“Revisiting Plato’s Cave: On the Proper Role of Lay People versus Experts in Politics”

Hélène Landemore
Yale University
helene.landemore@yale.edu

& Ryota Sakai
Chuo Gakuin University
sakai.ryota@gmail.com

Prepared for presentation at the PSA symposium on modeling epistemic diversity, New Orleans, November 17, 2024

Abstract: This paper revisits a debate in epistemic democracy about the value of the Diversity Trumps Ability Theorem (Hong and Page 2004 and Page 2008) in supporting the claims of epistemic democrats in favor of more inclusive decision-making processes in politics. We conduct a systematic review of DTA results and conclude that while they generally support the epistemic claims of deliberative democrats, they also support reintroducing experts in certain contexts. We use these results to complicate Plato's metaphor of the cave by identifying different areas within it where ordinary citizens, experts, or a mix of both are in a better position to make decisions.

"Revisiting Plato's Cave"

1. *Introduction*

Plato believed that philosophers, namely the wise, would make the best rulers for society. He powerfully illustrated this belief with his famous allegory of the cave, in which ordinary people are stuck watching shadows in the darkness while only the trained philosophers ascend to the light and the sight of real things outside the cave. A full 2,400 years later, even as the commitment to democracy has become the norm, scholars are still debating the place of experts versus amateurs in politics. While inclusion of all on an equal basis has been defended as the "right" thing to do, many believe that on instrumental grounds it is a mistake and that we would be better off with, e.g., "10% less democracy"(Jones, 2020), restriction of voting rights based on knowledge requirements (Brennan 2016), the replacement of voting with simulated popular preferences augmented with experts' beliefs (Brennan in Brennan and Landemore 2021), and at the very least a rebalancing of our institutions in favor of more expertise (Moore, 2014, 2016, 2017).

Recent work in "epistemic democracy," however, argues for democracy, and sometimes even more of it than what is currently available, precisely on instrumental, knowledge-related grounds (e.g., Anderson 2008, Estlund 2009, Landemore 2013, Knight and Johnson 2011, Goodin and Spiekermann 2023). Giving voice and votes to ordinary citizens is not just the right thing to do, it might well be the smart thing to do, and indeed possibly "ecologically rational" given the structural uncertainty of political matters (Landemore 2014). A lot of that work appeals to a mathematical theorem, called the Condorcet Jury Theorem, to model the properties of judgment aggregation at the large scale and buttress the idea that "more is smarter." Some of that work also appeals to a more recent and more controversial mathematical result, called the "diversity trumps ability" theorem (DTA, as defended in Hong and Page 2004 and Page 2008) to model and explain the epistemic properties of deliberative problem-solving (e.g., Anderson 2006, Landemore, 2012, 2013).

According to the DTA, under the relevant conditions, a diverse group of problem-solvers (as often generated from a random draw from a pre-existing sample) is more likely to solve problems successfully than any subset of problem-solvers, including those involving the "smartest" problem-solvers (as defined by their individual probability of solving the problem).

The DTA offers an (admittedly crude) modelization of deliberative problem-solving, whereby diverse agents arrayed on a rugged landscape need to identify a global maximum of a set of points. They are "diverse" because they have different ex ante given probabilities of identifying different local optima (they use different "heuristics") and only have the global optimum in common (meaning once they see it, they can recognize it but they may not be able to identify it ex ante). The beauty of collective problem-solving among these agents is that when they pool their knowledge, taking turn to reveal what they think is the highest point, they will nudge each other until someone correctly identifies the global optimum and everyone can see it. We are better off with a group of agents using different heuristics, even if their probability of getting to the right answer first is lower, because too similar agents (using the same heuristics) might get stuck at a local optimum, even if their probability of getting it right is higher.

Like the CJT before it, Hong and Page's DTA model has been criticized on many grounds, including for its empirical irrelevance (Estlund 2009), its triviality (Brennan 2020; Romaniega, 2023), and its inadequacy in terms of its modeling (Thompson, 2014; Grim *et al.*, 2019; Reijula and Kuorikoski, 2021). While we reject certain criticisms (of triviality and empirical irrelevance for example), we accept that relying on one model alone does not sufficiently buttress the validity of the claim that authors build on the DTA, namely that more generally, for political purposes, many heads are better than a smart few. In other words, relying on a single model makes the argument vulnerable to model-specific assumptions, distortions, and artifacts—non-real phenomena resulting from the modeling process. In order to be able to generalize properly, we need to test the robustness of the claim that "diversity trumps ability" across many replications of the model.

The purpose of this article, therefore, is to make progress on our understanding of when and where democracy proves a superior decision-rule through a more comprehensive understanding of the model supporting the DTA theorem, as well as its extension by Landemore into a NTA (Numbers Trump Ability) theorem. To achieve this goal, we conducted a systematic review of the families of DTA models, i.e., models that replicate or improve on the original DTA models proposed by subsequent studies, thereby enhancing the generalizability of the model analysis and obtaining reliable results.

From this perspective, the study responds to several criticisms against epistemic democrats who have cited the DTA model to support democracy. The significance of this article is that it provides reliable analytical results regarding DTA that can serve as a touchstone for determining which, between the 'more democracy' school and the 'less democracy' school, should be chosen from an epistemic perspective. What we find is that Plato's allegory of the cave needs to be complicated: while it is true that experts are most reliable in a context of risk (where there is enough "light" and the probability distributions are known and computable), the balance of power should shift toward the larger group of non-experts as uncertainty, defined as being in the dark about the underlying probability distributions of reality, comes to dominate. The Allegory of the Cave proves apt in ways that Plato may not have fully anticipated: politics is a continuum from darker to more well-lit domains of questions. In the cave itself, decision-making should go to ordinary citizens first and by default, as we navigate the darkest and primary level of political uncertainty. But as the darkness lifts and uncertainty turns to complexity of an increasingly predictable kind, we should introduce experts in the mix and start delegating more power to them. This means, interestingly, that both Brennan and Landemore (Brennan and Landemore 2021) may be partially right and partially wrong in equal measures, just in different areas of the cave. When we exit the cave and reach into the light of self-evident truth, however, we exit politics altogether.

This paper is organized as follows. Section 2 introduces the debate over the DTA model and the merit of multiple-models analysis. Section 3 presents the research questions relate to the recent debates over the DTA model analysis in epistemic democracy, while section 4 describes our systematic review procedure. The results of the review are presented in section 5, and section 6 discusses and defends the arguments of the epistemic democrats. Finally, section 7 presents a new interpretation of Plato's allegory of the cave, suggesting that the optimal composition of participants for problem-solving varies depending on the brightness of the cave, or the predictability of the issue.

2. Theoretical background

Debate over the DTA theorem

In the field of political theory on epistemic democracy, there has been a debate between advocates of democratic decision-making and advocates of expert decision-making (Brennan and Landemore, 2021). This debate is triggered by the mathematical model analysis of collective intelligence. The model analysis has shown that the optimal composition of problem-solving group participants depends on the predictability of the issue being discussed. The analysis of the DTA model, recognized as a mathematical model of collective intelligence generated by deliberation, showed that when the predictability of the task is low, a cognitively diverse composition of participants, such as that achieved by drawing citizen lotteries, is superior (Hong and Page, 2004, Page, 2008).

Several scholars advocate greater inclusion in democratic decision-processes based on this model analysis (e.g., Anderson 2008, Landemore, 2013, Vermeule 2009). For instance, Landemore argued that democratic decision-making with diverse participants epistemically outperforms other nondemocratic decision-making procedures such as experts judgment or benevolent dictator's judgment "for most political problems" (Landemore, 2013, p. 3). This argument was based on her assumption that political questions are "a diverse, unpredictable set of difficult collective problems" (Landemore, 2014, p. 187). She further builds on the DTA to formulate her own Numbers Trump Ability theorem (NTA), whereby a proxy for cognitive diversity is assumed to be greater inclusiveness of the decision-process, allowing her to build a case for democracy on the epistemic properties of greater inclusion, on equal terms, in the deliberative process leading to political decisions.

Hong and Page's DTA theorem, however, has been criticized on many fronts. One of them is triviality. On some level, this result is indeed as trivial as critics make it out to be, as it produces conclusions already contained in the premises, namely that when a group of agents with a different skill set is better than a group of agents with exactly the same skill set. But we find this charge of triviality unfair, since all theorems are, by definition, tautological that way. The Condorcet Jury Theorem similarly contains in its premises the conclusion that under the specified assumptions for an infinity of voters majorities are always right. Yet it is still surprising to most people to go from the premises—specifically the premise that voters are barely smarter than a coin toss at making predictions--to that conclusion—at the limit the majority is virtually certain to be right, because the human mind tends to think linearly and what is mathematically trivial can still defy common sense. Similarly, or so we think, the "triviality" of the DTA goes against, in our view, our faith in experts and the idea that those who are better at a given task would form a superior group of decision-makers than more mediocre people who think differently. Our common sense is not necessarily wired to understand group properties and the benefits of difference, as opposed to individual properties and the benefits of more of the same.

