Article

Philosophy of the Social Sciences
2020, Vol. 50(3) 195–214

DOI: 10.1177/0048393120917635
journals.sagepub.com/home/pos

Mathematical Models and Robustness Analysis in Epistemic Democracy: A Systematic Review of Diversity Trumps Ability Theorem Models

Ryota Sakai[1]

Abstract

This article contributes to the revision of the procedure of robustness analysis of mathematical models in epistemic democracy using the systematic review method. It identifies the drawbacks of robustness analysis in epistemic democracy in terms of sample universality and inference from samples with the same results. To exemplify the effectiveness of systematic review, this article conducted a pilot review of diversity trumps ability theorem models, which are mathematical models of deliberation often cited by epistemic democrats. A review of nine models extracted from 352 papers exemplifies the effectiveness of robustness analysis supplemented by systematic review in epistemic democracy.

Keywords

epistemic democracy, robustness analysis, systematic review, diversity trumps ability theorem, deliberative democracy

Received 1 June 2019

[1]Asahi University, Mizuho, Japan

Corresponding Author:
Ryota Sakai, Faculty of Law, Chuo Gakuin University, 451, Kujike, Abiko 270-1196, Chiba, Japan.
Email: sakai.ryota@gmail.com

1. Introduction

Why democracy? Why should judgment by diverse people prevail over experts' judgment? To answer these questions, political philosophers proposed a variety of reasons for democracy including equality and fairness. Epistemic democrats, a group of political philosophers who advocate democracy for its truth-tracking function, use mathematical models to show a reason for democracy. To buttress this argument, some epistemic democrats cite the deductive results of mathematical and computational models, such as the Condorcet jury theorem (CJT) and the diversity trumps ability (DTA) theorem (e.g., Landemore 2013).[1]

Concerns about the internal and external validity of model analysis have been the central criticism of epistemic democracy. For instance, Lu Hong and Scott E. Page's (2004) original DTA model analysis, a mathematical model of deliberation often cited by epistemic democrats, has been criticized for possibly containing defects that ruin its internal validity, such as its biased range of observations, indefensible assumptions, and errors (Grim et al. 2018; Holman et al. 2018; Thompson 2014). Moreover, the original DTA model has drawn a number of criticisms on its external validity (Brennan 2016; Caplan 2012). Meanwhile, researchers in the philosophy of science have proposed that multiple-model analysis accompanied by robustness analysis improves the reliability of model analysis (Weisberg 2013, 174). Recently, groups of researchers have suggested that the application of robustness analysis over mathematical models is an effective method for trustworthy model analysis in epistemic democracy (Klein, Marx and Fischbach 2018).

However, the drawbacks of robustness analysis in epistemic democracy have been overlooked. Although this article endorses the effectiveness of robustness analysis in epistemic democracy's model research, it criticizes two aspects of robustness analysis procedures—the absence of a guideline for universal sample collection and inference from samples with the same results. These potentially induce biased recommendations when robustness analysis is used in epistemic democracy. Meanwhile, the systematic review method has been a popular approach in medicine and the social sciences to alleviate such biased recommendations from the arbitral sample collection and inference (Petticrew and Roberts 2006). Although systematic review has been rarely discussed in the literature on the robustness analysis of mathematical models in epistemic democracy, robustness analysis may more adequately function as a reliable analysis method if it is supplemented by systematic review.

[1]A reviewer's suggestion helped the revision of this paragraph.

This article aimed to partly supplement the procedure of robustness analysis of mathematical models in epistemic democracy with systematic review. The article conducted a pilot review of nine DTA-related models extracted from 352 papers, and found the possibility that in many issues, the deliberation of a mixed group of experts and laypersons was epistemically superior. The review result differed from the recommendation that epistemic democrats made; layperson groups outperformed. This article argues that the result exemplifies the effectiveness of robustness analysis supplemented by systematic review in epistemic democracy. The rest of the paper is organized as follows. Section 2 discusses the drawbacks of robustness analysis in epistemic democracy. It presents the rationale why it failed to produce the trustworthy result that epistemic democrats aimed for. Section 3 introduces the systematic review method as a supplementary tool for robustness analysis. Section 4 conducts a pilot case study for the systematic review of DTA models. Sections 5 and 6 present and discuss the results of the systematic review, and Section 7 concludes.

