is reached.

It is useful to outline a brief history of epistemic iteration for temperature (Chang 2004: 47–48). In the first stage, scientists used the bodily sensation of hot and cold to judge the temperature of water. In the second stage, scientists used the observed correlation between sensations of hot and cold and changes in fluid volume to invent thermoscopes, using the changes in fluid volume to observe the changes in temperature more accurately. At the beginning of the third stage, scientists used boiling point as a fixed point to invent a numerical scale to quantify temperature. The invention of thermometers allowed for mathematical theorizing about temperature. This shows that scientists revised their explicit judgements about the water's temperature with their implicit judgement that the thermometer is accurate. At the end of the third stage, scientists observed that the boiling point was not as fixed as previously thought. So, they swapped the boiling point of water for the steaming point of water to gain a more stable fixed point and construct a more accurate thermometer. This shows scientists also revised their judgements about the thermometer's accuracy with their judgements about the temperature of the water at boiling point and at steaming point.

Of course, scientists use judgements about correspondence to guide the process of epistemic iteration. In other words, they select measurements they judge are likely to measure the true temperature of the water and repeatedly refine them to avoid various potential errors. Nevertheless, scientists cannot easily evaluate how successful their judgements about correspondence really are. They cannot easily evaluate how well their selected measurements really do represent the true temperature of the water or how successfully their repeated refinements really do avoid various potential errors. So, scientists cannot easily evaluate how well their refined judgements really do correspond with the true temperature of the water. For example, they previously judged that boiling-point-based thermometers corresponded well, but they later judged that they corresponded poorly and that steaming-point-based thermometers corresponded better. Consequently, rather than judge how successful their judgements might be in terms of *correspondence*, Chang argues that scientists should judge success in terms of *coherence*. Scientists need not judge whether their best efforts to track the truth and avoid error succeed in making their judgements about the water's temperature *correspond* better with the true temperature of the water. More modestly, they need only judge that their best efforts to track the truth and avoid error succeed in making their judgements about the water's temperature more *coherent* with their reflective judgements regarding the thermometer's accuracy.

The equilibrium solution is an imperfect but attractive solution. Suppose that scientific knowledge becomes refined enough for scientists to judge that the measurement procedure typically measures the variable's value accurately. The measurement procedure is very useful because it does what scientists wish it to do. A thermometer that chemists typically judge to measure the temperature of the water

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

accurately does what chemists wish it to do. Consequently, Reiss argues that whether the procedure measures the variable's *true* value eventually becomes an idle rather than an active problem. As explored next, similar 'coherentist' solutions are available to epistemic democrats.

7.2 The Expert Evaluation Solution

Epistemic democrats can defend a similar 'coherentist' solution. A third type of solution to the fundamental problem of evaluation is to concede that judgements about the reliability of real democratic states are dependent on how experts in research institutions evaluate them. As Estlund and Landemore say, 'by 'correct or right decision' here, or the 'truth' can be meant an array of things, from objective truth of the matter (about facts or morality) to a more intersubjective, culturally-dependent and temporary construct (about more socially constructed facts or moral questions)' (Estlund and Landemore 2018: 13). In particular, epistemic democrats can evaluate whether their judgements about the reliability of democracy are *coherent* with expert research on the inputs into the democratic process and expert research on the outcomes of democratic decisions. Epistemic democrats should modestly pursue a reflective equilibrium between expert research on the quality of the democratic process and the correctness of democratic decisions. In general terms, the first step is to revise epistemic democratic judgements about which decisions are correct in light of expert research on the inputs into the democratic process. The second step is to revise epistemic democratic judgements about the quality of the democratic process in light of expert research on the outcomes of democratic decisions. The third step is to repeat the process until an equilibrium is reached. I will call this the expert evaluation solution.