Other critics still complain about the alleged inadequacies of its modeling (Reijula and Kuorikoski, 2021; Romaniega, 2023; Thompson, 2014). Others still argue that the epistemic landscape of the model is set in a way that disadvantages agents with high ability (Holman et al., 2018; Grim et al., 2019; Reijula and Kuorikoski, 2021), and thus, it is unsurprising that groups that include the best agents alongside other agents outperform groups of the best agents (Thompson, 2014).

These criticisms indicate that relying on one mathematical model alone does not sufficiently guarantee the generalizability of the model analysis. How do we extract trustworthy implications from a set of multiple mathematical models that may contain distortions and artifacts?

Philosophical cornerstone for conducting robustness analysis and systematic reviews

To address this issue, political theory can turn to philosophers of science, who have defended the benefit of multiple-model analysis. In keeping with his belief that cognitive diversity is a good thing, Scott Page himself defends the superiority of "multiple models" (Page, 2018) over unique ones. If multiple models converge on similar conclusions, then we can have more faith in the latter.

In this article we conduct a robustness analysis of existing mathematical models, along with a systematic review of related model analysis studies. A systematic review allows for systematic and neutral sample collection. Robustness analysis integrates the results of the collected samples and can reveal the reliable parts of the model analysis (Wimsatt, 2007, p. 46). In other words, robustness analysis is a mechanism to reduce the risk of being affected by model-specific distortions and artifacts by having multiple backup systems in parallel. These approaches are attempts to permit the coexistence of a group of models with several different assumptions and parameters and then analyze the commonalities identified among them.

A number of modified versions of the DTA model have been presented. Although the existence of multiple criticisms and modified models of the DTA suggests that the original DTA theorem is not perfect, the variety of these modified models becomes a valuable resource for enhancing the reliability of our DTA model analysis. The coexistence of diverse models forms a philosophical cornerstone for conducting robustness analysis and systematic reviews of models.

Although multiple model analyses of DTA have been conducted to date, it was not previously known what synthesizing these individual studies would reveal about the whole-picture view of DTA. The pioneering analysis aimed at gaining an overall picture of DTA, as conducted by Sakai (2020), was undertaken with a limited number of studies. Our new review expands on this by increasing the number of studies analyzed from 352 to 2,662. It also refines the target model by focusing solely on DTA rather than other collective intelligence models. Not only methodologically, but also theoretically, the former review was conducted without focusing on how the review results would contribute to the “more democracy or less democracy” debate within epistemic democracy. Additionally, it minimally considered uncertainty in politics and policy.

3. Research questions

Research Questions:

This article was designed to answer the following research questions.

RQ1: What is the epistemically optimal composition of participants for problem-solving?

RQ2: What is the appropriate definition of ability?

RQ3: What is the appropriate definition of diversity?

RQ4: Is the 'diversity trumps ability' effect attributable to diversity, or is it a result of random-choice algorithms?

RQ5: Do numbers trump ability?

RQ1 is related to the debate among political theorists over the scope of democratic decision-making. The dispute between the ‘more democracy’ school, which supports the democratic participation of diverse citizens, and the ‘less democracy’ school, which sets limits on democratic participation and leaves room for expert judgment, has not been settled in political theory (Brennan and Landemore, 2021). Reliable analytical results related to DTA could serve as a touchstone for determining which school should be chosen from an epistemic perspective.

In terms of RQ2 (“What is the appropriate definition of ability?”), it is necessary to explore how expertise is defined in the DTA model. Hong and Page (2004) define an individually high-ability group for a given task and compare it with a diverse group. The definition of an individually high-ability group in the Hong and Page model has been criticized because it differs from the usual definition of experts (Holman *et al.*, 2018). For Hong and Page, experts have expertise in past events that cannot be applied to future events. Yet, we generally expect that the experts of yesterday will be the experts of tomorrow. Hong and Page do not. Therefore, it is necessary to investigate how ability is defined within the families of DTA models and to examine its impact on the results.

RQ3 asked “What is the appropriate definition of diversity?” Diversity in the DTA model is interpreted as diversity in prediction heuristic that agents have (Page, 2008). Hong and Page interpreted a randomly selected group from the agent pool a “diverse” group. However, this begs the question of whether a group constructed in this way is really diverse. Indeed, the evaluation of diversity depends on the measurement criteria, but there is controversy about how to measure diversity (Singer, 2018; Hankins, Muldoon and Schaefer, 2023). Therefore, we investigated how diversity is measured in the families of DTA models and how different diversity metrics affect the interpretation of DTA.

Relative to RQ4 (“Is the 'diversity trumps ability' effect attributable to diversity, or is it a result of random-choice algorithms?”), the results of the DTA model have been criticized as being the artifacts of a random selection algorithm and not the effects of diversity (Thompson, 2014).

While it is well known that random selection algorithms epistemically perform well, DTA modeling also employed a random selection algorithm. This creates confusion for readers in distinguishing whether the epistemic performance of the DTA model is attributable to the random selection algorithm or to diversity. In model analysis, it is necessary to distinguish between essential causal relationships and nonessential artifacts that are created in the process of model construction. We concluded that rather than relying on a single model's analysis, we could identify model-specific artifacts by evaluating the analyses of multiple models (Weisberg, 2013). Thus, we resolved to conduct a robustness analysis of the DTA model using a neutral and broadly collected sample through a systematic review to help answer this research question.

RQ5 asks whether numbers trump ability because Landemore expanded on the implications of the DTA model to propose the NTA theorem. Landemore posits that the easiest way to ensure diversity is to expand the number of participants in collective decision-making. If the NTA model is accurate, the political inclusion of citizens can be supported for its instrumental merit and thus enhance the practice of inclusive democracy (Landemore, 2014, p. 188). Relatedly, Hong and Page pointed out that problem-solving by a number of people is one of the conditions for the formation of DTA. The number of people required to solve a problem also depends on the difficulty of the issue and the abilities of the participants (Page, 2008, p. 162). However, neither Hong and Page nor Landemore have yet provided a formal mathematical demonstration for the NTA theorem; thus, we needed to look for evidence in the families of DTA models.

These five research questions address the optimal ratio of high-ability and diverse participants, the proper definitions of ability and diversity, the internal validity of DTA simulation modeling, and the NTA theorem. We proposed to answer these research questions by synthesizing the analytical results of the families of DTA models as accumulated through twenty years of follow-up research.

4. Research Methods

To answer the research questions mentioned in the section above, we conducted an updated systematic review of DTA-related models. In a systematic review, Sakai (2020) conducted a robustness analysis of DTA-related models extracted from 352 articles published between 2004 and 2018 in the Web of Science database. Our new review expanded the number of studies to be analyzed from 352 to 2,662. This paper will conduct a more extensive database, including the most recent studies (Scopus: 2004–2023). Since a significant number of new studies have been published since the previous review, this more extensive review allows for more reliable results to be presented.

Research Guidelines

In accordance with the systematic literature review of collective intelligence carried out by Suran, Pattanaik, and Draheim (2020), we followed Kitchenham and Charters's (2007) guidelines for performing systematic literature reviews of software engineering.

Search Strategy

As this article focused on the DTA model rather than collective intelligence in general, we used forward reference list checking (Gough, Oliver, and Thomas 2012, p. 126). Hence, we collected all the articles that cited Hong and Page's 2004 paper, which proposed the DTA model, and Page's 2008 book (both the original edition and the new edition), which helped disseminate the model. Although systematic reviews often use keyword searches, since this article is a review of the DTA model, we collected studies that cited the abovementioned works by Hong and Page. We expected that researchers very likely have cited these works when conducting original research on the DTA model. By adopting this approach, we sought to avoid collecting too many studies that did not address the DTA model.

Academic Database

As the first research on the DTA model was published in 2004, we searched the Scopus database for articles published between 2004 and December 2023 (search date: January 29, 2024). Scopus is the largest indexer of research content globally, indexing titles from more than 7,000 publishers. It includes journals, books, and conference papers (Elsevier, 2024).

Identified Records

In total, 1,122 records (articles, books, and conference reports) were identified that cited Hong and Page (2004)[1]. The records that cited the original edition of Page's (2008) book were 1,079, and those that cited its new edition were 755. We removed 294 duplicates from the combined 2,956 records, which left 2,662 unique records.

Record Selection

To identify the records relevant to our research questions, we followed a two-phase selection process. One of the researchers evaluated the identified records and selected those that were most pertinent to our research questions. Then, the researcher conducted a test-retest process

[1] In accordance with the Cochrane project, we call each work "record" until its assessment has been completed. Assessed works are called "studies".

