MATHEMATICAL ANALYSIS OF THE HONG-PAGE FRAMEWORK

ÁLVARO ROMANIEGA

ABSTRACT. This paper provides a mathematical analysis of the Hong-Page framework, a widely cited model in the collective intelligence literature in economics and philosophy that is often invoked in support of the claim that diversity trumps ability. We find that the framework exhibits significant limitations and that the claims made by Hong and Page are not fully substantiated by mathematical analysis. Our critique proceeds in three parts. First, we examine the mathematical theorems and show that they primarily restate the initial assumptions, relying on unrealistic premises that yield unrealistic outcomes of limited practical relevance. Moreover, these theorems can be simplified substantially and small adjustments to the framework lead to contradictory findings. Second, we investigate the simulations, which feature more prominently in the literature, and show that they only support the original claim under specific parameter choices—the only configurations presented in the original work by Hong and Page. Furthermore, when the framework is adapted to allow a more realistic or balanced comparison between ability and diversity, the simulations produce opposing results. Finally, we argue that the framework has broader shortcomings related to its focus on the initial point, characterization of the problem domain, and other underlying assumptions. In particular, it presumes agents with superrational capacities while simultaneously exposing them to simple framing effects that cannot be addressed. Overall, these issues constrain the framework's ability to offer meaningful insights into collective intelligence and may lead to misleading conclusions. We conclude by proposing directions for a more robust framework that should be built from scratch.

CONTENTS

Date: March 9, 2025.
TBD.

1. Introduction

In the intersection of economics, philosophy and political theory, one theorem has emerged as particularly influential: the Hong-Page Theorem. This theorem is claimed to show that under certain conditions, a diverse group of problem solvers has the potential to outperform a group of high-ability problem solvers. The theorem has been instrumental in shaping discourse around the importance of diversity in decision-making and problem-solving contexts. For instance, as pointed out by [GS20],

> It has been cited in NASA internal documents, offered in support of diversity requirements at UCLA, and appears in an *amicus curiae* brief before the Supreme Court in support of promoting diversity in the armed forces (Fisher v. Univ. of Texas 2016).

It has also been used in support of "diversity" instead of experts. For instance, in medicine, to support plural and alternative views instead of evidence-based medicine; see references in [Gri+19]. One notable application of the Hong-Page Theorem is found in the work of Hélène Landemore, who uses it as the cornerstone of her political proposal for an "Open Democracy." Landemore's thesis argues that greater cognitive diversity in a collective decision-making process not only enhances its epistemic properties, achieving the epistemic superiority of democracy, but also serves as a normative benchmark.

However, the theorem has not gone unchallenged. Mathematician Abigail Thompson has been among the most vocal critics, casting doubt on the use of the Hong-Page Theorem in this way and raising concerns about the validity of the conclusions drawn from it. In this paper, we aim to shed light on the Hong-Page Theorem, presenting our view that its findings may not hold the significant social implications often attributed to them and its central insight may be less informative than widely believed. By doing this, we aim to reveal the misconceptions propagated through their widespread acceptance, especially in their application to real-world socio-political phenomena, such as in Hélène Landemore's "Open Democracy" proposition and her epistemic argument for democracy, see [Rom23; Rom25] and references therein for more details.

Thompson's article ought to have highlighted the problems with the Hong-Page theorem. However, this has not been the case given its widespread acceptance, several rejoinders (particularly focusing on the simulation part) and papers building on Hong-Page's work. For instance, after the publication of [Tho14], Landemore continues to base her argument on the theorem, [BL21], for unreplied rejoinders to Thompson's critique see [Kue17] or [Sin19] and for examples of articles building on Hong-Page's work see [Wey15; Hol+18; Gri+19; Sak20; HMS23; Nie+24]. It is even included in the Stanford Encyclopedia of Philosophy [GS20]. For a recent review[1] of the literature on the Hong-Page theorem see [LS24]. Here, we aim to address some of these criticisms[2] and examine the use of the theorem by delving deeper into its theoretical aspects. Our new foundational critiques cast serious doubt on the Hong-Page theorem, as well as on some of the works that rely on it and their corresponding (mis)applications.[3]

[1] This review is problematic for several reasons, which will be discussed in different sections of this paper. For now, we note that it contains significant mischaracterizations of the work [Rom23] as well as clear omissions. For instance, although [Rom23] devoted considerable effort to analyzing Landemore's misapplication of the theorem and her so-called Numbers Trump Ability (NTA) Theorem (see Section 5, in particular Section 5.5), and this critique was clearly announced in the title, abstract, and introduction, the authors of [LS24, Table 3] claim that the NTA theorem is not addressed in [Rom23], as indicated by the column "Numbers trumps ability?" which reads "Not reported." They then conclude that the papers reviewed in [LS24] find evidence supporting Landemore's thesis, with no mention of the critique.

[2] We do not consider all critiques to be incorrect. For instance, we agree with the central thesis advanced in [Sin19].

[3] But not all. For instance, we agree with the analysis in [Gri+19], which we consider a far more rigorous treatment and a worthwhile starting point for understanding the simulation framework. Although we share their skepticism regarding the

Therefore, in this article, we introduce new arguments that underscore the theorem's simple insight and its misapplications. While Thompson's critique focused on illustrating the theorem's simplicity through specific examples, offering amendments to the theorem's statement, and assessing the role of simulations, this article examines several critical dimensions more deeply:

- First, at a mathematical level,
  - Demonstrating the theorem's simplicity by introducing and proving straightforward and non-probabilistic versions of it. These versions encapsulate all arguments related to diversity or ability, devoid of hypotheses not tied to diversity or ability, e.g., introducing clones and a random selection of them.
  - Highlighting the theorem's simplicity by offering a concise proof, based on standard mathematical results, thus calling attention to how the deployment of complex mathematics can sometimes obscure the straightforward truths.
  - Unveiling further, arguably unrealistic, consequences of the hypotheses of the theorem that cast doubt on their plausibility. For example, the random group always arriving at the correct solution unanimously.
  - Explicitly defining all the hypotheses and auxiliary assumptions which helps to illuminate how they seem to naturally lead to a predetermined conclusion, even if this was not the original intention of the theorem itself. This approach will be partly illustrated through the use of "counterexamples" to the theorem (without these assumptions), i.e., diversity does not trump ability when some technically tailored hypotheses are removed.
  - Enhancing the discussion further, we adopt a framework akin to the original, yet grounded in more empirically valid and theoretically sound assumptions (for example, excluding the use of clones). This refined approach maintains a formal resemblance to the original theorem, to the extent that it could be viewed as a sophisticated enhancement. However, it arrives at fundamentally different conclusions, demonstrating that when realistic hypotheses are used, ability significantly outweighs diversity.
  - Investigating other theorems by Hong and Page, particularly the "Diversity Prediction Theorem," to highlight similar issues. This theorem is often cited with the "Diversity Trumps Ability" Theorem.
- Second, analyzing the simulations in the Hong-Page framework. In particular,
  - We will show that the theorem does not explain the simulations and that the simulations are explained by a simple fact: by Hong and Page's design, the random group will be made of enough able agents (not far from experts), the best-performing agents will be similar to each other and the problems are such that they tend to favor the random group. As a result of this design, the random group will outperform the best group.
  - We will show that the simulations only support the original claim if the parameters are specific (the only cases presented in the original article [HP04]).
  - We will show that modifications of the parameters, which are somehow arbitrary, show that ability trumps diversity.
  - We will show that slight modifications of the framework or different frameworks to make a fairer comparison between ability and diversity or introduce realism show that ability trumps diversity. These new frameworks will be based on our previous analysis of the theorem, Section 4.

---

theorem's real-world applicability, our outlook is more pessimistic; we find the framework largely unhelpful and harbor deeper concerns about its misapplications in political philosophy.

- Finally, not taking the framework as given, but showing that the framework is flawed from a more general perspective:
  - The framework has an incorrect focus on the starting point, while it should be on the problem itself.
  - With the framework we could always design simple deliberations (even with a single agent) that always solve any problem.
  - The movement across the domain of the problem are unrealistic; agents can easily jump from the worst solutions to the best one (which should be a difficult task) but can be incapable of reaching the best solution when starting from a near-optimal solution, which should be a trivial task.
  - Agents are assumed, on the one hand, to be superrational (capable of evaluating the quality of any proposal, giving a complete ordering of the solutions for any problem. For instance, rank 2000 proposals based on their value with a 100% accuracy and unanimously) but, on the other hand, they are subject to simple framing effects that cannot be removed by any means (if we reorder the solutions in a different way, they will reach a different conclusion).
  - The shape of the problem has recieved a considerable amount of attention in the literature, but it is not an intrinsic property of the problem. In particular, we can reorder the solutions so that the easiest problem becomes the hardest and the hardest problem becomes the easiest.

The paper is organized as follows. We begin with a thorough dissection of the Hong-Page "Diversity Trumps Ability" Theorem in Section 2. We delve into the definitions that underpin these theorems and carefully examine the assumptions relating to the problems and problem solvers. We then derive and discuss a series of trivial corollaries from these assumptions, concluding with a concise analysis of other related results and a simplified proof of the theorem.

In Section 3, the paper takes a (more) critical turn, presenting a series of counterexamples when some arbitrary hypotheses are removed that challenge the robustness of the Hong-Page Theorem. We start by challenging the necessity of the injective function and the "unique best agent" assumption, eventually moving towards a discussion on the performance and selection of clones. We conclude that what the theorem requires is not "diversity", but the existence of a more "able" problem solver who can improve upon areas where others fall short.