It is useful to outline a brief hypothetical example.[9] Epistemic democrats may start with the reported facts and their preferred standard of correctness to judge that the latest war was correct or at least that it was not a primary bad (alongside a number of other significant outcomes of the last election). Epistemic democrats can then use expert research to develop a more accurate judgement about the reported facts and use public deliberation to develop a more refined standard of correctness. They revise their explicit judgements about the correctness of the latest war with their implicit judgement that the expert research is accurate and that the public deliberations were reasonable. However, experts may later show that their previous research was not accurate. The research is now known to have omitted significant facts or to have contained significant falsehoods. Alternatively, the public deliberations were not as reasonable as previously thought. Intense feelings of fear or anger dominated them. So, epistemic democrats later rely on more informed research and less fearful deliberations to develop a more accurate and reasonable judgement about the correctness of the latest war.

Similarly, epistemic democrats may start with the reported facts and their preferred theorem to judge that the last election was reliable. They then refine their explicit judgements about the reliability of the last election with their implicit judgement that

[9]I avoid actual cases to avoid disputed descriptions of divisive decisions that may complicate rather than clarify the deeper theoretical claim under consideration.

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

the expert research on various inputs is accurate. However, experts may later show that their previous research was not accurate. So, they later rely on new research to develop a more accurate judgement about the reliability of the last election. If epistemic democrats judge that the last election was reliable but the latest war was a primary bad, they can reevaluate whether the latest war really was a primary bad. Alternatively, they may reevaluate whether the last election really was reliable. Similarly, if epistemic democrats judge that the latest war was correct but the last election was unreliable, they might reevaluate whether the latest war really was correct. Alternatively, they could reevaluate whether the last election really was unreliable.

Epistemic democrats may use judgements about correspondence to guide their process of reflective equilibrium. In other words, they select expert research they judge is likely to represent the true reliability of democracy. Nevertheless, epistemic democrats cannot easily evaluate how successful their judgements about correspondence really are. They cannot easily evaluate how well the expert research they selected *really does* represent the true reliability of democracy. So, epistemic democrats cannot easily evaluate how well their reflective judgements *really do* correspond with the true reliability of democracy. For example, they may use expert research to judge that the last election was reliable, but later expert research shows that the latest war was not correct, and so the last election was significantly less reliable than previously thought. Consequently, rather than judge how successful their judgements might be in terms of *correspondence*, epistemic democrats should judge success in terms of *coherence*. Epistemic democrats need not judge whether their best efforts to track the truth succeed in making their judgements about the reliability of a real democratic state *correspond* more closely with its true reliability. More modestly, they need only judge that their best efforts to track the truth succeed in making their judgements about the reliability of a real democratic state more *coherent* with their reflective judgements regarding the inputs and outputs of the democratic process.

Without independent access to the quality of the inputs into democratic mechanisms, the measurement problem in the philosophy of science allows political philosophers to see that coherentist evaluations of democratic reliability are more attractive than the dominant 'procedure-first' evaluations. Nevertheless, whether coherentist evaluations of democratic reliability are attractive *all things considered* remains an open question and warrants much more work. As explored next, coherentist evaluations are much more difficult to deploy in politics than in science because of the need for public justification and the potential for such justifications to interact with the political behaviour of democratic citizens.

It is plausible to presume that many sincere people will be reluctant to vote and deliberate competently if they believe that the democratic state does not perform well (Cohen 1986: 37). Whatever complex intrinsic and extrinsic motivations to vote and deliberate competently there might be, if voting and deliberating competently is seen as ineffective, then many sincere people will not do it: dropping a few competent votes and voices into a sea of incompetent votes and voices will do little good. So, the expert evaluation solution should publicly justify the reliability of expert research in order to reassure sincere people that voting and deliberating competently does indeed contribute towards a generally reliable democratic process. Put simply, public justification is needed for motivation.