(Kitchenham and Charters, 2007, p. 33) by checking a random sample of the selected records after the first review to see if the decisions to include or exclude them were consistent.

Selection Phase 1. In this phase, we studied the titles and abstracts of the identified records and extracted records containing the word "model" or "simulation." Records that contained "modeling" in their titles and abstracts were included. This selection phase chose simulation results as the main topic of this article. After completion, 642 records were selected.

Selection Phase 2. In this phase, we selected records on the emergence of collective intelligence with original simulation results of hill climbing, epistemic landscape, and similar models. We included studies that meet either of the inclusion criteria shown in Table 1. Given the criticisms leveled against the internal validity of the original DTA model, this phase aimed to select a set of records that would answer the question whether the existence of the "diversity trumps ability" phenomenon, as suggested by the DTA model, can be confirmed by mechanism models similar to the DTA model or modified versions of it. Importantly, network models were not included. We will address some of the important implications of network models in the discussion section (e.g., Smaldino et al., 2023). We excluded 458 records that did not meet criteria (1) or (2), 40 records that did not meet criterion (3), and 92 records that did not meet criterion (5); 52 studies were kept.

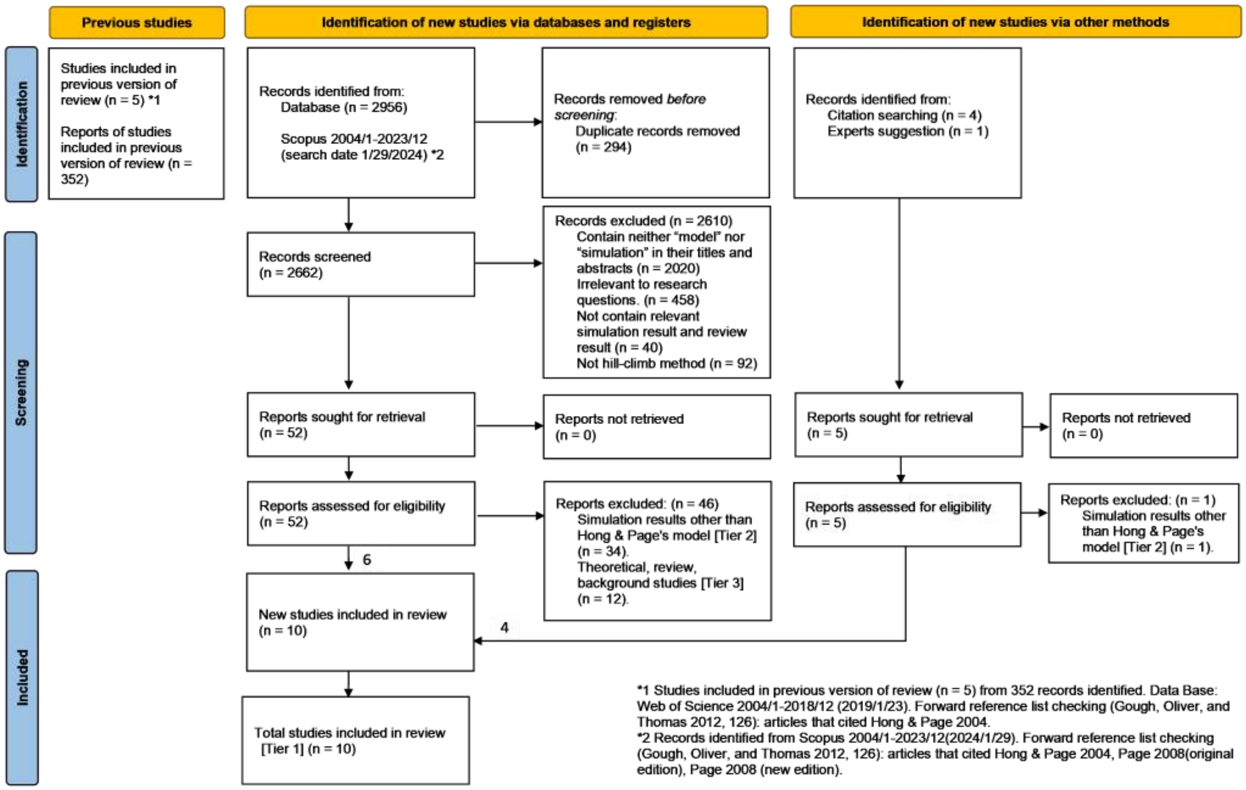
Table 1 Inclusion Criteria

Inclusion Criteria: (1) Studies related to at least one aspect of our research questions. (2) Studies of the emergence of collective intelligence. (3) Studies based on original simulation results or a review of studies using simulation. (4) Studies using the following models: hill climbing, epistemic landscape, and similar models.

Manually Added Records

We then scrutinized the reference lists of the selected studies to find any related articles that were possibly missed in our initial search. We identified four such articles that met our inclusion criteria (Grim *et al.*, 2018; Takahashi *et al.*, 2013; Thompson, 2014; Weymark, 2015). Moreover, we manually added one recent article that we knew of (Romaniega, 2023), which had been published on arXiv, a preprint server, but was not yet recorded in Scopus. We evaluated these articles using the same criteria to determine their inclusion on the short list. The selection process is illustrated with a PRISMA flow diagram in Figure 1.

Figure 1 PRISMA flow diagram



Study Quality Assessment

This phase evaluated the relevance of the selected studies. We classified studies with replicated simulations of Hong and Page's DTA model or mathematical analysis as Tier 1. After completing this assessment, 10 studies satisfied the quality criteria (6 studies from database search and 4 studies from manual search). These studies are the sources of the answers to our research questions. The short list of Tier 1 studies is shown in Table 2.

Table 2 List of Selected Studies

ID	Study Title	Author	Publication Type
Original model			
S1	Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers	Hong & Page 2004	Journal Article
Replication models			
S2	Does (mis)communication mitigate the upshot of diversity?	Hankins K., Muldoon R., Schaefer A. 2023	Preprint Server Paper
S3	Modeling cognitive diversity in group problem solving	Reijula and Kuorikoski 2022	Proceedings
S4	The diversity-ability trade-off in scientific problem solving	Reijula and Kuorikoski 2021	Journal Article
S5	Diversity, ability, and expertise in epistemic communities	Grim et al. 2019	Journal Article
S6	Representation in models of epistemic democracy	Grim et al. 2018	Journal Article
S7	Diversity and democracy: Agent-based modeling in political philosophy	Holman et al. 2018	Journal Article
S8	Diversity, not randomness, trumps ability	Singer 2018	Journal Article
Mathematical analysis			
S9	Fatal mathematical errors in Hong-Page Theorem and Landemore's epistemic argument	Romaniega 2023	Preprint Server Paper
S10	Cognitive Diversity, Binary Decisions, and Epistemic Democracy	Weymark 2015	Journal Article
S11	Does diversity trump ability? An example of the misuse of mathematics in the social sciences	Thompson 2014	Journal Article

Then, we classified studies with simulation results on the existence of the DTA phenomenon (results from models other than Hong and Page's model) as Tier 2. In this case, 35 studies satisfied the quality-assessment criteria. These studies are useful to confirm the existence and characterization of the DTA phenomenon from a broader perspective (i.e., beyond that of the DTA model). Tier 3 studies provide the theoretical background for the discussion; we identified 12 studies.

Comparison with Results of Previous Systematic Review

When examining these 10 Tier 1 studies, we found that five of them had already been included in a previous systematic review of DTA (Sakai, 2020), while five new studies were added to our shot list. These new studies were published between 2018 and 2023.

This systematic review covers a broader group of studies compared to the abovementioned previous review of the DTA. In the previous review, five studies remained as primary studies. All of them were on the short list of the present review. This suggests the comprehensiveness of this article's review.

Data Extraction

We extracted the following data items from the 10 selected studies: title, author, date, journal type, topic and setting, research question, type of model, main findings, comments on quality, and generalizability. Based on these data, we answer our research questions in the next section.

5. Results

In this section, we show that many of the selected studies support the DTA results, as presented by Hong and Page (S1), based on different modeling and assumptions (such as deliberation dynamics, generalizable definition of expert, measurement of diversity, and others). Therefore, the characteristics of the DTA phenomenon presented in Hong and Page's study (S1) are robust. However, a deeper examination of these models shows a more complete picture of DTA that Hong and Page did not uncover. The synthesis of the results of this review (Table 3) arguably complicates the picture of Plato's allegory of the cave presented 2,400 years ago.