The critique deepens in Section 4 with the proposition of a new Hong-Page style theorem: "Ability Trumps Diversity". This section revises the original assumptions to allow a fair comparison between ability and diversity, ultimately culminating in the presentation of this theorem. In Section 5 we shift focus onto the "Diversity Prediction Theorem" and the "Crowds Beat Averages Law", discussing their results, pointing out the asymmetric role of "ability" and "diversity" and highlighting a basic mathematical error made by Page in advocating for diversity. This is a related theorem often cited with the "Diversity Trumps Ability" Theorem.

In Section 6 we analyze in great detail the simulations in the Hong-Page framework with the same focus as the previous sections. In particular, there are two levels of analysis. If we use the same framework as the original one, we will see that the simulations only support the original claim if the parameters are chosen in a specific way (the only cases presented in [HP04]), otherwise ability trumps diversity. If we modify the framework to make more realistic or fairer comparison between ability and diversity, we will see that ability trumps diversity. Thus, if the original simulation framework had any value to derive conclusions for a real-world problem, it would likely support the claim that ability trumps diversity, contrary to the original claim. This phenomenon was common to the mathematical analysis of the theorem, but we see that in the simulations is even worse because the original claim is only supported for a specific set of parameters while maintaining the same framework. In any case, we do not believe that Hong-Page framework is useful to derive conclusions for a real-world problem. This is why in Section 7 we analyze the framework from a more general perspective, showing that the framework is flawed from its roots[4].

In conclusion, we show that the problems with the Hong-Page framework are not merely the ones related to a simplified model with imperfect assumptions (as all models are), but rather a framework that fundamentally fails to capture the essential relationships among diversity, ability, and effective group problem-solving. That is, it would be analogous to analyzing the motion of objects with a model that assumes there is no gravity, not just a simplified model like assuming there is no friction. The conclusions derived from a no-gravity model would be of little relevance to the real world, but those from a no-friction model would be a good starting point. Hong-Page's framework is more like a no-gravity model. The framework's rigid structure and unrealistic assumptions about how agents interact with problems make it unsuitable for understanding how to optimally compose groups to tackle real-world challenges. The framework misses crucial aspects of human expertise and group dynamics that are vital for actual decision-making and problem-solving. Furthermore, it leads to unrealistic conclusions that are not supported by empirical evidence or a minimal amount of realism. All in all, this model seems to be a good example of a model that is not effective and should be avoided. It provides little insight, logical evidence or even evidence to update our priors in a Bayesian way, in contrast to other theorems on the field [Rom22], about how to optimally compose groups to solve real-world problems. All work based on this framework needs to be rethought, as it stands on shaky ground.

A final caveat, this critique should not be taken as a dismissal of the importance of diversity (which we consider to be one important epistemic factor among others) in decision-making, but rather as a call to address the misuse of mathematics in these contexts. It urges us to consider the rigorous and nuanced approach required when applying mathematical theories to sociopolitical constructs. As such, this paper makes a contribution to the ongoing discourse on collective intelligence, fostering a deeper understanding of the mathematical theorems used.

[4]Thus, although the Hong-Page theorem is often invoked to support democracy, if it were truly applicable, it would favor an epistocratic arrangement. A similar dynamic occurs with other theorems frequently cited in the literature—such as the Condorcet Jury Theorem (CJT) and the Miracle of Aggregation [Rom22]. As discussed in Remark 2.11 and Section 7, however, the Hong-Page framework is not a sound starting point, unlike the CJT. Moreover, from a purely epistemic perspective, all of these theorems naturally suggest the need for ability thresholds or epistemic weights to ensure their applicability, again, more in favor of epistocracies. See [Rom22] for a lengthier discussion.

2. The "Diversity Trumps Ability" Theorem

This section introduces the key definitions and assumptions utilized in our analysis of the Hong-Page problem-solving framework.

2.1. Definitions.
Here, we define what constitutes a problem solver and the nature of problems they tackle.

Definition 2.1 (Problem solver)**.** Let X be a set of possible states of the world. For simplicity, X is assumed to be a finite set. A *problem solver* is a function $\phi : X \to X$. We will consider several problem solvers throughout this analysis. The set of problem solvers is denoted by Φ.

Definition 2.2 (Problem)**.** For each problem solver ϕ, define the value function as $V_\phi : X \to [0, 1]$ as a function mapping states of the world to real numbers between 0 and 1. The *problem* is finding the maximum of V_ϕ. Given a probability measure on X with full support, ν, the expected value of the performance of each agent is given by

$$\mathbb{E}_\nu(V_\phi \circ \phi) = \sum_{x \in X} V_\phi(\phi(x))\,\nu(x)\,. \tag{2.1}$$

Let's discuss the intuition behind the problem-solving process. Given an initial state[5] x from the set X, a problem solver aims to find a solution to the problem by mapping x to $\phi(x)$. In other words, the problem solver transforms the input x into a potential solution, hoping that this transformation will maximize the value of the function V_ϕ.

2.2. Problem assumptions.
Here, we state the assumptions implicit in the Hong-Page framework. In particular, we are going to impose some conditions on the set Φ and on the problem following Hong-Page's assumptions.

Assumption 1 (Unique problem)**.** All problem solvers share the same value function, i.e., $\forall \phi, \phi' \in \Phi$, $V_\phi = V_{\phi'} =: V$.

In other words, any two problem solvers evaluate the problem space in the same way. This assumption simplifies the analysis by ensuring that all problem solvers have a consistent evaluation criterion for the problem. See Remark 2.3 for why this assumption is not shared in standard political philosophy.

Remark 2.3. Assumption 1 appears to restrict the scope of problems to those of a technical nature, where the diversity of values might not be significant. However, the domains to which the theorem has been proposed for application, such as democracy, embody a profound component of political philosophy in which a diversity of values is unavoidable. Rawls, in particular, highlights "reasonable pluralism" as a fundamental element of his concept of a well-ordered society. Gaus further critiques this idea by arguing that reasonable pluralism not only encompasses varied conceptions of the good life but also extends to divergent views on justice itself. Gaus asserts [Gau10]:

> Kant, Hare, and Rawls all believed that if we set up the constraints on moral legislation properly, the optimal eligible set will be a singleton. Only one rule can meet the constraints on right, properly understood. However, in our model, reasonable pluralism is taken seriously: Members of the Public are applying their interpretations of the test based on different evaluative standards and beliefs. There is no reason to suppose that, except at a very abstract level (§17), the results of the application of the same tests, by different Members of the Public, based on different evaluative

[5] This is the equivalent of the guess value or starting value in an optimization algorithm, but see Section 7 for why this is not a good analogy.

> standards, will yield the same answer. As we saw, having diverse evaluative standards, they will rank the proposals differently based on these standards. There is no reason to suppose that, out of all this diversity, Members of the Public will converge on the same proposal.

In the context mentioned above, this could be interpreted as, according to Rawls, the V_ϕ functions may differ, yet there exists a unique optimum, x^*, such that $V_\phi(x^*) = 1$ for all "reasonable" ϕ, whereas for Gaus, not only do the V_ϕ functions do not equal a common V, but there also does not exist the singleton $\{x^*\}$. Consequently, in both cases, the assumption is not met.

This recognition of pluralism as an intrinsic aspect of social structure raises questions about the theorem's applicability without additional qualifications. Nonetheless, this analysis primarily examines the mathematical premises and their wider implications, temporarily setting aside discussions on value diversity to concentrate, for the moment, on the mathematical critique.[6]

◇

Assumption 2 (Unique solution)**.** There exists a unique[7] state of X that maximizes the value function V, i.e., $\exists_{=1} x^* \;/\; V(x^*) = 1\,.$

In other words, there is exactly one optimal solution to the problem that maximizes the value of the function V. This assumption allows us to focus on finding the unique solution.

Assumption 3 (Strictly increasing problem)**.** V is injective, i.e., if $V(x) = V(x')$, then $x = x'$. That is, we can order X as $\{x_1, \dots, x_{|X|}\}$ such that

$$V(x_1) < \ldots < V(x_{|X|})\,.$$

In other words, this assumption implies that the problem has a well-defined ordering of potential solutions. The original article did not state explicitly that the value function V is one-to-one. However, this assumption is necessary for the theorem to hold, as Thompson pointed out, [Tho14], see Section 3.1 for more details.

2.3. **Problem solver assumptions.**

Assumption 4 (Everywhere ability in problem solvers)**.** For all problems solvers $\phi \in \Phi$ it holds that $V(\phi(x)) \geq V(x)$ for every $x \in X$. In particular, $\phi(x^*) = x^*\,.$

This assumption states that all problem solvers are able to improve (with non-strict inequality) the value of any state. Furthermore, by Assumption 3, if $\phi(x) \neq x$, $V(\phi(x)) > V(x)$. In other words, if a problem solver is applied to a state, the value of the state will never decrease. Furthermore, if the agents transforms the state, the value will strictly increase.

Assumption 5 (No improvement, idempotence)**.** $\forall \phi \in \Phi$, $\phi \circ \phi = \phi\,.$

This assumption states that problem solvers are idempotent. In other words, applying a problem solver to a state twice will have the same effect as applying it once, without affecting the value of the candidate/potential solution.

Assumption 6 ("Difficulty", imperfect problem solvers)**.** For every problem solver ϕ there exists a state x such the optimal solution, x^*, is not returned, i.e., $\forall \phi \in \Phi\; \exists x \in X \;/\; \phi(x) \neq x^*\,.$

---

[6] Thus, applying the theorem to democracy makes little sense, as argued in [LS24]. The theorem assumes there is no value diversity, but as we have seen, value diversity is fundamental to political philosophy. Therefore, political applications of this theorem are flawed from the start.