2. Background

2.1. From Single-Model Analysis to Multiple-Model Analysis

Epistemic democrats have surprisingly placed little emphasis on multiple-model analysis. Political philosophers on epistemic democracy have referred to a single-model analysis to build their philosophical argument. For instance, Hélène Landemore based her philosophical argument mainly on the single-model DTA analysis provided by Hong and Page (Landemore 2013). Some political philosophers have introduced model competition. Elizabeth Anderson supported the Deweyan model of democracy and criticized the CJT and DTA models based on her philosophical argument on ideal democracy (Anderson 2006). Others have sought an integrated model providing a basis for multiple models. Marcus Pivato introduced the mean partition rule model that explains many common voting rules, including CJT and the wisdom of crowds (Pivato 2017).[2]

The present article proposes a new approach to epistemic democracy research, namely, multiple-model analysis, with the following benefits for epistemic democrats.[3] (a) It allows a pragmatic selection of multiple models

[2]For the mathematical definition of wisdom of crowds, see Pivato (2017, 58).

[3]Recently, Page proposed a similar concept called *many model thinking* (Page 2018). However, his proposal does not aim to provide a systematic model analysis methodology or an improvement in epistemic democracy.

to meet the multiple (incompatible) purposes of model analysis, such as theoretical explanation and prediction (Levins 1966, 431); (b) it captures many aspects of a phenomenon, giving researchers a full picture of the phenomenon (Weisberg 2013, 104); and (c) it provides sources of robustness analysis, allowing researchers to ensure the trustworthiness of the model analysis (Wimsatt [1981] 2007, 46).

2.2. A Gift from the Philosophy of Science: Robustness Analysis

The benefit of multiple-model analysis is enhanced by robustness analysis. Robustness analysis is a method of distinction of the core mechanism of mathematical models from the mere artifacts of these models. Here, artifacts are regarded as by-products of the idealization of the modeling process (Wimsatt [1981] 2007, 46). Accordingly, when properly designed, robustness analysis protects us from the possibility that a model analysis may result in an erroneous implication because of such by-products. It can improve the credibility of epistemic arguments for democracy.

2.3. Drawbacks of Robustness Analysis

However, we can identify drawbacks in robustness analysis procedures that induce selection bias. These drawbacks are related to two procedures: (a) the absence of a structured guideline for universal sample collection and (b) the procedure that makes the inference from the samples with the same results.

Michael Weisberg provided the following advice for sample collection: "during this stage, it is important that they [theorists] collect a sufficiently diverse set of models so that the discovery of a robust property does not depend in an arbitrary way on the set of models analyzed" (Weisberg 2013, 158). However, the actual procedure of universal sample collection has seldom been explicitly proposed by robustness analysis proponents. The absence of such guidance has inhibited informing epistemic democrats who have focused on single-model analysis to induce philosophical implications for democracy.

The second procedure has commonly been used by leading robustness analysis scholars (Kuorikoski, Lehtinen, and Marchionni 2010, 545; Levins 1966, 423; Wimsatt [1981] 2007, 44). As a typical case, Weisberg proposed a two-step procedure: "The procedure begins by examining a group of models to determine if they all predict a common result, the robust property. The second step involves analyzing the models for the common structure which

generates the robust property" (Weisberg 2013, 158).[4] Nevertheless, protection from selection bias is not ensured by this two-step procedure. Commencing a robustness analysis with the identification of a common result can violate the general advice on sampling that "selection should allow for the possibility of at least some variation on the dependent variable" (King, Keohane, and Verba 1994, 129). Accordingly, it may induce not only selection bias but also indeterminacy in causal inference.

Epistemic democrats are especially at a high risk of committing selection bias. They aim to justify democracy by citing certain mathematical models suggesting democracy's epistemic success and its favorable conditions. They instantiate mathematical models with favorable conditions for democracy. Take the CJT case, for example. They instantiate the model with, on average, above 50% of correctness among voters together with the independence of their votes. By showing favorable conditions, epistemic democrats argue for democracy and sometimes call for the amendment of political conditions (Estlund 2008; Landemore 2013). However, they overlook the possibility that the same combination of conditions leads to different results. Classical CJT suggests that in above 50% competence and independence conditions, the majority decision by a very large number of voters becomes infallible (Ladha 1992). Yet, with a slightly modified version of assumptions, problem-specific competence, and realistically modified independence assumptions, a very large number of voters with above 50% competence cannot be infallible (Dietrich and Spiekermann 2013).[5] Such negative cases with similar conditions as positive cases can be excluded from the research scope of the orthodox procedure of robustness analysis that makes the inference from the samples with the same results. Accordingly, it sometimes fails to give trustworthy recommendations on epistemic arguments for democracy, as this CJT case suggested.

3. Proposed Method and Systematic Review

3.1. An Amendment to Robustness Analysis: Introduction of Systematic Review

To improve the validity of model analysis, epistemic democrats can use the systematic review procedure for supporting robustness analysis. This article

[4]Weisberg applies the same procedure to assess at least three criteria of robustness: parameter permutations, structure alternations, and alternation of mathematical representations and simulation languages (Weisberg 2013, 160-66).