Table 3 The synthesis of the systematic review of the DTA models (link to the table)

No.	Study	Model families	Modifications	Deliberation Dynamics	Diversity trumps ability? Issue predictability			Number trumps ability?	Modeling factor1: Definition of experts	Modeling factor2: Definition of diverse	Modeling factor3: Measurement of	Modeling factor4: number of agent	Modeling factor5: communication noise	Other features and notes
					Low	Moderate	High							
S1	Hong & Page 2004	DTA (Ringworld model)	Original	Relay Tournament *1	Diversity (Relay) Diversity (Tournament)	NR NR	NR NR	NR	Best-performing agents	Randomly selected agents	HP-diversity	10, 20	Perfect communication without noise and cost	
S2	Hankins K., Muldoon R., Schaefer A. 2023	DTA (Ringworld model)	(1)Communication occurs with error (2)Different measurement of diversity	1. Relay 2. Tournament Hybrid (Relay + Tournament)	1. Diversity (Relay) 2. Diversity (Tournament) 3. Diversity (Hybrid)	1. Ability (Relay) 2. Weakly Diversity (Tournament) 3. Weakly Diversity (Hybrid)	1. Ability (Relay) 2. Indifferent (Tournament) 3. Indifferent (Hybrid)	NR	Best-performing agents	Randomly selected agents	(1) Minkowski (Manhattan) distance (2) Hamming distance	9	(1)Perfect Communication (2)Fixed Error (3)Normally Distributed Error (4)Exponentially Distributed Error (5)Poisson Distributed Error	Model is based on S8
S3	Reijula and Kuorikoski 2022	1:DTA(Ringworld model) 2: DTA(Stairway landscape model) 3: Binary string model	Computationally replicate and expand Hong and Page's findings by using the binary string model	Relay	1. Diversity (Ring world model) 2. Diversity *2 (Stairway landscape model) 3-1. Ability (Binary string model with fixed start point and fixed heuristic order) 3-2. Indifferent (Binary string model with random start point and	1. NR (Ring world model) 2. Indifferent *2 (Stairway landscape model) 3. NR (Binary string models)	1. NR (Ring world model) 2. Weakly ability *2 (Stairway landscape model) 3. NR (Binary string models)	1. NR (Ring world model) 2. NR (Stairway landscape model) 3. Mixed result (Binary string model)	Best-performing agents	Randomly selected agents	HP-diversity	10	Perfect communication without noise and cost	(1) Increasing marginal returns to an added problem solver. (2) Utilized, mutantis mudandis, the assumptions made in S1.
S4	Reijula and Kuorikoski 2021	DTA (Stairway landscape model)	Introduced model for medium-hard problems (Stairway landscape model)	Relay	Diversity	Indifferent	Weakly ability	Yes	Best-performing agents	Randomly selected agents	HP-diversity	10	Perfect communication without noise and cost	Conditions are same as S1.
S5	Grim et al 2019	DTA (Ringworld model)	(1) Define expertise as portability of ability. (2) Conduct parameter sweeps across smooth landscapes.	1. Relay 2. Tournament	1. Diversity (Relay) 2. Diversity (Tournament)	1. Ability (Relay) 2.)Mixed result (Tournament)	1. Ability (Relay) 2. Ability (Tournament)	1. Yes (Relay) 2. Yes (Tournament)	Transportable expertise	Randomly selected agents	C-diversity	9	Perfect communication without noise and cost	(1) Larger Heuristic Pools trumps ability [Fig. 4, 7] (2) Overlapping data and results as S7. (3) Base ABS model as S8.
S6	Grim et al. 2018	DTA (Ringworld model)	Introduce representational structure	Relay Tournament	Diversity (Relay) Diversity (Tournament)	NR	NR	NR	Randomly selected agents	Best-performing agents	C-diversity	9	Perfect communication without noise and cost	(1) Modeling results show that representation can preserve and even slightly amplify the epistemic virtues of collaborative search. (2) Base ABS model as S8.
S7	Holman et al. 2018	DTA (Ringworld model)	(1) Define expertise as portability of ability. (2) Conduct parameter sweeps across smooth landscapes. (3) Present the best composition of group members	1. Relay 2. Tournament	1. Diversity (Relay) 2. Diversity (Tournament)	1. Mostly of experts with few diverse participants (Relay) 2. Equal mix of experts and diverse participants (Tournament)	1. Mostly of experts with few diverse participants (Relay) 2. Equal mix of experts and diverse participants (Tournament)	1. Yes (Relay) 2. Yes (Tournament)	Transportable expertise	Randomly selected agents and Max-diversity group	C-diversity	3,6,9	Perfect communication without noise and cost	(1) Best composition of groups (Relay) Exp. 90% Ran. 10% (Tournament) Exp. 45% Ran. 55% * except unpredictable landscape (2) Base ABS model as S8.
S8	Singer 2018	DTA (Ringworld model)	(1)Introduced the measurement of coverage diversity (2) Agent-based Ringworld model	Relay	Diversity	NR	NR	NR	Best-performing agents	Randomly selected agents	C-diversity	10	Perfect communication without noise and cost	
Mathematical Analysis of DTA														
S9	Romaniega 2023	DTA	Mathematical replication	Relay	Ability	NR	NR	NR	Best-performing agents	Randomly selected agents	NR	NR	Perfect communication without noise and cost	
S10	Weymark 2015	DTA	Binary decision	Relay	DTA model is irrelevant when decisions are binary	NR	NR	NR	Best-performing agents	Randomly selected agents	NR	NR	Perfect communication without noise and cost	
S11	Thompson 2014	DTA	Mathematical replication	Relay	Random > Diversity > Ability	NR	NR	NR	Best-performing agents	Randomly selected agents	NR	NR	Perfect communication without noise and cost	

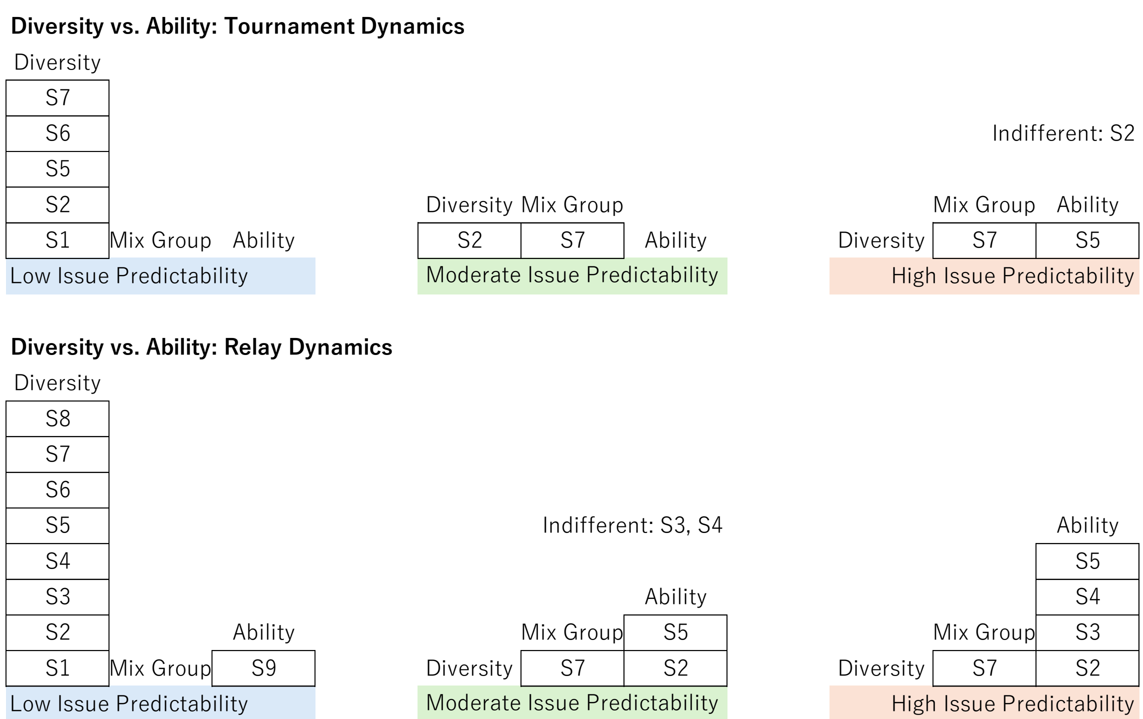
NR = not reported.

RQ1: What is the epistemically optimal composition of participants for problem solving?

Diversity trumps ability

In a similar model setting to the model analysis by Hong and Page (S1), the outperforming of ability by diversity was confirmed by replications of the DTA model (S2, S3, S4, S5, S6, S7, S8). In Figure 2, it is notable that studies reporting the advantage of diversity in cases of low issue predictability are frequent. All studies, including the model analyses of Hong and Page, that used simulation to analyze the less predictable domains of issue predictability confirmed that diversity outperformed ability (S1, S2, S3, S4, S5, S6, S7, S8).