[7] Hereafter, the subscript in the existential quantifier denotes the cardinality of the subset of elements satisfying a given property. For instance, $\exists_{=1}$, sometimes denoted as $\exists!$, means that there is only one element satisfying the required condition.

In other words, by hypothesis, for every agent there are instances where they fail to find the optimal solution.

Assumption 7 ("Diversity", sufficient unstuck problem solvers)**.** For every state $x \neq x^*$ there exists a problem solver ϕ such that it returns a different state of the world, i.e., $\forall x \in X \backslash \{x^*\}\ \exists\ \phi \in \Phi\ /\ \phi(x) \neq x$.

In other words, this assumption ensures[8] a "diversity" of problem solvers in Φ, with at least one problem solver capable of making progress from any non-optimal state.

Assumption 8 (Unique best problem solver)**.** $|\arg\max_{\phi\in\Phi}\{\mathbb{E}_\nu(V \circ \phi)\}| = 1$, i.e., there is only one best-performing agent according to (2.1).

In other words, this assumption states that there is only one problem solver that performs best on average. There is only one problem solver that is the most likely to lead to the optimal state.

2.4. **Problem solver interaction assumptions.** Here, we turn to the interaction between problem solvers.

Assumption 9 (In series deliberation)**.** The problem solver or agents $\Phi' := \{\phi_1, \ldots, \phi_{N_1}\}$, when working together to solve the problem starting at x are equivalent to the following sequence[9]

(1) First, a problem solver i_1 such that $x_1 := \phi_{i_1^x}(x) \neq x_0 := x$.
(2) Second, a problem solver i_2 such that $x_2 := \phi_{i_2^x}(x_2) \neq x_1$.
(3) Inductively, a problem solver i_j such that $x_j := \phi_{i_j^x}(x_{j-1}) \neq x_{j-1}$.

This stops at x_τ, returning that value, such that it is a fixed point for all elements of $\{\phi_1, \ldots, \phi_{N_1}\}$ (all the agents are stuck at the same point, unanimity).

Several remarks are in order. There can be multiple sequences arriving at the same point. The fixed point exists as x^* is a fixed point for all elements of Φ by assumption. The group performance is tantamount to composition of the functions in a proper way:

$$\phi^{\Phi'}(x) := \phi_{i_\tau^x} \circ \ldots \phi_{i_1^x}(x)\,.$$

In other words, this assumption states that a group of problem solvers can be thought of as a sequence of agents that takes turns applying the problem solvers in the group to the current state, such that we approach the optimal value. The group will stop when it reaches a state that is a fixed point for all of the problem solvers in the group, i.e., unanimity.

Note that i_x^k is not defined, and unless x^* is the only common fixed point of all agents, the final group output could be undefined. We will explore some specifications for simulation analysis later, but they are not needed for what follows. This is the case because the Hong-Page framework leads to comparing groups where the only common fixed point is the global optimum, or groups with only one different agent. To clarify the situation[10], let us prove the following proposition.

Proposition 2.4. *Assumption 9 is equivalent to returning an element of X belonging to the set of fixed points of all the problem solvers in the group.*

[8] As Hong and Page put it.

[9] In the simulation part, Section 6, we will use this assumption to simulate the group deliberation allowing an equality instead of an inequality for every agent.

[10] I would like to thank a referee of an earlier version of this article for pointing out that the following equivalence was not clear as he/she pointed out that the condition was different from the one appearing in [Wey15], although it is the same condition.

Proof. Assumption 9 implies that $\phi^{\Phi'}(x) = x_\tau \in \bigcap_{i=1}^{N_1} \phi_i(X)$, so $\phi^{\Phi'}(x)$ is a fixed point for all the problem solvers in the group. Conversely, for any $x \in \bigcap_{i=1}^{N_1} \phi_i(X)$ there exists at least one sequence of agents such that $x_\tau = x$. Indeed, $\phi^{\Phi'}(x) = x$ is a fixed point for all the problem solvers in the group. □

Thus, specifying the sequence of agents in Assumption 9 for every x is equivalent[11] to specifying the the fixed point chosen by the group. Consequently, specifying the probability distribution over the sequence of agents is equivalent to specifying the probability distribution over the choosen fixed points, which is denoted as η in the original paper [HP04] or $\eta_{\Phi'}$ in [Wey15]. But, as we will see later, this probability distribution is not needed for the analysis of the Hong-Page theorem nor for the analysis of the simulations. Our sequence formulation is more useful for the analysis of the simulations, see Section 6, and will be useful to enhance the theorem and introduce disagreement, see Section 4.3.

Assumption 10 (Clones)**.** There exists an infinite amount of identical copies of each agent $\phi \in \Phi$.

This assumption implies that an infinite number of each problem solver is available. This condition is necessary for the theorem to hold, as will be discussed in Section 3.3 and 3.4. Essentially, when the probability apparatus is introduced by Hong and Page, the selection criterion implies that an agent can be chosen multiple times, with repetition, indicating that multiple copies of each agent should be available.

2.5. **An example.** To give an example, let $X = \{a, b, c, d\}$ and $\Phi := \{\phi_1, \phi_2, \phi_3\}$ such that

x	$V(x)$	$\phi_1(x)$	$\phi_2(x)$	$\phi_3(x)$
a	$1/4$	b	a	b
b	$1/2$	b	c	b
c	$3/4$	d	c	c
d	1	d	d	d

.

For instance, this means that $V(a) = \frac{1}{4}$ or $\phi_2(b) = c$. Then, the hypotheses of the theorem are satisfied. Indeed, the problem is:

(1) **Problem**. The task is to maximize the value $V(x)$. Looking at the table, the objective is to find d, which has a value of 1.
(2) **Problem solver**. We have three problem solvers, ϕ_1, ϕ_2, and ϕ_3. These solvers transform an initial state x (one of the values a, b, c, d) into another state, which could be the same or different.

The hypotheses:

(1) **Unique problem evaluation**. All problem solvers use the same value function $V(x)$ to assess the desirability of a solution.
(2) **Unique solution**. There is a singular best solution, which is d, having a value $V(d)$ equal to 1.
(3) **Strictly increasing problem**. The values in $V(x)$ are distinct and can be ranked in increasing order: $V(a) = \frac{1}{4} < V(b) = \frac{1}{2} < V(c) = \frac{3}{4} < V(d) = 1$.
(4) **Everywhere ability in problem solvers**. When applied to any state, a problem solver either maintains the value or improves it. For instance, ϕ_1 strictly increases the value of the states $\{a, c\}$ and maintains the value of $\{b, d\}$.

[11] x choosen with a probability distribution ν', which might be different from ν, but otherwise stated, we will assume that it is the same as ν. In particular, we will assume this throughout the simulation in Section 6.

(5) **No improvement, idempotence**. If you apply a problem solver to a state repeatedly, the outcome remains the same. For example, applying ϕ_1 to a gives b and, if applied again, gives b too.
(6) **"Difficulty", imperfect problem solvers**. Each solver has some limitations. For instance, neither ϕ_1 nor ϕ_3 can transform b to d.
(7) **"Diversity", sufficient unstuck problem solvers**. For every non-optimal state, at least one problem solver can make progress. For example, for state b, ϕ_2 can improve it to c.
(8) **Unique best problem solver**. Let's compute the expected value for each problem solver using the uniform distribution over $\{a, b, c, d\}$. As we see, ϕ_1 is the best problem solver.
(9) For ϕ_1:
   - $\mathbb{E}(V \circ \phi_1) = \frac{1}{4} \cdot (V(b) + V(b) + V(d) + V(d)) = \frac{1}{4}\left(\frac{1}{2} + \frac{1}{2} + 1 + 1\right) = \frac{3}{4}$.
   - $\mathbb{E}(V \circ \phi_2) = \frac{1}{4} \cdot (V(a) + V(c) + V(c) + V(d)) = \frac{1}{4}\left(\frac{1}{4} + \frac{3}{4} + \frac{3}{4} + 1\right) = \frac{11}{16}$.
   - $\mathbb{E}(V \circ \phi_3) = \frac{1}{4} \cdot (V(b) + V(b) + V(c) + V(d)) = \frac{1}{4}\left(\frac{1}{2} + \frac{1}{2} + \frac{3}{4} + 1\right) = \frac{11}{16}$.

(10) **In series deliberation**. If solvers work together, they iteratively improve a state. For instance, consider the sequence of problem solvers $\phi_1, \phi_2, \phi_1, \ldots$. Starting with a, if we first apply ϕ_1 and then ϕ_2, we reach the state c. If we apply ϕ_1 again, we reach d.
(11) **Clones**. Imagine having infinite copies of each ϕ. This would mean we can always use any solver as needed, even if multiple instances of the same solver are required. As we said, when the probability apparatus is introduced by Hong and Page, the selection criterion implies that an agent can be chosen multiple times, with repetition, indicating that multiple copies of each agent should be available.

2.6. **Trivial corollaries from the assumptions.** By construction (i.e., from the assumptions and *not* the theorem), we have the following corollaries. Note that no profound or even standard mathematical results are needed; we just need the assumptions mentioned above combined with trivial arithmetic and trivial properties of sets.