[5]Although the classical and modified conditions are not identical, the fact that the realistic modification of the classical condition leads to different results is practically important.

makes two proposals for robustness analysis in epistemic democracy: (a) pursuing the actual sampling phase of robustness analysis with a systematic sample collection procedure and (b) initiating the analysis from the identification of differences in results. Regarding (a), robustness analysis with a systematic sample collection minimizes the problem of selection bias. It also helps researchers use a large pool of sample models in the literature. Regarding (b), the proposal alleviates the shortcomings of the underlying logic of inference in robustness analysis that may induce selection bias and indeterminacy in causal inference. To make these recommendations, one can use the systematic review method.

3.2. Systematic Review

Systematic review supplements robustness analysis. Systematic review refers to "a review of the research literature using systematic and explicit, accountable methods" (Gough, Oliver, and Thomas 2012, 2). Systematic review protocols ensure the transparency and neutrality of sample selection. The review process excludes an arbitrary selection of samples and an arbitrary attribution of the causal mechanism. The process synthesizes collected information by the methods which are not limited to the identification of common results.

One may think that systematic review is about meta-analysis, "a statistical method for combining the numerical results of studies" (Gough, Oliver, and Thomas 2012, 205). Yet, a systematic review is not limited to meta-analysis; instead, it is used for two purposes, hypothesis generation and hypothesis testing, which use different methods of synthesis formation. The configuration type of synthesis aims to form "a mosaic, in which the findings from each study are slotted together to form a coherent whole" for hypothesis generation (Gough, Oliver, and Thomas 2012, 51), whereas the aggregation type uses meta-analysis for hypothesis testing. As this article aims to refine hypothesis generation in epistemic democracy, it provides a configuration-type synthesis of systematic reviews on DTA models and provides a full picture of the phenomenon.

4. A Pilot Case Study for the Systematic Review of Mathematical Models: The DTA Model

The DTA model is one of the commonly cited agent-based simulation models on deliberative problem solving. Roughly speaking, agents with different strategies collaboratively climb hills on an epistemic landscape looking for the highest peak, which represents the epistemically superior solution. One of the originators of the DTA model, Page, argued that "this theorem is no mere metaphor or cute empirical anecdote that may or may not be true ten years

from now. It's a logical truth" (Page 2007, 162). Basing on a single-model analysis of the original DTA, political philosophers have claimed the epistemic success of deliberative democracy (Gaus 2016; Landemore 2013). Do these statements still hold?

4.1. Study Design

The systematic literature review protocol was prepared. This protocol guides the development of review questions, search strategies, and inclusion and exclusion criteria.[6]

4.2. Review Questions

The following research questions were established for the systematic review:

Research Question 1: What are the commonly stated definitions of experts and expert exploration patterns?
Research Question 2: What are the commonly stated definitions of layperson and layperson exploration patterns?
Research Question 3: What kinds of deliberation dynamics are assumed?
Research Question 4: Does diversity trump ability?
Research Question 5: What is the best composition of participants for an epistemically superior deliberation?

4.3. Information Sources and Search Strategies

This article analyzed models selected from 352 studies in the Web of Science database that were published from January 2004 to December 2018 and that cited Hong and Page's (2004) original DTA article (searched on January 23, 2019). This constituted forward reference list checking (Gough, Oliver, and Thomas 2012, 126). All models were manually screened by both their titles and their abstracts. In some cases, the contents of the papers were manually checked.

The bibliographical research starting from the original DTA model could potentially be biased in support of the original DTA analysis. However, the result presented below is contrary to the original DTA result, which partly supports the validity of the bibliographical research in this case.

[6]See Gough, Oliver, and Thomas (2012, 8).

4.4. Inclusion and Exclusion Criteria

The inclusion criteria used for this systematic review were studies conducted on the DTA phenomenon, original articles that have been published in peer-reviewed journals, studies conducted with simulation analysis, and studies conducted with deliberation-type dynamics as an information aggregation mechanism.

The majority of the studies were non-simulation research, so they were excluded. Excluded studies were philosophical arguments and commentaries of the model that neither developed a new model nor amended the original DTA model. Importantly, studies with other information collection dynamics such as majority voting were excluded from this systematic review of deliberation models.

This study shortlisted nine DTA-related model analyses from the 352 studies published between 2004 and 2018 that cited Hong and Page's (2004) original DTA article. This study added the epistemic landscape model (ELM) study by Michael Weisberg and Ryan Muldoon (2009), an agent-based simulation model similar to DTA on scientific progress via the collaboration of diverse types of scientists, as a useful comparison case with DTA. Although the original ELM study did not cite Hong and Page's original DTA study, modified versions of the ELM model (e.g., Alexander, Himmelreich, and Thompson 2015; Thoma 2015) often cite the original DTA model study. Eventually, the shortlisted model studies were either DTA or ELM extensions.