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

Unfortunately, expert research on the reliability of real democratic states is very difficult to publicly justify, partially because it is a case of what philosopher of science Ian Hacking calls interactive knowledge (Hacking 1999: 32). Interactive knowledge does not merely describe the characteristics of the subject of study, but the description itself interacts with and actively changes the characteristics of the subject of study. In other words, if the expert research is publicly justified, it can motivate more sincere people to put competent votes into the democratic process and actively improve the performance of that very democratic state in return. However, if it is not publicly justified, more sincere people may become reluctant to put competent votes into the democratic process, actively reducing the performance of that very democratic state. So, in politics, the true reliability of real democratic states remains a persistently active problem because it constantly interacts with the subject of study in a way that the true accuracy of a thermometer in science does not.

It is useful to explore Goodin and Estlund's Condorcetian coherentist solution for evaluating the reliability of real democratic states. They rely on the Condorcet Jury theorem as a proven mathematical truism to infer voter competence from the size and character of the democratic majority (Goodin and Estlund 2004: 133–135). Goodin and Estlund assume a principle of charity that presumes people are competent until proven otherwise (Goodin and Estlund 2004: 136). However, the Condorcet Jury theorem shows that competence is ineffective if people are dependent or insincere (Goodin and Estlund 2004: 137–138). First, Goodin and Estlund judge that the independence of supermajorities is suspicious. For example, a 98% supermajority signals that the average voter competence is either 0.98 or 0.02. So, dependence probably explains the size of a supermajority rather than an implausibly high or implausibly low average voter competence. Second, Goodin and Estlund judge that the sincerity of majorities is suspicious if votes strongly correlate with interests. For example, if most white voters voted one way and black voters the other, votes probably tracked race rather than the truth. So, insincerity probably explains the character of the majority. Once experts prove that the size of the majority does not signal dependence and its character does not signal insincerity, they can infer that the majority was probably competent and the decision was probably correct from the standpoint of justice.

Unfortunately, this clever Condorcetian coherentist solution remains too imprecise for experts to publicly justify the reliability of real democratic states. First, opinion leaders probably do harm voter independence to *some* degree. However, as with the formal solution, it remains very difficult for experts to evaluate if opinion leaders harm elections to a *significant* degree without independent access to the correctness of the resulting decision. For instance, perhaps a critical mass of voters did sheepishly follow opinion leaders in previous elections. Dependence was global enough to convert an otherwise independent and subsequently correct majority into a dependent and subsequently incorrect majority. Alternatively, maybe only an insignificant few truly followed opinion leaders and made no significant difference to the correctness of the resulting decision. Second, a complex web of entangled interests is typically present in high-stakes political questions, but it remains very difficult for experts to evaluate if narrow interests harm voter sincerity in elections to a significant degree. For example, maybe narrow interests did dominate enough voters in previous elections to convert an otherwise sincere

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

and subsequently correct majority into an insincere and subsequently incorrect majority. Alternatively, perhaps insincerity was too local to be significant. Whatever the preferred analysis might be, this clever Condorcetian coherentist solution may discover the *presence* of dependence or insincerity in previous elections, but it is too imprecise to show its *significance* for the overall quality of the election and the overall correctness of the resulting decision.

Third, experts should not presume that laypeople are competent *in politics*. They should expect that generally competent laypeople still have specific incompetencies, especially if a specific competency is very difficult to develop. So, however plausible a general principle of charity might be, it is not plausible to presume that laypeople are competent in politics any more than it is plausible to presume that laypeople are competent in computer science, mechanical engineering or advanced mathematics. Experts should publicly justify if and when laypeople are politically competent rather than merely presume it.

Whatever the coherentist solution might be, it remains very difficult for epistemic democrats to publicly justify expert research on democratic reliability. Perhaps the credentials of the experts provide good indirect evidence that expert research is reliable. Expert research really does track the true reliability of real democratic states. It is possible that very coarse judgements are put into expert research, potentially based on hopelessly incomplete information or grossly reckless reasoning. So, the credentials of the experts may reassure the public that expert research is refined enough to count as reliable research.