Corollary 2.5. *All members of Φ working together can solve the Problem $\forall\, x \in X$.*

Proof. This follows straightforwardly from Assumptions 3, 4, 7, and 9 such that,

$$V(x) < V\left(\phi_{i_1^x}(x)\right) < \ldots < V\left(\phi_{i_n^x}(x_{n'-1})\right) = 1\,,$$

for some $n' \leq |X|$. Indeed, by Assumption 7 there is an agent which transforms the previous state, starting at x and assuming it is not x^*, to a different one. By Assumption 3 and 4, the value is strictly superior to the one of the previous state. This can be done till they reach x^*. □

Corollary 2.6. *There exists a state $x \in X$ such that $\forall N_1 \in \mathbb{N}$, N_1 "clones" of the the best performing agent cannot solve the Problem.*

Proof. By Assumptions 5 and 9, N_1 "clones" of the best performing agent work in the same way as a single clone alone. By Assumption 6, no agent can solve the problem alone for some x, so the corollary follows straightforwardly. □

So, just "rearranging" the assumptions we can trivially (again, no profound mathematics, just basic arithmetic and trivial properties of sets) prove the following version of the "theorem" containing the main conclusion. The advantage of this formulation is that no clones are needed, compare with Theorem 2.10.

Theorem 2.7 (Basic Hong-Page theorem)**.** *Given the Assumptions 1-9, all problem solvers (or a selected subset of them) working together perform better than the best problem solver (in the sense that there is a state*

x such that the best problem solver cannot solve but the whole group can). Note that the best problem solver can be included in the first group.

Proof. This follows straightforwardly from the assumptions outlined above. Specifically, a proper subset of Φ, as described in Corollary 2.5, can always solve the problem, whereas ϕ^* cannot, as indicated in Corollary 2.6. Note that, according to Assumption 4, including more agents does not worsen the state in terms of its value; the state either improves or remains the same. $\square$

Again, this result is a simple corollary given the way we have formulated our hypotheses. Let us explain the proof in words. By assumption, we have arranged for the best agent not to always solve the problem. On the other hand, by mere assumption, for every state, there is an agent that can reach a better state. As the number of states is finite, they will reach the optimum in a finite number of steps. Note that the "diverse group" includes the best agent. The fact that the "diverse" group" outperforms the best agent is trivial, as they include additional agents and, by hypothesis, they do not worsen the solution and improve it for some states.

Actually, we do not need the whole group Φ to outperform the best performing agent. As $\mathbb{E}\,(V;\Phi) = 1$, it is evident that there exists[12] a $\tilde{\Phi}_0 \subseteq \Phi$ such that

$$\mathbb{E}\left(V;\tilde{\Phi}_0\right) > \mathbb{E}\left(V;\{\phi^*\}\right) < 1\,. \tag{2.2}$$

That is the "selected group" of the statement of Theorem 2.7. To connect with the original statement of Hong and Page, let us formulate the following version.

Theorem 2.8 (Deterministic Hong-Page's Theorem)**.** *Given the hypotheses above and two natural numbers $0 < N_1 \leq N$:*

- $(\mathcal{A}_1)$ ***"Ability-diversity" assumptions****. The assumptions given above, Assumption 1-9.*
- $(\mathcal{A}_2)$: ***"Counting clones" assumptions****. Assumption 10 and we choose two groups from a selected pool with N clones of Φ that makes the following possible. In the first group there are N_1 clones such that $\tilde{\Phi}_0 \subset \{\phi_1^R, \ldots, \phi_{N_1}^R\} =: \Phi_R$, and, in the second, there are N_1 clones of the best performing agent, Φ_B.*

Then, the performance of Φ_R is better than Φ_B in the sense of (2.1)*.*

Proof. It is trivial by the corollaries given above. Indeed, by (2.2), $\mathbb{E}_\nu(V \circ \phi^{\Phi_R}) > \mathbb{E}_\nu(V \circ \phi^{\Phi_B})$. Note that by Corollary 2.6, $\mathbb{E}_\nu(V \circ \phi^{\Phi_B}) < 1$ as $\exists_{\geq 1}\, x$ such that $V(\phi^{\Phi_B}(x)) < 1$. $\square$

In other words, as detailed in Appendix A, the theorem has the following structure with an evident proof:

Pseudo-Theorem 2.1. *Hypotheses*: Let there be a pool of agents such that:

(1) they never degrade the solution and where there is always an agent who can improve the solution,
(2) the best performing agent is imperfect by hypothesis,
(3) as it follows from these two points above, we can choose agents randomly until they outperform the imperfect best performing agent.

Thesis: Consequently, this "random" group will outperform the best performing agent.

Remark 2.9. In Theorem 2.8 two points warrant attention. Firstly, we could have consider a subset Φ_0 such that $\forall\, x \in X$, agents reach x^* consistent with Assumption 9, i.e., $\mathbb{E}\left(V;\tilde{\Phi}_0\right) = 1$, so (2.2) is trivially satisfied. For simplicity, one might prefer considering Φ over Φ_0, as both groups consistently achieve the correct solution. Unneeded elements can be omitted without loss of generality.

[12] See Remark 2.9 for several comments on different definitions of Φ_R.

Given the uncertainty in the selection process, the only definitive method to include all indispensable agents is to encompass all. Whether we take into account $\Phi_0 \subset \Phi_R$ or $\Phi \subset \Phi_R$ has no bearing on our arguments.

Secondly, in [HP04] and above, N_1 is an integer such that

$$\mathbb{E}(V; \Phi_R) > \mathbb{E}(V; \{\phi^*\}),$$

which translates to $\tilde{\Phi}_0$ being a subset of Φ capable of outperforming the best-performing agent. As we said, this subset's existence is a straightforward outcome of Corollary 2.5. Yet, selecting N_1 such that $\mathbb{E}(V; \Phi_0) = 1$ provides a more encompassing result (as seen in Proposition 2.12), which inherently contains the result in [HP04]; if an N_1 exists where the correct solution is invariably attained, then due to its monotonic nature, there must also be an N_1 where the alternative imperfect group is outperformed. Again, the distinction between considering $\tilde{\Phi}_0 \subset \Phi_R$ or $\Phi_0 \subset \Phi_R$ is mostly irrelevant. For ease and based on the aforementioned discussion, we will also use Φ or Φ_0 in subsequent sections. Further details can be found in Appendix A. $\diamond$

So, what did Hong and Page prove in their article? Essentially, they demonstrated that assumption $\mathcal{A}_2$ holds almost surely (after defining some probability measures). See their proof of Lemma 1 and Theorem 1 in [HP04]. This is a not very difficult probabilistic claim that has nothing to do with either ability or diversity, which are contained in the assumptions $\mathcal{A}_1$. It is a probabilistic fact that can be shown, regardless of whether the objects considered are diverse agents, incapable problem solvers, balls in a box, or mathematical functions in a Hilbert space[13]. Nevertheless, this is the heart of proof of their article published in the Proceeding of the National Academy of Sciences. But one might question, given that we have shown that Theorem 2.7 (a trivial restatement of the assumptions) encapsulates all the information regarding diversity and ability, what is the necessity of introducing clones? This is why Thompson say that the theorem "is trivial[14]. It is stated in a way which obscures its meaning. It has no mathematical interest and little content." We can compare the previous version, Theorem 2.8, with the original statement:

Theorem 2.10 ([HP04])**.** *Given the assumption above, let μ be a probability distribution over Φ with full support. Then, with probability one, a sample path will have the following property: there exist positive integers N and N_1, $N > N_1$, such that the joint performance of the N_1 independently drawn problem solvers exceeds the joint performance of the N_1 individually best problem solvers among the group of N agents independently drawn from Φ according to μ.*

As noted by Thompson [Tho14], the theorem as originally stated was false because Assumption 3 was not included (see Section 3.1 for details). Note, too, that "N_1 individually best problem solvers" refers merely to clones of the best problem solver (which is unique under the assumptions), rather than, say, the top two problem solvers by expected value (which would perform better). This restriction is imposed by hypothesis, even though Page denies it in [Pag07] when disseminating the theorem. On p.158 he states:

> The diversity-trumps-ability claim assumes lower average ability for the collection of diverse problem solvers. It also allows for variation among the collection of the best problem solvers. We're not assuming that every person in the group of the best is identical. We're only assuming that they're good.

---

[13] That is, the probability apparatus is introduced to ensure that with probability one, we have the groups described in $\mathcal{A}_2$, as further discussed in Section 3.4. This approach could be applied to any object, unrelated to diversity, and would be simplified by directly stating $\mathcal{A}_2$ or, even better, presenting a clone-free version of it, as in Proposition 2.7. In Section 4, detailed in Section 4.3, we will employ probability in a manner pertinent to an actual deliberative process.

[14] See Remark 2.11 on why the same cannot be said of other theorems, like the Condorcet Jury Theorem.

This is *clearly false*: they choose N such that the best-performing agents are merely clones of the best-performing agent, and hence identical.

More precisely, we need a new assumption about the selection criteria for the members of the comparison groups employed in the argument.

Assumption 11. By hypothesis we have the following.

(1) The first group is selected randomly from an infinite pool of clones of the elements in Φ. The group size N_1 can be adjusted as required.
(2) Similarly, the second group is chosen independently, but from an identically distributed set of clones of the elements in Φ of size N_1. This selection process follows the stipulations that:
   - the group size N can be adjusted as required,
   - the selection allows for the repetition of the best problem solvers.