5. Results

5.1. Does Diversity Trump Ability?

On the first and the second research questions, many studies on DTA models defined experts as individually best-performing agents and laypersons as randomly selected agents with diverse strategies. We focus on two major questions. Does diversity trump ability? What is the epistemically best composition of deliberation groups for many issues? The review result revealed that the DTA phenomenon suggested by the original DTA model is robust only for problem-solving cases about unpredictable issues. The unpredictability or difficulty of a problem in the DTA model is represented as the ruggedness of the landscape in a computer simulation. The landscape of the original DTA was extremely rugged with randomness (Holman et al. 2018, 263).

Seven studies provided results for collective problem solving with unpredictable issues, that is, with an extremely rugged landscape. Among them, six studies identically agreed with the result that the DTA phenomenon is robust

for problem-solving cases about highly unpredictable issues (Grim et al. 2018; Holman et al. 2018; Hong and Page 2004; Singer 2018; Thoma 2015; Weymark 2015).[7] The result held even when the definition of "experts" was altered from a problem-specific knowledge holder to a transportable knowledge holder (Grim et al. 2018; Holman et al. 2018). A modified ELM version also gave support to the DTA phenomenon holding only for a rugged landscape (Thoma 2015, 466). This ELM result held only for the case in which agent movement is restricted, that is, a limited range of agent movements causes functionally similar effects of unpredictability. Generally, we can find much evidence of the robustness of the DTA phenomenon in cases of very unpredictable issues (difficult problems) in the social epistemology literature.

5.2. The Best Composition of Groups for Modestly Predictable Issues

Outside of the highly unpredictable range, some studies supported a mixed population of experts and laypersons as the best composition of problem-solving groups. Although the original DTA model study does not answer this question, three out of the nine studies reviewed answered it. Existing studies, although few, suggested that a mixed group of experts and laypersons is possibly the best composition for collective problem solving on modestly predictable issues. Two DTA studies suggested a mixed group as the epistemically best composition (Grim et al. 2018; Holman et al. 2018). These two studies modified the original expert representation, with experts being transportable knowledge holders. Moreover, one study on ELM with a medium movement range (a moderately difficult problem) also suggested a mixed group as the best (Thoma 2015).

Tournament dynamics, such as roundtable deliberation in which all participants simultaneously deliberate, favored more laypersons in mixed populations; the appropriate composition of the problem-solving group was about a 40:60 expert to layperson ratio (Grim et al. 2018; Holman et al. 2018, 270). Relay dynamics, such as bulletin-board online deliberation in which discussions take place within one's mind and one's best judgment is handled to the next deliberator, benefited from a primarily expert group with a small portion of laypersons (Holman et al. 2018, 270). Other studies on ELM suggested

[7]Thompson's negative result that the DTA result is a mere artifact of modeling was criticized (Singer 2018; Thompson 2014).

about a 50:50 expert to layperson ratio as the epistemically best composition for relay dynamics (Thoma 2015, 468). This sensitivity of simulation analysis suggested that the ratio of experts and laypersons may possibly change depending on model settings.

5.3. Synthesis

Table 1 summarizes the synthesis of the results of this systematic review. The original DTA model study obtained a DTA result under the condition of issue unpredictability. The extended DTA version and related ELM model studies suggested that a mixed group trumps both the expert and layperson groups under the condition of modest issue predictability.

5.4. Review Conclusion

The more realistic factors were included in the DTA-related models, the more likely it was that mixed groups of experts and laypersons would perform the most strongly. We found that a combination of experts and laypersons produced epistemically superior results when the tasks and issues were not fully unpredictable. This result was confirmed across many related but different models, indicating that this result is relatively robust. As politics and policy issues are somewhere in between fully contingent issues (lottery) and fully predictable ones (technical problem), we should question the validity of the original DTA result for political philosophy.

6. Discussion

This article aimed to partly supplement the procedure of robustness analysis of mathematical models in epistemic democracy with the systematic review method. Epistemic democrats have found diversity to be a major source of epistemic success in problem solving based on the original DTA model (Gaus 2016; Landemore 2013). By contrast, robustness analysis with systematic review found the possibility that the mixture of diversity and expert knowledge is a major source of epistemic success for most issues. The results of this pilot study support the argument that robustness analysis of mathematical models with systematic review aids epistemic democracy by alleviating the possibility of narrow-sighted and biased recommendations.

6.1. Contributions to Epistemic Democracy

The proposed method helps improve the credibility of model analysis and draw a full picture of the target phenomenon. First, it increases the validity of

Table 1. Synthesis of the Systematic Review Results on DTA-Related Models.