Figure 2 The frequency of studies supporting "diversity trumps ability"



Moreover, the DTA result achieved structural robustness, which means that the result is unaffected by "changes to the model's mechanistic attributes" such as model's procedure and new state variables (Weisberg, 2013, p. 160). "Using this approach, the theorist can probe which parts of the causal structure represented by her model are really essential for the production of an observed behavior of the model" (Weisberg, 2013, p. 161). We found that the DTA result is robust under the model's structural variations, such as deliberation dynamics, communication noise, representation system, and epistemic landscape structure.

First, the DTA result was confirmed to hold independent of deliberation dynamics: sequential (relay dynamics) and simultaneous deliberation (tournament dynamics) (S5, S7). Second, the result that diversity trumps ability remained for both the relay and tournament dynamics, even when errors were added to the communication (S2: Hankins et al., 2023). Third, the DTA result held true in the representational system, where the population is divided into subgroups, the results of the subgroup search are handed to the representatives, and the representatives of each subgroup bring these results to a higher-level meeting body. The representative structure, incidentally, outperformed direct participation. The DTA result was more striking in the representational system with tournament dynamics (S6: Grim et al., 2018). Lastly, the DTA theorem found support among three family models of DTA that differ in their epistemic landscape structures: the Ringworld model (S1: Hong & Page, 2004), Singer's agent-based simulation model (S2, S5, S6, S7, S8), and the Stairway model (S3, S4).

We should note that one study reported that the DTA result did not hold for the Binary string model (S3: Reijula and Kuorikoski, 2022). As "the simulation model in Hong and Page (2004) can be seen as a simplified version of the Binary string model" (S3: Reijula and Kuorikoski, 2022, p.1864), it may potentially reduce the robustness of the DTA across different models.

The DTA result was structurally robust under the same conditions as in Hong and Page's study (S1). However, the analysis by Hong and Page left large areas of issue predictability unexplored.

The whole picture of Plato's cave

The above-mentioned robustness of the DTA theorem is limited to problem solving over a domain of issues with very low predictability. The epistemic landscape used by Hong and Page (S1) has extremely low predictability, with no relationship to the elevation of adjacent points. In this case, the epistemic landscape becomes an unstructured rugged landscape. Holman and his colleagues stated that "the extremely rugged landscape used by Hong and Page represents problem-solving in a case where none of the agents possess any understanding of the problem" (S7: Holman et al., 2018, p. 263).

Further analysis revealed that the DTA result did not achieve parameter robustness. What results can be expected when the predictability of the issue is higher than in the original Hong and Page modelling scenario? As with more regularity in the problem structure, issue predictability increases, and the epistemic landscape becomes smoother, with more consecutive rises and falls of the slopes. The replication of several DTA models shows that as the predictability of the issue changes from low to high, the results go from diversity trumps ability to ability trumps diversity.

In some simulation analyses (two of which are studies by overlapping authors), the results differed depending on the method of deliberation, i.e., deliberation dynamics (S2, S5, S7).

Relay dynamics: Sequential individual deliberation

In the model analysis by Hong and Page (S1), the results from relay dynamics were mainly reported from one agent to the next (as opposed to pooled among all of them, like in a more typical human deliberation). In relay dynamics or sequential deliberation, indeed, the first agent pushes the solution to the maximum value it can climb, and then the second agent takes over to improve the solution, and so on, until no one in the group can improve the solution. This is equivalent to deliberation on an online bulletin board or editing process of Wikipedia where contributors edit articles one after another.

When using relay dynamics, diversity trumped ability in areas of low issue predictability (S1, S2, S3, S4, S5, S6, S7, S8). However, when the predictability of the issue was moderate, there was no difference (S3, S4), or ability was weakly superior (S2, S5, S7). Moreover, in areas where the predictability of the issue was high, ability trumped diversity (S2, S3, S4, S5, S7).

In relay dynamics, the influence of ability is greater because the results from previous problem-solving agents influence later problem solving. Some studies focusing on optimal participant composition have reported that many experts mixed with a small number of diverse participants outperformed in both highly predictive issues and moderate predictable issues (S7: Holman et al., 2018).

Tournament dynamics: Simultaneous deliberation

In tournament dynamics or simultaneous deliberation, which is closer to what real human deliberation looks like, all agents participate in a simultaneous search for the highest solution, starting from the same point. In the second round, all agents search for a higher solution, starting from the solution with the highest elevation in the first round. This is repeated until no one can improve the solution any further.

When using tournament dynamics, diversity trumped ability in domains where the issue predictability was low (S2, S5, S6, S7). When the predictability of the issue was moderate, diversity was weakly superior (S2: Hankins et al., 2023), or the results were mixed (S5: Grim et al., 2019). In areas of high issue predictability, there were no differences (S2: Hankins et al., 2023), or ability outweighed diversity (S5: Grim et al., 2019). Some studies have reported that the optimal participant composition, a roughly equal group of experts and diverse participants, outperformed in both domains of high issue predictability and moderate issue predictability (S7: Holman et al., 2018).

Simultaneous deliberation, as expressed by tournament dynamics, weeds out inferior solutions within a round; therefore, errors are less likely to propagate (S2: Hankins et al., 2023, p. 23). Hence, it is likely that the advantages of inclusive participation outweigh the disadvantages associated with including participants of diverse abilities. This is compatible with Landemore's argument that deliberation weeds out the bad arguments and therefore the benefits of including even less educated or even misguided voices in the off chance they contribute some key insight or information outweigh the costs (Landemore 2013: 90-97).

The combination of both dynamics was also examined (S2: Hankins et al., 2023). When relay dynamics was used within subgroups and tournament dynamics was used between subgroups, diversity trumped ability in areas where issue predictability was low. When the predictability of the issue was moderate, diversity was weakly superior. There was no difference in the domains where the predictability of the issue was high.

Hong and Page (S1) argued that the results of tournament dynamics were similar to those of relay dynamics: “We considered environments in which a collection of agents attempt to find better solutions to the problem either sequentially or simultaneously. Our findings do not seem to depend on which structure was assumed” (S1: Hong & Page, 2004, p. 16386). As noted above, this conclusion holds when the predictability of an issue is low. However, replication studies of the DTA model showed that in other areas, overall, ability tends to dominate in relay dynamics, and diversity dominates in tournament dynamics (S2, S5, S7).

In summary, regarding the diversity versus ability issue, diversity trumped ability in areas of low issue predictability and ability trumped diversity in areas of high issue predictability. The result depended on deliberation dynamics: the tournament dynamic that corresponds better to actual deliberation did give a slight advantage to diversity in the intermediate areas. These results are robust, as similar results were derived from several types of modified models of DTA.

Optimal composition of discussion groups

Beyond the dichotomous discussion framework of diversity and ability, is it possible to read the optimal participant composition from the model analysis? In a study that attempted to do this, the optimal participant composition was identified according to issue predictability and the deliberation dynamics (S7: Holman et al., 2018).

In tournament dynamics (like real deliberation), the optimal participant composition was a diverse population in areas of low predictability, but in areas of moderate and high predictability, a population that included approximately an equal proportion of experts and diverse participants was effective (50 % - 70 % of diverse participants with 30 – 50 % of experts out of 9 agents) (S7: Holman et al., 2018).

In relay dynamics (discussion in a step-by-step manner, where each agent's deliberation is influenced by the preceding ones), the optimal participant composition was a diverse group in the low predictability domain, but in the moderate and high predictability domains, it was a group that included a number of experts and few diverse participants (10 – 20 % of diverse participants with 80 – 90 % of experts out of 9 agents). Interestingly, the inclusion of a small number of diverse participants in the high issue predictability domain was effective even in relay dynamics (S7: Holman et al., 2018).

Real-world (non-intentional) applications or verifications of this insight include the Brussels Regional Parliament and the French-speaking Brussels Parliament (FBR), whose so-called "deliberative commissions" are parliamentary committees composed of 45 citizens (selected via democratic lottery) and 15 Members of Parliament (MPs), as well as the G1000 process of mini-publics, where citizens make up 75% of the total participants, with the remaining 25% equally divided between civil servants, politicians, and employers (OECD, 2020, p. 47). In these latter cases, note that the practitioners go for a high ratio of "diverse" agents versus "experts" (2/ versus 1/3 in the former case and ¾ versus ¼ I in the latter). Presumably this is because in the real world we need to factor in the need for a critical mass of the "diverse" agents to be reached so that they have enough confidence to maintain epistemic peerhood with the experts.

In short, the proportion of diverse participants in the optimal participant mix for problem solving is greatest in the low issue predictability domain, but the proportion of experts increases as issue predictability increases.

RQ2: What are the appropriate definitions of ability?

Definition of the best-performing agents

Many studies have defined best-performing agents according to Hong and Page (S1) as "the agents with the highest expected values" (pp. 16386–16387) (S2, S3, S4, S8). For example, one study defined a best-performing agent as the "agents with the highest average competence on the landscape" (S2: Hankins et al., 2023, p. 12).