Remark 2.11. One could say that the Condorcet Jury Theorem (CJT) is as trivial[15] as Theorem 2.10 and subject to similar critiques as the one presented above, but this would be an unfair comparison. The Hong-Page theorem and the CJT both try to contribute to our understanding of collective decision-making, yet they do so in markedly different ways. The Hong-Page theorem, despite its initial presentation of mathematical sophistication, as we have seen, essentially boils down to a self-evident proposition once its layers are peeled away. Its hypotheses—assuming agents never degrade a solution and that there exists always an agent who can improve upon it, coupled with the selection of agents until they surpass an already acknowledged imperfect best performer—do all the heavy lifting. Once these assumptions are made explicit, the theorem's conclusion becomes almost self-evident. Mathematics is not needed to understand that if it is assumed that the best agent is imperfect and the "random group" can be made perfect, the random group will outperform the best agent. Furthermore, when the theorem is adjusted to include more realistic characteristics, the original conclusion shifts dramatically, see Section 4, indicating the frailty of its insights under slight modifications.

In stark contrast, the CJT, even in its most basic form, delivers insights into the epistemic strength of democracy through collective decision-making. The theorem demonstrates how, under certain conditions, a large group of voters, each with a probability of being correct greater than chance, can collectively make decisions that converge to the correct choice with increasing certainty as the group size grows. While the proof of CJT utilizes probabilistic tools that may require university-level mathematics to fully grasp, it remains accessible and mathematics is essential to fully capture this insight. Moreover, the CJT has been the foundation for extensive refinement and expansion in

[15] [LS24] egregiously conflate the notion of a theorem being trivial—that is, easy to understand—with being tautological ($p \to q$ is a tautology). They say:

> 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 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 in 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.

For example, while Condorcet's theorem (say, for non-homogeneous voters) or Arrow's theorem are tautological in the sense that their conclusions are built into their premises, arriving at and understanding these theorems requires nontrivial reasoning and subtle logical steps. In contrast, the Hong-Page theorem is not only tautological but also straightforward, rendering it easy to grasp (cf. Section 2.7.1). Although it might not be trivial for those not well-versed in mathematics, its fundamental idea is still trivial (cf. Pseudotheorem 2.1). This is not the case for other theorems, which require nontrivial reasoning and subtle logical steps beyond proper mathematical formalization. Although the fact of being trivial is subjective, we can agree that $p \to p$ is trivial and Arrow's impossibility theorem is not. The Hong-Page theorem is more like $p \to p$, cf. Proposition 2.7.

political theory and economics literature, adapting to more complex scenarios such as correlated votes and non-homogeneous preferences [Rom22] and references therein, using mathematics to understand the conditions under which the insight works, thereby enhancing its applicability and relevance[16].

The comparison between the Hong-Page theorem and the CJT is thus not just in their mathematical complexity but in the depth of insight each provides. While the former wraps a simple fact in unnecessary mathematical additions and dramatically unstable under changes of the tailored hypotheses, the latter offers a compelling argument for the power of collective decision-making. Here, mathematics, in contrast to natural language, is essential for fully understanding the insight and the conditions for applicability in more complex cases. This contrast underscores the importance of critically examining the assumptions and implications of mathematical theorems in social sciences, distinguishing between those that offer genuine insights and those that can be reduced to simple facts when the unnecessary mathematics are removed. $\diamond$

2.7. Other results and simpler proof.

Proposition 2.12. *Assuming the conditions of Theorem 2.10 with N_1 large enough, with probability one,*

(1) the randomly selected group of N_1 problem solvers will invariably converge on the correct solution without any disagreement and unanimity,

(2) the "random group" always contains the best-performing agent.

These facts explain that they can always outperform the best problem solvers.

Proof. This is straightforward from Corollary 2.5 and that, following the first part of Assumption 11, the first group includes a copy of Φ or Φ_0 (Remark 2.9) μ- almost surely. It is also Lemma 1 in [HP04]. There is unanimity as, for every state $x \in X$, the group solution is x^*, where everyone accepts as a solution, $\phi(x^*) = x^*$. Even if the group solution is not x^*, there is unanimity as the group solution is the local optima of every agent, see Assumption 9. The second statement follows from the Strong Law of Large Numbers, see *Simpler proof* in Section 2.7.1. $\square$

For instance, the following $\Phi := \{\phi_1, \phi_2, \phi_3\}$ such that

x	$V(x)$	$\phi_1(x)$	$\phi_2(x)$	$\phi_3(x)$
a	$1/4$	b	a	b
b	$1/2$	b	c	b
c	$3/4$	d	c	c
d	1	d	d	d

satisfies the hypotheses of the theorem, but, if the "random" group does not include the best performing agent[17], ϕ_1, then it cannot outperform ϕ_1.

That is, given the same hypotheses, one could simply select the N_1 of Proposition 2.12 and assert a stronger result. It is not just that the random group outperforms the best agent, but they also almost surely find the optimal solution. So, with the same hypotheses, one can prove a stronger result. Shouldn't one then fully explore the implications of these hypotheses and reject them if they lead to unrealistic conclusions? That is, given that $N_1 = N_1(\omega)$ is *not fixed* but rather a "stopping time" that halts when the desired members, as per the selector's or chooser's preference, are included, one might ask why the selector should stop merely to outperform the best agent if the

[16] However, as we have criticized elsewhere [Rom22], the theorem is also often misused to reach conclusions that it cannot support.

[17] Assumptions can be made to exclude the best performing agent, while ensuring that there is another agent that performs as the best one does when needed. Consequently, it is no surprise that the theorem still holds. However, this approach is purely ad hoc.

correct solution can always be reached when the hypotheses are met. This more comprehensive result trivially implies the one in Theorem 2.10. See Appendix A for more details.

2.7.1. *Simpler proof.* In fact, a simpler proof of the theorem can be constructed based on that simple fact. This approach also exposes *the theorem's triviality given its underlying assumptions*.

Proof of Theorem 2.10. First, by hypothesis (Assumptions 6 and 7), $\exists\, x_* \in X$, $\phi^*, \phi_* \in \Phi$ such that the best agent $\phi^*(x_*) \neq x^*$ and $V(\phi_*(\phi^*(x_*))) > V(\phi^*(x_*))$. By hypothesis (Assumptions 9 and 10), $V \circ \phi^{\{\phi^*,\, \phi_*\}} \geq V \circ \phi^*$, where the equality is strict for at least one point. Given that ν has full-support, $\mathbb{E}_\nu\left(V \circ \phi^{\{\phi^*,\, \phi_*\}}\right) > \mathbb{E}_\nu\left(V \circ \phi^*\right)$.

Second, we introduce the probabilistic selection of clones. By the Strong Law of Large Numbers (SLLN),

$$\mu\left(\omega \in \Omega \,:\, \bigcap_{\phi}\left(\lim_{N\to\infty} f^N(\phi) = \mu\left(\{\phi\}\right)\right)\right) = 1\,,$$

where $f^N(\phi)$ represents the frequency of appearance of ϕ when the size of the group of clones is N. The intersection is finite. For this full-measure set, we define $N_\phi = N_\phi(\omega)$ as the integer such that, if $N \geq N_\phi$, then $f^N(\phi) > \mu(\{\phi\})/2$. Following Assumption 11, we take $N_1 := \max\{N_{\phi^*}, N_{\phi_*}, \frac{2}{\mu(\{\phi^*\})}, \frac{2}{\mu(\{\phi_*\})}\}$ for the first event. By these definitions, at least one copy of ϕ_* and ϕ^* are included. For the second event, take $N \geq \frac{2}{\mu(\{\phi^*\})} N_1$, so there are more than N_1 copies of ϕ^* in the second group. The proof then follows from the first part of this argument. $\square$

An intuitive explanation goes as follows. The first paragraph of the proof in the simpler proof corresponds to the part of the theorem where diversity and ability are put into play. This essentially reduces to the following triviality: by assumption, there are two distinct agents - the best agent and another agent - and a state x_* for which the best agent does not provide the optimal solution. However, the other agent can improve upon the solution of the best agent for this state. This implies that the performance of a group consisting of the best agent and this additional agent surpasses the performance of the best agent alone, at least for some states. For other states, again by assumption, adding an agent does not worsen the situation, thus completing the deterministic clone-free part of the proof. Subsequently, we apply the strong law of large numbers to ensure that, under the setting of Assumption 11, the random group will always contain copies of these two agents, and the best-performing agents are all copies of the unique best-performing agent.

3. Removing technical hypotheses: Counterexamples

The theorem depends critically on certain assumptions that we are going to analyze now. In this section, we will refrain from critiquing certain empirical hypotheses, such as the assumption that agents share the same concept of problem-solving (Assumption 1, although see Remark 2.3), or that they can recognize the solution ($\phi(x^*) = x^*$). Such critiques largely pertain to the plausibility inherent in every model, and one could always defend these by invoking ideal conditions, much as one might assume frictionless systems in physics, but see Section 7. Although these critiques are adequate, a different critique, following a "Moorean style," is presented in [Rom23; Rom25] using Section 4. There, some empirical hypotheses (not the ones mentioned above) are revised, slightly modifying them to enhance their plausibility, which leads to contrary conclusions if one tries to apply the "Diversity Trumps Ability" theorem. However, in this section, we wish to focus on certain technical assumptions, often overlooked[18], that are essential for the theorem to hold. Without these assumptions, the theorem fails, giving our counterexamples. These technical assumptions, by their nature, involve facets of the model (not the underlying reality) that are difficult to verify, hence making it challenging to argue for their plausibility. This raises the question of why we should adopt these hypotheses, rather than others.