	Study	Model	Definition of Experts	Definition of Laypersons	Deliberation Dynamics	Diversity Trumps Ability? Issue Predictability: Little	Modest	High	Best Composition of Participants (for Modestly Predictable Issues)
The original DTA model									
1	Hong and Page 2004	DTA	Best-performing agents	Randomly selected agents	Relay	Yes	—	—	—
					Tournament	Yes[a]	—	—	—
Family of DTA									
2	Holman et al. 2018[b]	DTA	Transportable knowledge holders	Randomly selected agents	Relay	Yes	No	No	Experts 90%: Laypersons 10%
					Tournament	Yes	Mixed result[c]	Seldom[c]	Experts 40%: Laypersons 60%
3	Grim et al. 2018[b]	DTA	Transportable knowledge holders	Randomly selected agents	Relay	Yes	No	No	Experts 90%: Laypersons 10%
					Tournament	Yes	Mixed result[c]	Seldom[c]	Experts 40%: Laypersons 60%
4	Singer 2018	DTA	Best-performing agents	Randomly selected agents	Relay	Yes	—	—	—
5	Weymark 2015	DTA	Best-performing agents	Randomly selected agents	Relay	Yes	—	—	—
						(DTA model is irrelevant when decisions are binary)			

(continued)

Table 1. (continued)

	Study	Model	Definition of Experts	Definition of Laypersons	Deliberation Dynamics	Diversity Trumps Ability? Issue Predictability: Little	Modest	High	Best Composition of Participants (for Modestly Predictable Issues)
6	Thompson 2014	DTA	Best-performing agents	Randomly selected agents	Relay	No	—	—	—
						(Random algorithm causes) (Random > diversity > ability[d])			
Related models in other research subjects									
7	Weisberg and Muldoon 2009	ELM	Following a successful strategy (followers)	Unvisited patch search strategy (mavericks)	Relay	—	—	Yes[e]	—
8	Thoma 2015	ELM	Following a successful strategy (modified version[f])	Unvisited patch search strategy (explorers[g])	Relay	Yes	No	No	Experts 50%: Laypersons 50%
9	Alexander et al. 2015	ELM	Following a successful strategy (modified version[f])	Unvisited patch search strategy (mavericks)	Relay	—	—	No	—
						(Hill-climb > Follower > Maverick)			

Note. DTA studies evaluated the group performance of 9 to 20 agents based on the expected value of the group's stopping points. Meanwhile, ELM studies evaluated the group performance of 10 to 400 agents mainly based on time, the percentage of significant approaches explored, or the total number of approaches explored. DTA = diversity trumps ability; ELM = epistemic landscape model; — = not reported.

[a]There was no reported difference in results between relay and tournament.

[b]Research was conducted by overlapping authors.

[c]A large number of members and heuristics enhanced the DTA result in tournament dynamics.

[d]It was reported that randomness trumps diversity, but the diverse group in this study is criticized for its insufficient diversity (Singer 2018).

[e]This study is criticized for modeling experts too weak with errors (Alexander et al. 2015; Thoma 2015).

[f]Different modifications were employed.

[g]The ELM model agents are named differently.

epistemic democracy's model analysis in terms of its transparency and neutrality. This effect is due to the universal sample collection procedure for robustness analysis that is ensured by the guideline of the systematic review method. Accordingly, it alleviates the criticism on the trustworthiness of epistemic democracy's model analysis.

Second, the proposed method can draw a full picture of the target phenomenon. The proposed method allows synthesis methods that are not limited to the inference from the samples with common results. Indeed, the review result suggested that the popular view among epistemic democrats that amateur-only problem solving is epistemically the most effective can be an oversimplification. This was due to the epistemic democrats' limited range of observations. The review result suggested that deliberative problem solving by a mixed population of experts and laypersons is an interesting topic that has been given little attention by epistemic democrats.

6.2. Ranking Scheme of the Trustworthy Model Analysis

Robustness analysis supplemented with systematic review provides a relatively credible foundation for political philosophers. The methods of sample selection and result synthesis vary in trustworthiness from less trustworthy to trustworthy. This ranking scheme is useful in identifying a trustworthy baseline on which a philosophical argument can be established (Cartwright and Hardie 2012).[8]

Table 2 illustrates the ranking scheme. Single-model analysis may become less trustworthy because it sometimes falls short in identifying unforeseeable inappropriate assumptions and errors.[9] Even robustness analysis compromises its trustworthiness when it collects sample models based on their common results. Therefore, as this article proposes, researchers can supplement robustness analysis by collecting models and synthesizing the results using the systematic review method. In the final stage, the robustness of a given model analysis could be checked through meta-analysis. Meta-analysis of mathematical models in some cases is better described as a future goal because not many researchers have conducted meta-analysis of mathematical models to avoid hidden bias and we must wait until a sufficient number of

[8]Notice that this ranking does not exclude the possibility that a well-established and well-tested single-model analysis is trustworthy.