Unfortunately, this is easier said than done. First, suppose many sincere people believe that the experts in research institutions do the best anyone can, but that expert research remains incapable of evaluating the reliability of real democratic states well. The most credentialed experts frequently remain too unrefined and still get it significantly wrong. With many more variables and value judgements at play, good political judgements about the reliability of democratic states are exponentially more difficult than good scientific judgements about the accuracy of a thermometer. Second, suppose many sincere people believe that the experts in research institutions become motivated to misevaluate the democratic state. They become motivated to provide unreliable evaluations. Independently of external interests corrupting expert research, the institutional incentives to which the experts must react in order to survive and succeed professionally potentially induce a willingness in them to confirm rather than correct subtle institutional biases (especially if they prioritize their private interests in career progression over the public interest in reliable research). So, many sincere people potentially believe that institutionally biased experts often remain too unrefined and get it significantly wrong. Worse, external interests potentially do corrupt expert research. Cynical experts – or sincere experts in cynical institutions – may deceive too many laypeople into mistaking incorrect decisions that advance narrow interests or partisan ideologies for the correct decisions.

Unless expert evaluations of real democratic states are publicly justified, a critical mass of sincere but reluctant people are inclined to put incompetent votes into otherwise good elections and produce bad decisions as a result. Unfortunately, it remains very difficult for epistemic democrats to evaluate the quality of democratic mechanisms or the correctness of democratic decisions from the standpoint of

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

justice in light of expert research. Without independent access to when democratic mechanisms are truly reliable or to which democratic decisions are truly correct, it remains very difficult to publicly justify expert evaluations of the reliability of real democratic states.

In order to avoid the evaluation problem, political philosophers might be tempted to retreat from competence and return to prioritizing some conception of fairness. For example, relational egalitarians ground legitimacy in creating a community in which citizens stand in relations of equality with each other (Anderson 1999; Kolodny 2014; Viehoff 2014). Perhaps relational equality is enough for a *legitimate* democracy. Nevertheless, relational equality is not necessarily enough for a *competent* democracy. A relational egalitarian democracy may not perform as well as an epistocracy or a fair coin flip. Worse, a relational egalitarian democracy might become complicit in multiple primary bads, from unjust wars to easily avoidable famines, epidemics, and genocides, as well as collapsing various economies and governments. In that case, its catastrophic incompetence from the standpoint of justice should call its acceptability into question. So, independently of whether a relational egalitarian democracy is *legitimate*, relational egalitarians should also want to evaluate the *competence* of a relational egalitarian democracy.[10] As a result, the evaluation of competence is not easy for non-epistemic democrats to avoid, whatever their preferred conception of legitimacy might be.

8. Conclusion

The core contribution of this paper is to promisingly cross-fertilize political philosophy with the fundamental problem of measurement in the philosophy of science to show that epistemic democrats should significantly revise how to evaluate democracy. Chemists cannot evaluate the temperature of water if they lack independent access to the accuracy of the thermometer. However, chemists also cannot evaluate the accuracy of the thermometer if they lack independent access to the temperature of the water. By analogy, epistemic democrats also cannot easily evaluate the correctness of democratic decisions from the standpoint of justice if they lack independent access to the quality of the democratic process. However, it is also very difficult for epistemic democrats to evaluate the quality of the democratic process if they lack independent access to the correctness of democratic decisions.

An arbitrary solution to the fundamental problem of measurement says chemists can decide that the thermometer *defines* the temperature of the water by fiat. However, an arbitrary solution fails because empirical research rather than *a priori* definitions should determine the accuracy of thermometers. Similarly, a formal solution to the fundamental problem of evaluation says access to the quality of democratic mechanisms regarding their ability to track the truths of justice can depend on evaluating their inputs – the votes and voices – rather than their outputs – the resulting decisions. However, a formal solution is defective.