Remark 3.1. The distinction between empirical and technical assumptions might seem somewhat arbitrary, but it nonetheless serves a useful purpose in our analysis. For instance, if we apply the theorem to a jury in a criminal trial, how do we model clones of the jurors and how do we select them from (infinite) groups? Similarly, as we will see, the values of V (apart from $V(x^*) = 1$, the right option) are important for the theorem to hold; if certain conditions are not met, then the thesis of the theorem fails. However, how could one verify that the hypotheses on V hold when V is not empirically observable? What is the value or how to determine it for the $x =$ being charged with different combinations of crimes while being innocent, $x^* =$ innocent? Recall also Assumption 1, V_ϕ could be more easily observed, but we need the "social" function V. $\diamond$

3.1. V **is an injection, Assumption 3.** This was pointed out by Thompson and we reproduce it here with minor modifications. This assumption was not originally in [HP98; HP04], making the theorem false.

Let $X = \{a, b, c, d\}$. Define $V(x)$ and three agents ϕ_1, ϕ_2 and ϕ_3 according to the table below:

x	$V(x)$	$\phi_1(x)$	$\phi_2(x)$	$\phi_3(x)$
a	$1/3$	d	c	b
b	$2/3$	b	c	b
c	$2/3$	c	c	b
d	1	d	d	d

The set of agents $\Phi = \{\phi_1, \phi_2, \phi_3\}$ satisfies all the hypotheses of the theorem. The agents ϕ_1, ϕ_2, ϕ_3 have average values $5/6, 9/12, 9/12$ respectively, so ϕ_1 is the "best" agent. Notice that all three agents acting together do not always return the point d, where the maximum of V occurs. Indeed all three agents acting together work only as well as ϕ_1 acting alone. Hence in this case, no group of agents can outperform ϕ_1, or, equivalently, multiple copies of ϕ_1, hence no N and N_1 exist which satisfy the theorem. See Remark 3.2 on why some replies given to Thompson on this regard are not satisfactory.

Remark 3.2. In real-life applications, the value of V can be highly uncertain. Therefore, it is sensible to assume that, in the case of two states, $x, x' \in X$, where it is estimated that $V(x) \approx V(x')$, we set

[18] For instance, see [Rom23; Rom25] and references therein for examples.

$V(x) = V(x')$ for practical purposes. This situation should not be disregarded as uncommon. Nevertheless, as argued in [Kue17], cited by Landemore:

> You don't fail to make it to the cashier in a grocery store when you are completely indifferent between buying one more apple or one more orange, nor do deliberators in a meeting fail to decide on some course of action if two options have precisely equivalent value. Adding a simple tie-breaking rule to the theorem is entirely sufficient to deal with the mathematical hiccup and move forward with the fundamental scientific question at hand.

This argument does not address the technical point raised by Thompson. The problem is not that we are indifferent between the solutions b or c, but rather that no one knows the solution if we do start at b or c (no one moves from these states to d). A tie-breaking rule only selects either b or c, but never d, representing a failure of the theorem. The fact that the value function is "indifferent" implies that the hypotheses (in particular, the "diversity" assumption) are not sufficient to guarantee that d is reached.

The thesis of the theorem still holds if we replace Assumption 7 with $\forall x \in X \backslash \{x^*\}\ \exists\, \phi \in \Phi\ /\ V(\phi(x)) > V(x)$. However, this adjustment only serves to render the theorem more trivial and misapplies the term diversity. This condition simply implies that for every state, there exists an agent that can strictly improve that state. It is unsurprising that in a finite number of steps, these agents reach the maximum, which, by hypothesis, the best problem solver cannot always attain. Consequently, this adjustment does fix the theorem, but at the cost of making it more trivial and highlighting that what the theorem requires is not "diversity", but the existence of a more "able" problem solver who can improve upon areas where others fall short. $\diamond$

3.2. **Unique best agent, Assumption 8.** To justify this assumption, Hong and Page write:

> Let ν be the uniform distribution. If the value function V is one to one, then the uniqueness assumption is satisfied.

This mathematical statement is false. Indeed, let us consider $X = \{a, b, c, d\}$. Define $V(x)$ such that $0 < V(a) < V(b) < V(c) < 1$, $V(b) < \frac{1}{2}(V(a) + 1)$ and n agents ϕ_1, ϕ_2 and ϕ_i according to the table below:

x	$V(x)$	$\phi_1(x)$	$\phi_2(x)$	$\phi_i(x)$
a	$V(a)$	a	c	$\phi_i(a)$
b	$V(b)$	b	c	$\phi_i(b)$
c	$V(c)$	d	c	$\phi_i(c)$
d	1	d	d	d

The set of agents $\Phi = \{\phi_1, \phi_2, \phi_i\}_{i=3}^{N}$ satisfies all the hypotheses of the theorem and are ordered according to its expected value. If we set

$$V(c) := \frac{1}{3}\left(V(a) + V(b) + 1\right),$$

then ϕ_1, ϕ_2 have the same "expected ability" under the uniform measure. Furthermore, now the theorem is false. Indeed,

$$\phi_1 \circ \phi_2(a) = d,\ \phi_1 \circ \phi_2(b) = d,\ \phi_1(c) = d,\ \phi_1(d) = \phi_2(d) = d\,.$$

In this case, no group of agents can outperform $\{\phi_1, \phi_2\}$, no N and N_1 exist which satisfy the theorem.

Remark 3.3. Here, we have demonstrated an example involving two agents possessing identical "expected abilities". Of course, in real-world applications, there would likely be uncertainty or variability in the value of $\mathbb{E}_\nu (V \circ \phi)$; thus, it would be prudent to consider an interval rather than a single point. In such circumstances, the top-performing agents might comprise multiple individuals with high probability. However, as demonstrated, the theorem may not necessarily hold in these scenarios. $\diamond$

3.3. **Clones performance.** As we saw, simply by the assumptions, one million Einsteins, Gausses or von Neumanns are the same as just one of them. Indeed, mathematically, by Assumption 10, $\{\phi, \ldots, \phi\}$ is a well-defined set of problems solvers such that

$$\phi^{\{\phi,\ldots,\phi\}} \overset{\text{Assumption 9}}{=} \phi \circ \ldots \circ \phi \overset{\text{Assumption 5}}{=} \phi.$$

In other words, by the way the deliberation is structured and the idempotency assumption, a million clones of the same agent are equivalent to one. Again, those are just the assumptions they arbitrarily made. But this may not make much sense if we want to apply it to real-life scenarios. More realistic versions could be:

- *Improvement*: $V \circ \phi \circ \phi \geq V \circ \phi$ (strict inequality for some points). In other words, if an agent (competent) produced a solution after a certain amount of time, say one hour, it would provide a better answer if it had one-million hours, or if a "clone" could pick up where he left off.
- *Work in parallel*[19]: $V \circ \phi^{\{\phi,\phi\}} \geq V \circ \phi$ (strict inequality for some points). In other words, one can imagine that a group of Einsteins would not work sequentially, always producing the same result, but would divide the work, resources, focus, etc. to produce a better answer once they have put all of their findings together.

It is not clear which is the best way to treat formally the performance of clones in their theorem. But as the previous arguments show, even in conditions of ideal theorizing, there are more reasonable assumptions that could have been chosen. However, these alternative assumptions do not guarantee the conditions required for the theorem to hold. Otherwise, as $N_1 \to \infty$, no group of agents could generally outperform $\phi, \overset{N_1}{\ldots}, \phi$; we cannot guarantee the existence of an N that would satisfy the theorem.

Remark 3.4. Following Jason Brennan's recurrent "magic wand" thought experiment, let's imagine we are confronted with an exceedingly difficult problem to solve, for instance, the Navier-Stokes Millennium Problem. Suppose we have a magic wand at our disposal that can create agents to solve the problem for us. Should we choose Terence Tao, or should we use the magic wand to create 100 Terence Taos working together to solve our problem? According to the assumption of the Hong-Page Theorem, this magic wand would be useless. $\diamond$

Remark 3.5. The critique still holds in the case of infinite X. Regarding the issue of clones, the following is stated in [HP04]:

> [...] we present a simpler version of our result where X is assumed to be finite. This finite version makes the insight more straightforward, although it comes at the cost of trivializing some intricate assumptions and arguments. For example, the group of the best-performing agents is proven below to be comprised of identical agents. This is an artifact of the finite version. In the general version under reasonable conditions, the group of the best-performing agents can be shown to be similar, not necessarily the same.

However, this explanation is far from accurate. Clones also appear in the less realistic case where X is not finite. This occurs because we have to take copies from Φ and, if ϕ has already appeared,

[19] As a technical note, now $\{\phi, \phi\}$ should be considered a multiset (the multiplicity distinguishes multisets).

it can appear again[20]. Moreover, the finite version of the model is neither sufficient nor necessary for proving that the group of the best-performing agents is comprised of identical agents.

In a scenario where X is finite, the best agents could be several different ones once we consider only one copy of each agent. This could be easily demonstrated by following our previous example from Section 3.2 or [HP98, Assumption 5']. In the version where X is not finite, according to Assumption 5 of their appendix, $B(\phi^*, \delta) \cap \Phi = \{\phi \in \Phi \mid d(\phi, \phi^*) < \delta\}$ could contain only one agent, namely ϕ^*. It should also be noted that a finite X represents a more realistic setup. Typically, rendering things continuous simplifies the analysis, as it allows us to use standard calculus, for example, but this is not the case here. It is less realistic to assume that agents have answers to an infinite set of elements than to a finite set.