[9]Single-model analysis provides a detailed description of the mechanism of a phenomenon, whereas a many-model comparison ensures the trustworthiness of a model analysis, which single-model analysis cannot identify.

Table 2. Ranking Scheme of the Trustworthy Model Analysis Method.

	Trustworthy----------Less Trustworthy			
Synthesis Method	Meta-analysis of Math Models	Robustness Analysis with Systematic Review	Robustness Analysis based on Samples with Common Results	Single-Model Analysis
Shortcomings	Currently not enough studies	No statistical analysis	Selection bias Indeterminacy	Selection bias Not trustworthy
Status	Future goal	This article's proposal	Existing proposals	Some epistemic democrats rely

model studies have been accumulated. Political philosophers can exploit the ranking scheme of the credibility of model analysis on which their philosophical arguments are built.

6.3. Sufficient Level of Confirmation for Political Philosophy

Although this article did not provide a confirmation in the real world, the confirmation of abstract models against realistic settings is sufficient for a certain purpose of political philosophy that epistemic democrats aim for. The realistic settings are not the real world nor an experimental system (Guala and Mittone 2005, 501) in which one can conduct empirical tests.

We can classify three purposes of robustness analysis from the literature: discovery, proper modeling, and confirmation. First, the discovery branch regards robustness analysis as a methodological tool for hypothesis generation (Odenbaugh and Alexandrova 2011, 759). Second, the proper modeling branch suggests that robustness analysis can be used as a methodological tool that ensures "*low-level confirmation*, confirmation of the fact that certain mathematical structures can adequately represent properties of target phenomena" (Weisberg 2006, 740; italics in the original). Third, the confirmation branch claims that

robustness analysis can confirm causal relations within mathematical model analysis (Lloyd 2010, 2015).[10] Some argue that robustness analysis confirms the relative importance of model components (Kuorikoski, Lehtinen, and Marchionni 2010, 543). The required confirmatory power level of robustness analysis depends on the purpose and research area in question.

The robustness analysis in political philosophy could have the different purposes mentioned above. The confirmation of abstract ideal models against realistic settings reconciles two of the four roles of political philosophy stated by John Rawls: reconciliation through rational explanation and providing a realistic utopia (Rawls 2001, 1-5). Rawls claimed that political philosophy functions if it shows us how political institutions are rational (Rawls 2001, 3). He also stated that political philosophy aims to show "what would a just democratic society be like under reasonably favorable but still possible historical conditions" (Rawls 2001, 4). The confirmation of abstract ideal models against realistic settings draws a link between these two roles of political philosophy through model analysis. If successful, it reconciles these two functions of political philosophy within a model analysis. This article exemplified this course of argument through a robustness analysis of an abstract mathematical model on democracy. The robustness analysis of mathematical models can confirm an abstract model against a realistic setting at most. Yet, for political philosophers, this seemingly insufficient confirmation is sometimes adequate.

6.4. Limitations

6.4.1. A sophisticated hypothesis-generating tool. While this article aims to contribute to the hypothesis-generating method of epistemic democracy, it does not aim at hypothesis testing. Systematic review in medicine and the social sciences often uses a large set of evidence that has already been confirmed against real data. However, provided that models are for hypothesis generation, robustness analysis with systematic review of models can use only a large set of hypotheses, not evidence. Then, model comparison does not provide confirmation against real data. Therefore, the robustness analysis eventually requires empirical confirmation, but this article argues that robustness analysis with modifications for the refinement of hypothesis generation in epistemic democracy is worth conducting.

[10]For criticisms of the confirmation branch of robustness analysis, see Woodward (2006), Odenbaugh and Alexandrova (2011), Parker (2011), and Justus (2012).

6.4.2. Similarity of the compared models and forward reference list checking. This article conducted forward reference list checking from the original DTA model. Forward reference list checking supports the identification of the similarity of the various models compared, as they are considered to be related (or, in some cases, contrasted) by a number of researchers. Since the collected models have been manually checked as to whether they are related, forward reference list checking was a good starting point for collecting related models from the ocean of literature. As drawbacks, forward reference list checking may collect studies in favor of the original model and be affected by non-publication bias. It does not also collect related samples that did not cite the original study (e.g., Pöyhönen 2017; Zollman 2010). However, this article revealed that the result contradicted the expected bias, so such a concern was alleviated in this pilot analysis case on the DTA model. Moreover, although it was not based on universally collected samples per se, the pilot study could suggest different recommendations from the one based on a single-model study of DTA, therefore encouraging further research in epistemic democracy.