[10]If or when a highly incompetent democracy may become an illegitimate democracy exceeds the scope of this paper.

https://doi.org/10.1017/S0266267125100503 Published online by Cambridge University Press

Without independent access to the correctness of the resulting decisions, epistemic democrats lack an easy way to evaluate the quality of the votes and voices put into the democratic process precisely enough.

An absolute solution to the fundamental problem of measurement says chemists can know the temperature of the water at fixed points. However, an absolute solution fails since chemists cannot know the temperature of the water at fixed points and should not infer general accuracy from an accuracy at fixed points anyway. Similarly, a primary bads solution to the evaluation problem says epistemic democrats can know the correctness of particular decisions from the standpoint of justice in advance. However, a primary bads solution is defective. Epistemic democrats cannot easily know the correctness of particular decisions in advance and they should not infer that the democratic process is generally reliable from its mere avoidance of primary bads anyway. The central virtue of epistemic democracy is that the epistemic democratic evaluations of democratic mechanisms are seemingly more modest than moral evaluations of the resulting decisions. However, the evaluation problem questions whether the epistemic democratic evaluations of democratic mechanisms really are more modest than moral evaluations of the resulting decisions. So, epistemic democracy risks losing its central virtue.

An equilibrium solution to the fundamental problem of measurement says chemists can persistently revise their fallible knowledge about the temperature of the water and the accuracy of the thermometer. Similarly, an expert evaluation solution to the evaluation problem says epistemic democrats can persistently revise their evaluations of the correctness of democratic decisions and the quality of the democratic process from the standpoint of justice in light of expert research. Nevertheless, the expert evaluation solution is more complex than the equilibrium solution. Much more work is needed on how epistemic democrats can publicly justify expert research on the reliability of real democratic states in light of the evaluation problem. Otherwise, real democratic states risk a critical mass of sincere people believing that voting and deliberating competently is ineffective, and real democratic states may overproduce incorrect decisions from the standpoint of justice as a result.

Acknowledgements. I thank Professor Julian Reiss, Professor Nancy Cartwright and Dr Elizabeth Kahn for their rigorous feedback and I am grateful for the supportive research community at the Centre for Humanities Engaging Science and Society (CHESS) and on the 'Knowledge For Use' (K4U) research project during my time at Durham University. I am grateful to Professor Alasdair Cochrane, Dr Joshua Forstenzer and Dr Robbie Morgan for their generous feedback during my time at Sheffield University. I thank Professor Sabina Leonelli for her helpful feedback on the final drafts. I thank the two anonymous reviewers for their supportive and substantive comments. Any mistakes remain my responsibility. This research was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [Grant Agreement No. 667526 K4U]. This research was also supported by funding from the ERC under the European Union's Horizon 2020 research and innovation programme [Grant Agreement No. 101001145]. The content reflects only the author's views, and the ERC is not responsible for any use that may be made of the information it contains.

References

Anderson E. 1999. What is the point of equality? *Ethics* **109**, 287–337.
Anderson E. 2006. The epistemology of democracy. *Episteme* **3**, 8–22.

Austen-Smith D. and J. Banks 1996. Information aggregation, rationality, and the Condorcet Jury theorem. *American Political Science Review* **90**, 34–45.

Benson J. 2021. The epistemic value of deliberative democracy: how far can diversity take us? *Synthese* **199**, 8257–8279.

Berger W. and A. Sales 2019. Testing epistemic democracy's claims for majority rule. *Politics, Philosophy & Economics* **19**, 22–35.

Brennan J. 2014. How smart is democracy? You can't answer that question a priori. *Critical Review* **26**, 33–58.

Cartwright N. 2012. Presidential address: will this policy work for you? Predicting effectiveness better: how philosophy helps. Philosophy of Science 79, 973–989.

Cartwright, N. 2021. Rigour versus the need for evidential diversity. *Synthese* **199**, 13095–13119.