Furthermore, to reach the conclusion in this case, more predetermined hypotheses are needed, which cannot be justified in terms of being reasonable or intuitive. Basically, Assumption 5 in [HP98, Appendix] imposes that the best agents are clustered near the best-performing agent. They are clustered, in the sense of the distance metric defined above, in a way that guarantees that if they are close, the expected performance will be close too, [HP98, Lemma 2]. Then, in the proof, they choose the best agents as close to the best-performing agent as needed so that they do not always reach x^*, [HP98, Lemma 4]. ◇

3.4. **Selection of clones.** Similarly, the assumptions for clone selection appear to be somehow arbitrary, and it is not justified why these assumptions, that facilitate reaching the conclusion that "Diversity Trumps Ability" should be chosen over other perfectly reasonable and simpler ones:

- The choice of two independent groups seems arbitrary. Why not fix N and, from the same group, select a random subgroup of size N_1, as well as the best N_1 problem solvers, and then compare? In such a scenario, the theorem might not hold. Indeed, we need $N \gg N_1$ such that the Strong Law of Large Numbers (SLLN) applies, $\mu(\{\phi^*\})$ can be very small. However, a random group of N_1, Φ_{N_1}, agents might not include all the needed problem solvers of Φ, thus we cannot guarantee a probability of one, as the theorem does. That is, for $N > N_1$, it could be the case that, see Remark 2.9,
$$\mathbb{P}\left(\tilde{\Phi}_0 \subset \Phi_{N_1}\right) < 1\,.$$
- Permitting repetition is also arbitrary. We could, for instance, select the best problem solvers without allowing repetitions, as is standard in probability problems. Recall from Section 3.3 that adding a repeated clone is equivalent to adding nothing. This could prevent the paradoxical result that, by mere hypotheses, when choosing the best problem solvers from a group of size N is more beneficial when the group size is relatively small, i.e., for choosing the best it is preferable to have less options available. However, if we prohibit repetitions, then the theorem does not necessarily hold as the best problem solvers can include the ones (without taking repeated clones into account) of the "random" group, so no N and N_1 exist which satisfy the theorem.

We should note the general approach adopted by Hong and Page:

(1) They introduce randomness into their model by assuming an infinite number of clones for each agent.
(2) They invoke the SLLN (or similar results) to ensure that the frequency of appearance converges to the original probability, given by μ.
(3) As the group sizes involved in the SLLN could be large, N_1, N are not fixed, but also random. For instance, $N_1 = N_1(\omega)$ is chosen such that the "random" group contains enough agents to outperform ϕ^*. This cancels out the randomness previously introduced.

[20] With positive probability in the reasonable case of a finite Φ.

Note that the infinite pool of clones is not a proper pool (with a finite size,as in real life examples or the simulations) because we can take an infinite number of elements from it. Let us call it the infinite pool of clones. In other words, the actual process is the following. First, we form the "random" group taking elements from this infinite pool of clones until we get the desired elements (enough to outperform the best agent). Note that the number of elements of this group, N_1, is not fixed; it is not always 10 or 15, but it can be as large as needed depending on the situation. That is, this number is random and defined in such a way that cancels the randomness previously introduced: it is defined so that we always have a "random group" containing the elements we want. Once this group is formed, then, from an independent event, we make a large set of clones, of size N such that there are N_1 clones of the best agent. Again, N is not a fixed number, but as large as needed until the desired conclusion is reached. This has nothing to do with choosing from a given pool with finite elements, as often suggested. What is more, it is not just that the description is not accurate, but as we have seen, doing something similar to what they seem to suggest would make the theorem false. We cannot select the agents in a normal way, but we need an infinite amount of clones of each agent to be able to stop whenever we have reached the desired elements. That is the reason why we argued that they introduced the probability apparatus unnecessarily, which can be simplified by removing the probability part. The simplification reveals that the theorem is merely a restatement of the hypotheses.

All in all, the steps detailed above effectively eliminate the randomness that was introduced, transforming the framework into a *de facto* deterministic model. That is, after the "complex" probabilistic steps, with probability one, the groups being compared are, essentially[21], the ones in Theorem 2.7. But, as we saw, if the results were presented this way, the triviality would be manifest. In the next section, we will remove clones and explore the consequences of a fair comparison of ability and diversity in the Hong-Page setting.

[21] See Appendix A for more details.

4. New Hong-Page style theorem: Ability trumps diversity

The objective of this section is to make a fairer comparison of diversity and ability while maintaining the Hong-Page framework. We are going to state and prove a new version of the Hong-Page theorem such that the hypotheses are going to be of the same kind and as plausible (or even more as we will see, for instance, no need of clones and disagreement is possible) as the ones in the Hong-Page theorem. Nevertheless, we will reach the opposite conclusion, "ability trumps diversity". We are not claiming that this theorem has any social content; it simply shows how the justification of the theorem is driven by the selection of assumptions that might be overly strict or arbitrary.

The moral would be if we create two groups from the group in the original theorem – one in which we make the minimal reduction in ability while ensuring full diversity, and another in which we considerably reduce diversity while ensuring ability – the less diverse group would systematically outperform the fully diverse group. In other words, ability trumps diversity. First, we present a non-technical exposition, see Section 4.3 for all the mathematical details.

4.1. **The new assumptions.** We start with Hong-Page's framework, but redefining some of their assumptions to incorporate the potential for disagreement within a group of agents, introducing a realistic element into our model. We describe a situation where one agent's suggestion can be directly countered by another, leading to a deadlock, symbolizing a disagreement. Unlike the original Hong-Page Theorem's assumptions, where agreement is always achieved, our model accepts the possibility of disagreement, reflecting a more lifelike scenario where unanimous decisions might not always be possible. Clones are not needed now.

Also, we then introduce a probability measure to account for the likelihood of each agent influencing the decision process. This is done ensuring that every agent has a chance to contribute, emphasizing inclusivity in the decision-making process.

4.1.1. *The ability group.* We define an ability-focused group, Φ^{A}, with the following characteristics:

- **Ability:** Every agent in this group is capable of maintaining or improving the current state, ensuring that their contributions are always constructive.
- **Common knowledge:** There exists a subset of states, X_{CK}, where all agents in this group agree on the solution, showing non-diversity and competence.
- **Limited diversity:** While this group maintains some level of diversity, it is notably reduced outside of X_{CK}, with a unique agent providing distinct solutions in these scenarios.

4.1.2. *The diversity group.* Conversely, the diversity group, Φ^{D}, is characterized by:

- **Full diversity:** its wide range of perspectives, ensuring that for almost every problem state, there's an agent capable of offering a different perspective.
- **Minimal ability loss:** this group includes the minimal loss of ability possible. However, this will serve to demonstrate that this is enough to show that diversity does not trump ability.

4.2. **The theorem.** Theorem 4.3 shows that given the defined groups Φ^{A} and Φ^{D} with their respective properties, the group selected for its ability rather than diversity performs better in reaching the optimal solution.

This theorem underscores the importance of selecting problem solvers not just for their diverse perspectives but for their ability to constructively contribute to solving complex issues, challenging the notion that diversity trumps ability. See next section, Section 4.3 for a fully detailed exposition.

4.3. **Ability Trumps Diversity Theorem: mathematical analysis.** We give the precise statements and proof of the theorem mentioned in the previous section.

4.3.1. *The new assumptions.* Among a set of agents Φ, we select two finite groups with different properties. We are going to modify some assumptions, but the other remain the same. First, let us introduce the possibility of disagreement following Assumption 9 as:

$$\phi_{i^x_{j+1}}\left(\phi_{i^x_j}(x_{j-1})\right) = x_{j-1}\,, \text{ with } \phi_{i^x_j}(x_{j-1}) \neq x_{j-1} \text{ and } i^x_j \neq i^x_{j+1}\,.$$

A disagreement is a stopping point. In other words, if there is a cycle such that one agent proposes a new solution and other reverse back that solution there is a disagreement and that initial solution is given as the group solution. This is a simple model where disagreement is possible.

Remark 4.1. Note that in the original formulation of the Hong-Page Theorem, for any group of any size, even if they might not be able to reach the correct solution, Remark 2.9, there will be no disagreement in any case, as per Assumption 9. This always leads to unanimity, which is highly unrealistic. ◇

Let also μ_x be a probability measure such that if x is the previous solution, $\mu_x(\{i\})$ represents the probability that $\phi_i(x)$ is the next solution in the deliberation chain[22], see Assumption 9. This measure has full support; no one is silenced. The indices are chosen independently. Once x_0 is fixed, this defines a probability measure $\mathbb{P}$ on the possible paths. That is,

$$\mathbb{P}\left(x_k = x' \mid x_{k-1} = x\right) = \mu_x\left(\{i \mid \phi_i(x) = x'\}\right) > 0\,.$$

4.3.2. *The ability group.* Let us denote a group by $\Phi^{\mathrm{A}} = \{\phi_\alpha\}_{\alpha\in A}$ which is selected such that:

- *Ability*: $V \circ \phi_\alpha \geq V$. In other words, this group is chosen to ensure ability in the sense that each agent does not decrease the value of the initial state given. This is Assumption 4, but now is imposed by the selection of the group.
- *Common knowledge*: $\exists X_{\mathrm{CK}} \subset X$ such that:
  Non-diversity set: $\forall \alpha, \alpha' \in A$ we have $\phi_\alpha|_{X_{\mathrm{CK}}} \equiv \phi_{\alpha'}|_{X_{\mathrm{CK}}}$.
  Knowledge: $\phi_\alpha(x_c) = x^* \;\forall\, x_c \in X_{\mathrm{CK}} \;\forall\, \alpha \in A$.
  In other words, this group selection comes with a selection bias. The agents have a common knowledge that makes them similar; for the set X_{CK}, they all give the same solution. This is an extension of the second part of Assumption 4; agents are not only able to recognize that x^* is the solution, but they can do the same for other states $x \in X_{\mathrm{CK}}$. Note that $x^* \in X_{\mathrm{CK}}$.
- *Smaller diversity set*: $\forall\, x \in X \backslash X_{\mathrm{CK}}$, Assumption 7 holds. In other words, for this group, the original assumption of Hong and Page only holds outside the "larger" set X_{CK}, instead of x^*.
- More importantly, we diminish diversity in a second way. For all x in $X \backslash X_{\mathrm{CK}}$, there exists exactly one agent, ϕ^x, who provides a distinct answer, and the set of unique answers could be equidistributed, meaning it's not just one agent always giving the different answer. Formally, $\{x \mid \phi^x = \phi\} \leq \frac{|X|}{|\Phi^{\mathrm{A}}|} + 1$ for all $\phi \in \Phi^{\mathrm{A}}$. Therefore, if the ratio is small enough, two agents are quite similar, signifying a lack of diversity, i.e., following [HP98, Appendix], their distance is relatively small.