6.4.3. From pilot study to research project. This systematic review of the DTA model was only a pilot study. Although this study filed 352 papers, it could not review whole samples of related model studies. Simulation analysis of DTA and related models is still in progress. We must wait for the accumulation of simulation analysis of DTA to conduct a large-scale systematic review. Nevertheless, this article attempted to overcome this problem by contrasting DTA models with the ELM model, and it demonstrated the possibility of mixed-group deliberation, a new perspective for epistemic democracy.

7. Conclusion

This article argued that the systematic review method offers the procedure needed for the robustness analysis of mathematical models in epistemic democracy. It proposed two types of amendments to robustness analysis: commencing the sampling phase of robustness analysis with a systematic sample collection procedure and initiating the analysis from the identification of differences in results. These amendments alleviate the drawbacks of robustness analysis in epistemic democracy in terms of the universality of samples and the inference from the samples with the same results.

To exemplify the effectiveness of systematic review, this article conducted a pilot review on DTA models. Based on the analysis of nine models extracted

from 352 papers, we found the possibility that in many issues, the deliberation of a mixed group of experts and laypersons was epistemically superior. The review result differed from the recommendation that epistemic democrats made; the layperson group outperformed. This article argued that this exemplifies the effectiveness of robustness analysis supplemented with systematic review in epistemic democracy.

The article provided a pragmatic solution to the validity problem of model analysis in epistemic democracy. This paper's proposal uses robustness analysis neither for mere hypothesis generation nor as a confirmation against real-world phenomena. Rather, it proposes robustness analysis as a method of confirming abstract ideal models against realistic settings and as sufficing for the purpose of supporting democracy's epistemic benefits within political philosophy. Accordingly, political philosophers are advised to examine the result of robustness analysis with systematic review of mathematical models before they conduct a detailed single-model analysis or build philosophical arguments based on it, in order to know trustworthy conditions and mechanisms that are worth focusing.

Author's Note

The draft of this article was presented at the Asian Conference on the Philosophy of the Social Sciences (Nankai University, Tianjin, June 1, 2019). Some parts of the systematic review of the DTA models was presented at the fourth International Conference on Public Policy (Concordia University, Montréal, June 26, 2019).
Ryota Sakai is now affiliated with Chuo Gakuin University, Chiba, Japan.

Acknowledgments

I am grateful for the feedback of conference commentators and anonymous reviewers.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author(s) received financial support for the research of this article from Public Policy Studies Association Japan and Japanese Political Science Association.

References

Alexander, Jason McKenzie, Johannes Himmelreich, and Christopher Thompson. 2015. "Epistemic Landscapes, Optimal Search, and the Division of Cognitive Labor." *Philosophy of Science* 82 (3): 424-53. https://doi.org/10.1086/681766

Anderson, Elizabeth. 2006. "The Epistemology of Democracy." *Episteme* 3 (1-2): 8-22. https://doi.org/10.3366/epi.2006.3.1-2.8

Brennan, Jason. 2016. *Against Democracy*. Princeton: Princeton University Press.

Caplan, Bryan. 2012. "The Myth of the Rational Voter and Political Theory." In *Collective Wisdom: Principles and Mechanisms*, edited by Hélène Landemore, and Jon Elster, 319-37. Cambridge: Cambridge University Press.

Cartwright, Nancy, and Jeremy Hardie. 2012. *Evidence-based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.

Dietrich, Franz, and Kai Spiekermann. 2013. "Epistemic Democracy with Defensible Premises." *Economics & Philosophy* 29 (1): 87-120. https://doi.org/10.1017/S0266267113000096

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

Gaus, Gerald F. 2016. *Tyranny of the Ideal: Justice in a Diverse Society*. Princeton: Princeton University Press.

Gough, David, Sandy Oliver, and James Thomas, eds. 2012. *An Introduction to Systematic Reviews*. London: SAGE.

Grim, Patrick, Daniel J. Singer, Aaron Bramson, Bennett Holman, Sean McGeehan, and William J. Berger. 2018. "Diversity, Ability, and Expertise in Epistemic Communities." *Philosophy of Science* 86 (1): 98-123.

Guala, Francesco, and Luigi Mittone. 2005. "Experiments in Economics: External Validity and the Robustness of Phenomena." *Journal of Economic Methodology* 12 (4): 495-515. https://doi.org/10.1080/13501780500342906

Holman, Bennett, William J. Berger, Daniel J. Singer, Patrick Grim, and Aaron Bramson. 2018. "Diversity and Democracy: Agent-Based Modeling in Political Philosophy." *Historical Social Research* 43 (1): 259-84.