Case A. and A. Deaton 2020. *Deaths of Despair and the Future of Capitalism*. Princeton: Princeton University Press.

Cerovac I. 2020. *Epistemic Democracy and Political Legitimacy*. London: Palgrave Macmillan.

Chang H. 2004. *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.

Cohen J. 1986. An epistemic conception of democracy. *Ethics* **97**, 26–38.

Coleman J. and J. Ferejohn 1986. Democracy and social choice. *Ethics* **97**, 6–25.

Dietrich F. and K. Spiekermann 2013. Epistemic democracy with defensible premises. *Economics and Philosophy* **29**, 87–120.

Dryzek J. and C. List 2003. Social choice theory and deliberative democracy: a reconciliation. *British Journal of Political Science* **33**, 1–28.

Elgin C. 1996. *Considered Judgment*. Princeton: Princeton University Press.

Elgin C. 1997. *Between the Absolute and the Arbitrary*. Ithaca: Cornell University Press.

Elster J. 2000. Arguing and bargaining in two constituent assemblies. *University of Pennsylvania Journal of Constitutional Law* **2**, 345–419.

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

Estlund D. 1997. Beyond fairness and deliberation: the epistemic dimension of democratic authority. In *Deliberative Democracy: Essays on Reason and Politics*, eds J. Bohman and W. Rehg, 173–204. Cambridge, MA: MIT Press.

Estlund D. 1993. Making truth safe for democracy. In *The Idea of Democracy*, eds D. Copp, J. Hampton and J.E. Roemer, 71–100. New York: Cambridge University Press.

Estlund D. and H. Landemore 2018. The epistemic value of democratic deliberation. In *Oxford Handbook of Deliberative Democracy*, eds. A. Bächtiger, J. Dryzek, J.J. Mansbridge and M. Warren, 113–131. Oxford: Oxford University Press.

Goodin R. and D. Estlund 2004. The persuasiveness of democratic majorities. *Politics, Philosophy and Economics* **3**, 131–142.

Goodin R.E. and K. Spiekermann 2018. *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.

Hacking I. 1999. *The Social Construction of What?* Cambridge, MA: Harvard University Press.

Kolodny N. 2014. Rule over none II: Social equality and the justification of democracy. *Philosophy & Public Affairs* **42**, 287–336.

Landemore H. 2012. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton: Princeton University Press.

Perote-Peña J. and A. Piggins 2015. A model of deliebrative and aggregative democracy. *Economics and Philosophy* **31**, 93–121.

Peter F. 2013. The procedural epistemic value of deliberation. *Synthese* **190**, 1253–1266.

Reiss J. 2008. *Error in Economics: Towards a More Evidence–Based Methodology*. London: Routledge.

Schwartzberg M. 2015. Epistemic Democracy and its Challenges. *Annual Review of Political Science* **18**, 187–203.

Thompson A. 2014. Does diversity trump ability? An example of the misuse of mathematics in the social sciences. *Notices of the American Mathematical Society* **61**, 1024–1030.

Toole B. 2021. What lies beneath: the epistemic roots of white supremacy. In *Political Epistemology*, eds. E. Edenberg and M. Hannon, 76–96. Oxford: Oxford University Press.
Viehoff D. 2014. Democratic equality and political authority. *Philosophy & Public Affairs* **42**, 337–375.

Richard Beadon Williams is a Postdoctoral Researcher on the ERC-funded 'Philosophy of Open Science' research project in the Department of Science, Technology and Society, Technical University of Munich. His current work focuses on open science and public policy and how open science practices across diverse research environments can and should inform public policy. URL: www.sts.sot.tum.de/sts/personen/forschende/richard-williams/

Cite this article: Williams RB. Does this democracy work? Measuring reliability, Trojan horses and the fundamental problem of evaluation. *Economics and Philosophy*. https://doi.org/10.1017/S0266267125100503