Transportable expertise and smooth landscapes

On the other hand, it may be difficult to understand the best-performing agents in Hong and Page's discussion as experts "for the best performing agents to be considered to be experts, we would expect to see their skill as transportable: we would expect them to perform roughly as well on other related questions"(S5: Grim et al., 2019, p. 108). In rugged landscapes, competency in one epistemic landscape does not guarantee competency in another similar landscape, since there is no regularity or structure in epistemic landscapes. Therefore, the proper interpretation of DTA as presented by Hong and Page (S1) is that "in problems where there are no real experts, it is a diverse group that typically does best (S7: Holman et al., 2018 , p. 266)."

Note however that what the model allows us to realize is that our everyday use of the term expert may not always track real expertise but instead illicitly attribute to people who are good in some domain competence in a seemingly adjacent, but in fact completely different domain. For example, people are willing to pay asset managers who talk a good talk and display dazzling mathematical skills a full 1 to 2% of their wealth when there is in fact no evidence that these oratory and mathematical skills translate into any ability to outperform a basic index fund over the long term (Taleb 2005). Similarly, it is not clear at all that economists are the best at governing the economy or doctors best at deciding health-related policies. Yet we make those illicit assumptions all the time (despite our commitment to democracy, which mercifully protects us from such a way of making political decisions). One could argue that what Hong and Page force us to do is realize that no one is an expert unless their skills are demonstrably transportable, not just remarkable in one domain.

And indeed, one study suggested that "smoother landscapes do exhibit this form of transportability of best performing heuristics" (S5: Grim et al., 2019, pp. 108–109). Increasing the smoothness of the epistemic landscape would solve the problem of Hong and Page's (S1) modeling. Since the Ringworld model they used and its replications (S5, S7) observed only a small difference between diversity and ability, some studies introduced new epistemic landscapes that better measure the tradeoff between diversity and ability (S3, S4). The results of the model analysis with such modifications are shown in the former section (S2, S3, S4, S5, S7).

RQ3: What are the appropriate definitions of diversity?

Definition of diverse groups

Diverse groups are randomly selected agents from an entire participant population (S1: Hong & Page, 2004, p. 16385). Hong and Page interpreted that "diversity is the key to collective performance" of randomly selected agent groups (S1: Hong & Page, 2004, p. 16387). Most replications of the DTA model share a similar definition of diverse groups.

However, a randomly selected population is not necessarily a diverse population. Hence, it is necessary to evaluate whether a randomly selected population is diverse by establishing criteria to measure diversity. This task is also important in settling the academic dispute about whether the cause of DTA is diversity or the effect of random selection (Page, 2015; Thompson, 2014).

Some studies used the same criteria for evaluating diversity that Hong and Page used (S3, S4). Let's call this HP-diversity.

> Hong and Page's diversity (HP-diversity): HP-diversity essentially measures the lack of overlap of heuristics of members in the group: the higher the HP-diversity, the less overlap there is among the heuristics (S8: Singer, 2018, p. 4).

However, the absence of overlap in a set of heuristics does not equate to using as much of the heuristics present in the population as possible. Even when HP-diversity is maximal, in some cases, only about 60% of the heuristics present in the population are utilized (S8: Singer, 2018, p. 7).

Some studies use the coverage of heuristics that are used in a group from the population as a measure of diversity (S5, S7, S8). Let's call this C-diversity.

> Coverage diversity (C-diversity): C-diversity measures "the percentage of the heuristic numbers that are represented in any spot in any heuristic in the group" (S8: Singer, 2018, p. 7).

The criticism that the DTA results are artifacts created by the random selection algorithm has attracted much attention. This controversy can be resolved by comparing a maximally diverse population with a population created by random selection (S8, S11).

A result of the comparison of both criteria reported that "C-diversity can explain the success of random groups much better than HP-diversity" (S8: Singer, 2018, p. 8). Multiple studies using C-diversity suggest that the results of the DTA theorem are due to diversity, not random selection algorithms (S5, S7, S8).

In addition, one study evaluated DTA models using commonly known diversity criteria, such as Hamming distance and Minkowski distance (Manhattan distance) (S2: Hankins et al., 2023). Differences in diversity evaluation criteria can lead to differences in results under the assumption that miscommunication is caused by the presence of diversity (S2: Hankins et al., 2023, p. 12).

RQ4: Is the 'diversity trumps ability' effect attributable to diversity, or is it a result of random-choice algorithms?

Thompson analyzed with simulation and concluded that randomness, not diversity, trumps ability (S11: Thompson, 2014). She argued that the DTA result is an artifact of random algorithm, which often outperforms, and that we should be cautious about making the wrong attribution of causes.

However, this argument has now been refuted, as Thompson's result was due to insufficient diversity in modeling, and when re-tested with a sufficiently diverse group of agents, the results showed that diversity outperformed randomness (S8: Singer, 2018). Another study supports this finding that the theorem relies on diversity and not merely on randomness" (S4: Reijula and Kuorikoski, 2021, p. 898).

RQ5: Do numbers trump ability?

Numbers trump ability

There is one more result that we want to explore, which is Landemore's hypothesis that the DTA can be translated for politics into a Numbers Trump Ability theorem (NTA), since by default having more people in the mix is bound to increase the diversity of heuristics and skill sets (or cognitive diversity) brought to bear on any given political problem (Landemore, 2012, p. 261, 2013a, p. 1217). To translate her claim in the categories used in this paper, Landemore thus argues that as the number of participants increases, the heuristic coverage of the population (C-diversity) increases. Therefore, she suggests, "more inclusive deliberating groups are, all things otherwise equal, likely to be smarter than less inclusive ones. Simply put, the more, the smarter" (Landemore, 2013a, p. 1217). This claim of course is meant to hold "all things otherwise equal" because we have to assume that communication costs or value conflict do not increase with numbers for the result to hold. Landemore also concludes from the NTA that wherever it is impossible to include more people, the next best solution is to take a random sample, which maximizes cognitive diversity in expectation across a range of political problems given the uncertainty she assumes in politics.

Landemore never formalized or tested any of these claims. Several studies (S3, S4, S5, S7), however, have seemingly found support for the numbers trump ability theorem. Consider that the first assumption of the NTA is that heuristic coverage increases as the number of participants increases. One study supports this point: "As group size increases, the redundancy in the high-ability group increases more than in the random group. This suggests that when the group size is larger, random groups again begin to approach the full heuristic, which obviously is sufficient for climbing the stairway sequence" (S4: Reijula and Kuorikoski, 2021, p. 13). Another study suggested that larger heuristic pools trump ability for both relay dynamics and tournament dynamics (S5: Grim et al., 2019, Fig. 4, p.117n11).

For the Binary string model, a generalized version of Hong and Page's DTA model, one study reported mixed results on marginal returns to an added problem solver. The study found increasing marginal returns to an added problem solver across group size of 2-8. However, as group size increases, the probability of increasing marginal returns decreases. (S3: Reijula and Kuorikoski, 2022, p. 1866).

The second assumption of the NTA is that the side effects of inclusion do not overwhelm the positive effects of inclusion. Under Hong and Page's (S1) conditions, a study (S7: Holman et al., 2018) suggested that "Landemore's claim that group performance can go up even as the individual average goes down finds support" (S7: Holman et al., 2018, p. 273). Landemore presupposed that "deliberators are expected, in the ideal speech situation where there are no time and information constraints, to reach an uncoerced agreement on the "better argument" (Landemore, 2013b, p. 92). This presupposition fits well with the theoretical characteristics of the DTA theorem (S7: Holman et al., 2018, p. 272). The study found that the maximum-diversity group (C-diversity) outperformed all other groups with various compositions of experts and diverse participants on low predictable issues, regardless of deliberation dynamics (S7: Holman et al., 2018). We can create a maximum-diversity group by replacing expert members with diverse participants. This often lessens the average ability of the group; however, the DTA theorem "will hold even if it means 'dumbing down' the average performance of the group" (S7: Holman et al., 2018, p. 273). This is because the replacement of diverse agents with high-ability members reduces the heuristic cover in a group (C-diversity). Notice again that this argument holds for the theoretical extension of Hong and Page's (S1) original conditions, in which issue predictability is very low.

The more complex and difficult the issue, the more diverse people are needed (S4: Reijula and Kuorikoski, 2021, Fig. 4).

Mathematical analysis of the DTA

Mathematicians tend to criticize the DTA theorem (Table 3). They focus on the mathematical part of the DTA theorem by Hong and Page (S9, S10, S11). However, Hong and Page (S1) have stated that "*A Computational Experiment* reports simulation results establishing that a diverse group can often outperform a group of the best. *A Mathematical Theorem* explores the logic behind the simulation results" (Hong & Page, 2004, p. 16386). Page argued that:

> Some nonmathematicians have stated that Lu and I "proved mathematically that diverse groups of people always outperform groups of the best." Obviously, such a proof would be impossible. Instead, Lu and I have used mathematics to identify sufficient conditions for a result to hold, a technique widely used by social scientists (Page, 2015, p.10).