Hong, Lu, and Scott E. Page. 2004. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences of the United States of America* 101 (46): 16385-89. https://doi.org/10.1073/pnas.0403723101

Justus, James. 2012. "The Elusive Basis of Inferential Robustness." *Philosophy of Science* 79 (5): 795-807. https://doi.org/10.1086/667902

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Klein, Dominik, Johannes Marx, and Kai Fischbach. 2018. "Agent-Based Modeling in Social Science, History, and Philosophy: An Introduction." *Historical Social Research* 43 (1): 7-27.

Kuorikoski, Jaakko, Aki Lehtinen, and Caterina Marchionni. 2010. "Economic Modelling as Robustness Analysis." *British Journal for the Philosophy of Science* 61 (3): 541-67. https://doi.org/10.1093/bjps/axp049

Ladha, Krishna K. 1992. "The Condorcet Jury Theorem, Free Speech, and Correlated Votes." *American Journal of Political Science* 36 (3): 617-34. https://doi.org/10.2307/2111584

Landemore, Hélène. 2013. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton: Princeton University Press.

Levins, Richard. 1966. "The Strategy of Model Building in Population Biology." *American Scientist* 54 (4): 421-31.

Lloyd, Elisabeth A. 2010. "Confirmation and Robustness of Climate Models." *Philosophy of Science* 77 (5): 971-84. https://doi.org/10.1086/657427

Lloyd, Elisabeth A. 2015. "Model Robustness as a Confirmatory Virtue: The Case of Climate Science." *Studies in History and Philosophy of Science* 49:58-68. https://doi.org/10.1016/j.shpsa.2014.12.002

Odenbaugh, Jay, and Anna Alexandrova. 2011. "Buyer Beware: Robustness Analyses in Economics and Biology." *Biology & Philosophy* 26 (5): 757-71. https://doi.org/10.1007/s10539-011-9278-y

Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton: Princeton University Press.

Page, Scott E. 2018. *The Model Thinker: What You Need to Know to Make Data Work for You*. New York: Basic Books.

Parker, Wendy S. 2011. "When Climate Models Agree: The Significance of Robust Model Predictions." *Philosophy of Science* 78 (4): 579-600. https://doi.org/10.1086/661566

Petticrew, Mark, and Helen Roberts. 2006. *Systematic Reviews in the Social Sciences: A Practical Guide*. Malden: Blackwell.

Pivato, Marcus. 2017. "Epistemic Democracy with Correlated Voters." *Journal of Mathematical Economics* 72: 51-69. https://doi.org/10.1016/j.jmateco.2017.06.001

Pöyhönen, Samuli. 2017. "Value of Cognitive Diversity in Science." *Synthese* 194 (11): 4519-40. https://doi.org/10.1007/s11229-016-1147-4

Rawls, John. 2001. *Justice as Fairness: A Restatement*. Edited by Erin Kelly. Cambridge, MA: Belknap Press.

Singer, Daniel. 2018. "Diversity, Not Randomness, Trumps Ability." *Philosophy of Science* 86 (1): 178-91. https://doi.org/10.1086/701074

Thoma, Johanna. 2015. "The Epistemic Division of Labor Revisited." *Philosophy of Science* 82 (3): 454-72. https://doi.org/10.1086/681768

Thompson, Abigail. 2014. "Does Diversity Trump Ability? An Example of the Misuse of Mathematics in the Social Sciences." *Notices of the American Mathematical Society* 61 (9): 1024-30.

Weisberg, Michael. 2006. "Robustness Analysis." *Philosophy of Science* 73 (5): 730-42.

Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Weisberg, Michael, and Ryan Muldoon. 2009. "Epistemic Landscapes and the Division of Cognitive Labor." *Philosophy of Science* 76 (2): 225-52. https://doi.org/10.1086/644786

Weymark, John A. 2015. "Cognitive Diversity, Binary Decisions, and Epistemic Democracy." *Episteme* 12 (4): 497-511. https://doi.org/10.1017/epi.2015.34

Wimsatt, William C. (1981) 2007. "Robustness, Reliability, and Overdetermination." Reprinted in *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*, 43-74. Cambridge, MA: Harvard University Press.

Woodward, Jim. 2006. "Some Varieties of Robustness." *Journal of Economic Methodology* 13 (2): 219-40. https://doi.org/10.1080/13501780600733376

Zollman, Kevin J. S. 2010. "The Epistemic Benefit of Transient Diversity." *Erkenntnis* 72 (1): 17-35.

Author Biography

Ryota Sakai is formerly an assistant professor of political science and public administration at Asahi University, Japan and is now an assistant professor of political science and public administration at Chuo Gakuin University, Japan. His main interests are political theory and public policy. He has published on epistemic democracy, deliberative democracy, and evidence-based policy in public deliberation.