The criticisms directed at the mathematical part in Hong and Page's paper (2004) may not necessarily align with the authors' intended objectives.

6. Discussion

The purpose of this article was to provide a comprehensive understanding of the DTA model. We demonstrated that DTA is robustly established across multiple models for issues with low predictability, that the predictability and diversity of issues are inversely correlated. In addition, we found that several studies suggested that the NTA theorem proposed by Landemore also holds under the conditions where DTA is established. These results support the epistemic democrats' hypothesis that diversity aids collective problem-solving, which can provide an epistemic rationale for democratic decision-making.

DTA theorem was defended

Our systematic review demonstrated that all of the simulation studies confirmed that diversity outperformed ability when issue predictability was low. The epistemic function of this diversity is consistent with the studies that show that it is "transient diversity" that provides the epistemic performance in collective problem-solving (Smaldino *et al.*, 2023). Borg et al. define transient diversity as follows:

> Transient diversity refers to a process in which a community engages in a parallel exploration of different theories, which lasts sufficiently long to prevent a premature abandonment of the best of the available theories, but which eventually gets replaced by a consensus on the best theory (Borg *et al.*, 2019, p. 2; Zollman, 2010).[2]

When a group aims to reach a consensus while improving the quality of its solutions, increasing the transient diversity of the solutions can help meet this goal. This result has been demonstrated through several simulation models and formal models, including the DTA model. Transient diversity has the effect of enhancing the epistemic function of a group by preventing premature convergence in collective problem-solving. There are multiple methods for achieving transient diversity and avoid premature convergence, including diversity in the behavioral patterns of the agents involved in problem-solving, communication noise between agents, and the sparsity

[2] Smaldino et al. (2023) includes epistemic landscape models within the scope of transient diversity, whereas Borg et al. (2019) do not, highlighting a distinction between the two studies.

of the network structure through which the group exchanges information. These methods for increasing epistemic performance have been presented individually in the past. Today, these separate methods can be explained as part of a single mechanism known as transient diversity (Smaldino *et al.*, 2023).

The division of epistemic labor

Our new systematic review of the families of DTA models revealed that some parameter robustness analyses on issue predictability in DTA showed an inverse correlation between issue predictability and the contribution of diversity. These results suggest the possibility of optimizing participant composition based on the level of issue predictability.

There are situations where democratic decision-making is effective, while in others, decisions can be left to experts. Some epistemic democrats have proposed a division of epistemic labor between experts and lay citizens (Landemore, 2013, 2014, p. 190). By combining this idea with an uncertainty index based on the policy category, it may be possible to elucidate the optimal participant composition tailored to the specific category of policy. Baker et al. (2016) developed the economic policy uncertainty (EPU) index, which measures the frequency of co-occurrence of each policy area and the keyword “uncertainty” in newspaper reports (Table 4). This provides a proxy measure of policy-related predictability (p. 1593). Here, we crudely assume that a situation involving uncertainty in an issue leads to its low predictability. Moreover, when considered alongside Theodore J. Lowi’s (1964) classical argument on policy typology, notable characteristics can be highlighted (pp. 690–691).

Table 4 Economic Policy Uncertainty Index of the US and the possible decision makers

	Policy Categories										
No. of months	Health care	Entitlement programs	Trade policy	Sovereign debt, currency crises	Taxes	Regulation	Fiscal Policy	Financial Regulation	Government spending	National security	Monetary policy
462(1985/1-2023/6)	132.06	126.09	124.12	112.18	110.74	110.45	109.54	105.91	102.24	96.04	95.81
	Low predictability					Moderate predictability					High predictability
	Democratic decision-making					A mix of citizens and experts participants					Expert decision-making

First, it should be noted that redistributive policies, including healthcare and entitlement policies (e.g., Medicaid and food stamps), have low predictability in EPU index. Lowi (1964) pointed out that redistributive policies affect a wider range of people. This could explain why the predictability of redistributive policies was lower in the EPU data. In redistributive policies, the range of those affected by the policy is wider and the size of the stakeholder base is larger, thus making the policy less predictable. In deliberations on redistributive policies, democratic decision-making that includes cognitively diverse participants and access to diverse information about the policy's reach may contribute to policy formation and consensus building.

Second, the predictability of national security, and monetary policy was high. Lowi (1964) cited defense procurement and defense-related R&D as examples of distributive policies, in which the range of direct beneficiaries is narrow (p. 690). Returning to EPU data, national security policies are closely linked to defense procurement and defense-related R&D. While monetary policy is not necessarily categorized as a distributive policy by Lowi (1964), its characteristics may limit the range of direct beneficiaries. So the participation of a group of people who share specific knowledge in the policy area (a uniform group) in deliberations may contribute to policy formation.

Finally, the predictability of tax, regulation, and fiscal policy was moderate. This result is in line with Lowi's (1964) argument, which places the range of beneficiaries of regulatory policies between those of redistributive and distributive policies. Policy areas such as climate change and disaster prevention require institutional design by experts, as well as consensus and cooperation to which citizens can commit.

Study limitations

We acknowledge several limitations in this article. First, during the systematic review process, we could not completely rule out the possibility that relevant literature was unintentionally excluded. We have attempted to mitigate this risk by conducting a test-retest process.

Second, while we attempted to explore the epistemic function of diversity through the DTA model, we did not discuss other models of collective intelligence. Nevertheless, various models of collective intelligence have consistently supported the epistemic function of diversity in collective problem-solving (Smaldino et al., 2023). Therefore, our systematic review of the DTA model aligns with the broader literature on collective intelligence. The primary contribution of this paper lies in its focused examination of the DTA model, providing a deeper understanding of the conditions under which it is effective.

Third, while we have sought to improve the internal validity of the implications of the DTA model, we have not addressed its external validity. However, the foundation of external validity rests on the presence of internally valid causal mechanisms. Thus, our efforts contribute indirectly to enhancing the external validity of the DTA model. Some critics claim that DTA is a simple idealized model with no external validity (Brennan, 2014, 2016; Quirk, 2014; Somin, 2014).

However, for the purposes of explanation and conceptual clarification that are made easier to handle by idealization and modeling methods (Mäki, 2020; Greco, 2023), and therefore, these criticisms may be a category mistake that downplays the unique value of each analytical method (Landemore, 2014). Models can be seen as part fable or story (Johnson, 2021), or as "proof-of-concept exercises about the possible mechanism" (Reijula and Kuorikoski, 2019, p. 274), but, at the same time, the predictions that models yield set limits on what we can reasonably expect in the real world (Dowding and Lenine, 2023).

In this article, we have attempted to improve the internal validity of the DTA model. It is reported that there is a similarity between models and experiments in terms of their ability to control conditions (Mäki, 2005). Through this similarity, there is the potential for closer integration between model analysis and real-world experiments. Indeed, empirical data involving the participation of diverse stakeholders have already been used to test the effects of the DTA model as proposed by epistemic democrats, with results supporting the conclusions of the DTA (Aminpour *et al.*, 2021). In response to empirical studies like this, this paper offers a more reliable framework for hypothesis construction by moving beyond reliance on a single model analysis.

7. Conclusion

Twenty-four hundred years ago, Plato believed that philosophers or the wise would be the best rulers. However, epistemic democracy may help complicate Plato's allegory about who has a claim to rule inside the cave.

Through the systematic review of DTA models conducted above, we extracted common characteristics across different models. This review also allows us to specify the domain of questions where experts and citizens have the upper hand in politics. Experts are more reliable on issues where its predictability is high or technical and ordinary citizens are more reliable on issues where its predictability is low. Therefore, Plato's Cave needs to be revisited as follows: in a pitch-black cave (a situation with low predictability of issues, caused by uncertainty), identifying experts is impossible (another way to say this is that contrary to our temptation to assume that some people always know more experts do not exist over this domain of question), and problem-solving tends to be epistemically superior among diverse participants. Thus, in darkness, diversity outperforms ability. Near the exit of the cave (a situation with higher predictability of issues), visibility allows for problem-solving to be epistemically superior when conducted by experts. Indeed, in technically predictable challenges, ability surpasses diversity. In the mid-point of the cave, where a sense of dim light is perceived, a mix of experts and diverse participants sometimes tends to yield superior epistemic outcomes in problem-solving.

Thus, it turns out that Plato was wrong and that the proposal of the epistemic function of diversity and the use of diverse participants in different situations—which recent epistemic democracy theorists have tried to show with the aid of the DTA model—is more defensible.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number: JP22KK0210, JP23K12410, JP20K13415, JP23H00526, JP21K01325.

Author Contributions

Conceptualization: Hélène Landemore, Ryota Sakai.
Systematic Review: Ryota Sakai.
Writing –original draft: Hélène Landemore, Ryota Sakai.
Writing –review & editing: Hélène Landemore, Ryota Sakai.

Reference cited

Aminpour, P. *et al.* (2021) ‘The diversity bonus in pooling local knowledge about complex problems’, *Proceedings of the National Academy of Sciences of the United States of America*, 118(5). Available at: https://doi.org/10.1073/pnas.2016887118.

Anderson, E. (2008). An Epistemic Defense of Democracy: David Estlund’s Democratic Authority. *Episteme: A Journal of Social Epistemology*, 5 (1), 129–139.

Baker, S.R., Bloom, N. and Davis, S.J. (2016) ‘Measuring Economic Policy Uncertainty’, *The Quarterly Journal of Economics*, 131(4), pp. 1593–1636. Available at: https://doi.org/10.1093/qje/qjw024.

Borg, A. et al. (2019). ‘Theory-choice, Transient Diversity and the Efficiency of Scientific Inquiry’, *European Journal for Philosophy of Science*, 9 (2), 1-26.

Brennan, J. (2014) ‘How Smart is Democracy? You Can’t Answer that Question a Priori’, *Critical Review*, 26(1–2), pp. 33–58. Available at: https://doi.org/10.1080/08913811.2014.907040.

Brennan, J. (2016) *Against Democracy*. Princeton, NJ: Princeton University Press.

Brennan, J. and Landemore, H. (2021) *Debating Democracy: Do We Need More or Less?* Oxford: Oxford University Press (Debating Ethics).

Dowding, K. and Lenine, E. (2023) 'Models, Conceptual and Predictive: A Response to Johnson's Models-as-Fables', *Perspectives on Politics*, 21(1), pp. 254–263. Available at: https://doi.org/10.1017/S1537592721002000.

Estlund, D.M. (2008). *Democratic Authority: A Philosophical Framework*. Princeton, NJ: Princeton University Press.

Goodin, R.E. and Spiekermann, K. (2018). *An Epistemic Theory of Democracy*. Oxford, New York: Oxford University Press.

Gough, D., Oliver, S. and Thomas, J. (eds) (2012). *An Introduction to Systematic Reviews*. London: SAGE.

Greco, D. (2023) *Idealization in Epistemology: A Modest Modeling Approach*. Oxford, New York: Oxford University Press.

Hankins, K., Muldoon, R. and Schaefer, A. (2023) 'Does (mis)communication mitigate the upshot of diversity?', *PLoS ONE*, 18(3 March). Available at: https://doi.org/10.1371/journal.pone.0283248.

Holman, B. *et al.* (2018) 'Diversity and Democracy: Agent-Based Modeling in Political Philosophy', *Historical Social Research*, 43(1), pp. 259–284.

Hong, L. and Page, S.E. (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), pp. 16385–16389. Available at: https://doi.org/10.1073/pnas.0403723101.

Jones, G. (2020). *10% Less Democracy: Why You Should Trust Elites a Little More and the Masses a Little Less*. Stanford, CA: Stanford University Press.

Kitchenham, B. and Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.

Knight, J. and Johnson, J. (2011). *The Priority of Democracy: Political Consequences of Pragmatism*. Princeton, NJ: Princeton University Press.

Landemore, H. (2013) *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton, NJ: Princeton University Press.

Landemore, H. (2014) 'Yes, We Can (Make It Up on Volume): Answers to Critics', *Critical Review*, 26(1–2), pp. 184–237. Available at: https://doi.org/10.1080/08913811.2014.940780.

Levinson, S. (2014) ‘A Welcome Defense of Democracy’, *Critical Review*, 26(1–2), pp. 92–100. Available at: https://doi.org/10.1080/08913811.2014.907043.

Lowi, T.J. (1964). American Business, Public Policy, Case-Studies, and Political Theory. *World Politics*, 16 (4), 677–715.

Mäki, U. (2005) ‘Models are Experiments, Experiments are Models’, *Journal of Economic Methodology*, 12(2), pp. 303–315. Available at: https://doi.org/10.1080/13501780500086255.

Mäki, U. (2020) ‘Puzzled by Idealizations and Understanding Their Functions’, *Philosophy of the Social Sciences*, 50(3), pp. 215–237. Available at: https://doi.org/10.1177/0048393120917637.

Moore, A. (2014) ‘Democratic Reason, Democratic Faith, and the Problem of Expertise’, *Critical Review*, 26(1–2), pp. 101–114. Available at: https://doi.org/10.1080/08913811.2014.907044.

Moore, A. (2016) ‘Deliberative Elitism? Distributed Deliberation and the Organization of Epistemic Inequality’, *Critical Policy Studies*, 10(2), pp. 191–208. Available at: https://doi.org/10.1080/19460171.2016.1165126.

Moore, A. (2017) *Critical Elitism: Deliberation, Democracy, and the Problem of Expertise*. Cambridge University Press (Critical Elitism: Deliberation, Democracy, and the Problem of Expertise). Available at: https://doi.org/10.1017/9781108159906.

OECD (2020). *Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave*. Paris: OECD Publishing. Available from https://doi.org/10.1787/339306da-en.

Page, S.E. (2008) *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, And Societies*. Princeton, NJ: Princeton University Press.

Page, S.E. (2015) 'Diversity Trumps Ability and The Proper Use of Mathematics', *Notices of the AMS*, 62(1), pp. 9–10.

Quirk, P.J. (2014) 'Making it up on Volume: Are Larger Groups Really Smarter?', *Critical Review*, 26(1–2), pp. 129–150. Available at: https://doi.org/10.1080/08913811.2014.907046.

Reijula, S. and Kuorikoski, J. (2019) 'Modeling Epistemic Communities', in *The Routledge Handb. of Soc. Epistemol.* Taylor and Francis, pp. 240–249.

Reijula, S. and Kuorikoski, J. (2021) 'The Diversity-Ability Trade-Off in Scientific Problem Solving', *Philosophy of Science*, 88(5), pp. 894–905. Available at: https://doi.org/10.1086/714938.

Reijula, S. and Kuorikoski, J. (2022). ‘Modeling Cognitive Diversity in Group Problem Solving’, In: *Proc. Annu. Meet. Cogn. Sci. Soc.: Cogn. Diversity, CogSci*. The Cognitive Science Society, 1863–1869.

Romaniega, Á. (2023) ‘Fatal mathematical errors in Hong-Page Theorem and Landemore’s epistemic argument’. arXiv. Available at: http://arxiv.org/abs/2307.04709 (Accessed: 12 July 2023).

Sakai, R. (2020) ‘Mathematical Models and Robustness Analysis in Epistemic Democracy: A Systematic Review of Diversity Trumps Ability Theorem Models’, *Philosophy of the Social Sciences*, 50(3), pp. 195–214.

Singer, D. (2018) ‘Diversity, Not Randomness, Trumps Ability’, *Philosophy of Science*, 86(1), pp. 178–191. Available at: https://doi.org/10.1086/701074.

Smaldino, P.E. *et al.* (2023) ‘Maintaining Transient Diversity Is a General Principle for Improving Collective Problem Solving’, *Perspectives on Psychological Science* [Preprint]. Available at: https://doi.org/10.1177/17456916231180100.

Somin, I. (2014) ‘Why Political Ignorance Undermines the Wisdom of the Many’, *Critical Review*, 26(1–2), pp. 151–169. Available at: https://doi.org/10.1080/08913811.2014.907047.

Suran, S., Pattanaik, V. and Draheim, D. (2020). Frameworks for Collective Intelligence: A Systematic Literature Review. *ACM Computing Surveys*, 53 (1), Association for Computing Machinery.

Takahashi, S. et al. (2013). Agent-Based Simulation of Diversity and Organizational Performance. In: Agent-Based Approaches in Economic and Social Complex Systems VII. Agent-Based Social Systems. Tokyo: Springer, 43. Available from https://doi.org/10.1007/978-4-431-54279-7_3.

Taleb, N. (2005). *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*. New York: Random House Trade Paperbacks.

Thompson, A. (2014) 'Does Diversity Trump Ability? An Example of the Misuse of Mathematics in the Social Sciences', *Notices of the AMS*, 61(9), pp. 1024–1030.

Vermeule, A. (2009). Many-Minds Arguments in Legal Theory. *Journal of Legal Analysis*, 1 (1), 1–45.

Weisberg, M. (2013) *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press (Oxford Studies in Philosophy of Science).

Weymark, J.A. (2015). Cognitive Diversity, Binary Decisions, and Epistemic Democracy. *Episteme*, 12 (4), 497–511.

Wimsatt, W.C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, Mass: Harvard University Press.

Zollman, K.J.S. (2010) 'The Epistemic Benefit of Transient Diversity', *Erkenntnis*, 72(1), pp. 17–35.