Remark 4.2. For the theorem to function, we don't require a large X_{CK}, but having it large makes the agents less diverse. It could be just x^* as in the original theorem. ◇

[22] From here it is evident that we are not assuming "relay dynamics" as in the original Hong-Page framework, see Section 6. Nevertheless, [LS24, Table 3] argue that we are, i.e., that relay dynamics was used in the previous paper where this result was first shown, [Rom23]. The authors also say that according to Thompson, see [Tho14], "Random>Diversity>Ability", while we cannot find evidence of the last inequality in [Tho14].

4.3.3. *The diversity group.* Let us denote a group by $\Phi^{\mathrm{D}} = \{\phi_j\}_{j\in\mathcal{J}}$ which is selected such that the maximum diversity is guaranteed. More precisely, there is a unique $x^0 \in X$ such that:

- *Full-diversity with ability*: $\forall\, x \in X\backslash\{x^0, x^*\}$ there is a set of agents $\{\phi_{j_k^x}\}_{k=1}^{n_x} \subset \Phi^{\mathrm{D}}$ such that $\phi_{j_k^x}(x) \neq x$ and such that all the states that improve the state are the local optimum for some agent.
- *Minimal ability loss*: there is only one agent $\phi_{j_0} \in \Phi^{\mathrm{D}}$ and only one state x^0 such that $V\left(\phi_{j_0}(x^0)\right) < V\left(x^0\right)$. Note that this is the *minimal ability that can be lost.*

4.3.4. *The theorem.*

Theorem 4.3 (Ability Trumps Diversity)**.** *Let Φ^{A}, Φ^{D} as above with the given assumptions. Then, the ability group outperforms the diversity group almost surely*[23].

Proof. To prove this theorem, we need to compare the performances of the two groups. First, we consider the ability group Φ^{A}. Any agent from Φ^{A} does not decrease the value of the given state. Moreover, for any state in the non-diversity set X_{CK}, all agents in Φ^{A} will return the optimal solution x^*. Thus, following the measure $\mu_{x'}$, for any $x \in X$ there is a sequence of $n = n_x$ agents such that

$$V(x) \leq V\left(\phi_{i_1^x}(x)\right) \leq \ldots \leq V\left(\phi_{i_n^x}(x_{n-1})\right) = 1\,.$$

This is because, for any $x' \neq x^*$, $p_{x'} := \mu^A_{x'}\left(\alpha \in A \mid \phi_\alpha(x') \neq x'\right) > 0$. This holds true even in the worst-case scenario where all agents, except one, are stuck at that point. Thus being stuck has probability, by the subadditive property,

$$\mathbb{P}\left(\exists\, n_0, x' \neq x^* \;:\; \phi_{i_n^x}(x') = x' \;\forall\, n \geq n_0\right) \leq \sum_{n_0} \sum_{x' \neq x^*} \prod_{i=n_0}^{\infty} (1 - p_{x'}) = 0\,.$$

where the sum in x' is finite. Thus, with probability one, in a finite number of steps we have strict inequalities reaching x^*, returning this as the solution. Thus, for all $x \in X$, every path starting at x leads to x^*. Thus,

$$\mathbb{E}_{\mu^{\mathrm{A}},\nu}(V \circ \Phi^{\mathrm{A}}) := \sum_{x\in X} \nu(x)\mathbb{E}_{\mu^{\mathrm{A}}}\left(V \circ \Phi^{\mathrm{A}}(x))\right) = 1\,.$$

where $\Phi^{\mathrm{A}}(x) := \phi^{\Phi^{\mathrm{A}}}$.

Now, consider the diversity group Φ^{D}. This group is selected to maximize diversity and only allows minimal ability loss. However, there exists exactly one agent ϕ_{j_0} and one state x^0 such that $V\left(\phi_{j_0}(x^0)\right) < V\left(x^0\right)$. Let $x^{(-1)} := \phi_{j_0}\left(x^0\right)$ and let x such that $V(x) \leq V(x^{(-1)})$, being x the start. Similar as above, by finiteness, there exists some k such that the probability satisfies

$$\mathbb{P}\left(\exists\, \{i_j^x\}_j \mid \phi_{i_k^x}(x_{k-1}) = x'\right) > 0\,.$$

We have two possibilities:

[23]Note that we are not assuming any particular form of the function V; it works for all injective V. As we will see in Section 6, this will be important for the simulations as they depend on the "shape of the landscape." In particular, the shape will be related to the "predictability" of the problem, although see Section 7.4. Nevertheless, [LS24, Table 3] states that this theorem only applies for low-predictability problems, which is not true. We are also introducing disagreement, which is not present in the previous studies and is an important aspect of realism (unanimous agreement is not realistic), but the authors do not mention it in the column "Other features" of the table.

- If $x_{k-1} = x^{(-1)}$ and $V(x) \leq V(x^{(-1)})$, then, a "disagreement cycle" can be completed, $x_{k-1} = x^{(-1)} \to x^0 \to x^{(-1)}$, returning $x^{(-1)}$. This happens with probability

$$\sum_{k=1}^{\infty} \mathbb{P}\left(x_{k-1} = x^{(-1)}\right) \mathbb{P}\left(x_k = x^0 \mid x_{k-1} = x^{(-1)}\right) \mu_{x^0}^{\mathrm{D}}(j_0) > 0,$$

where $\mathbb{P}\left(x_k = x^0 \mid x_{k-1} = x^{(-1)}\right) = \mu_{x^{(-1)}}^{\mathrm{D}}\left(\{j \in J \mid \phi_j\left(x^{(-1)}\right) = x^0\}\right) > 0$, where we have used the full-diversity assumption.
- Also, if $x_{k-1} = x^0$ again, completes a disagreement cycle, $x^0 \to x^{(-1)} \to x^0$, returning x_0. Similarly, this happens with probability

$$\sum_{k=1}^{\infty} \mathbb{P}\left(x_{k-1} = x^0\right) \mu_{x^0}^{\mathrm{D}}(j_0) \, \mathbb{P}\left(x_{k+1} = x^{(-1)} \mid x_k = x^{(-1)}\right) > 0 \,.$$

As $V(x^{(-1)}) < V(x^0) < 1$, thus,

$$\mathbb{E}_{\mu^{\mathrm{D}},\nu}(V \circ \Phi^{\mathrm{D}}) := \sum_{x \in X} \nu(x) \mathbb{E}_{\mu^{\mathrm{D}}}\left(V \circ \Phi^{\mathrm{D}}(x)\right) < 1 \,.$$

□

5. The Diversity Prediction Theorem and the Crowds Beat Averages Law

5.1. **The results.** Hong and Page also present another theorem that would be useful to analyze. First, some definitions. Given a set of individuals labeled as $i = 1, \ldots, N$, we associate to each of them a signal or prediction of some magnitude, which has θ as true value. The squared error of an individual's signal equals the square of the difference between the signal and the true outcome:

$$\mathrm{SE}(s_i) = (s_i - \theta)^2 .$$

The average squared error is given by

$$\mathrm{MSE}(\underline{s}) = \frac{1}{n}\sum_{i=1}^{n}(s_i - \theta)^2 ,$$

with $\underline{s} := (s_1, s_2, \ldots, s_n)$. The collective prediction is

$$c = c(\underline{s}) = \frac{1}{n}\sum_{i=1}^{n} s_i .$$

Predictive diversity of the collective is defined as:

$$\hat{\sigma}(\underline{s}) = \frac{1}{n}\sum_{i=1}^{n}(s_i - c)^2 . \tag{5.1}$$

This is simply a (biased) estimation of the variance. Two trivial theorems can be deduced. The first, a particular version of the Pythagoras Theorem:

Theorem 5.1 (Diversity Prediction Theorem)**.** *The squared error of the collective prediction equals the average squared error minus the predictive diversity:*

$$SE\left(c\left(\underline{s}\right)\right) = MSE(\underline{s}) - \hat{\sigma}(\underline{s}) .$$

Proof. This is quite standard, but let us give a proof using the (generalized) Pythagoras Theorem. In $\mathbb{R}^n$ we can define the standard Euclidean or l^2-norm. If $\underline{c} = (c, \ldots, c)$ and analogously for $\underline{\theta}$, then $\langle \underline{s} - \underline{c}\, , \underline{\theta} - \underline{c} \rangle_{l^2} = 0$ so the Pythagoras Theorem gives

$$\|\underline{s} - \underline{\theta}\|_{l^2}^2 = \|\underline{\theta} - \underline{c}\|_{l^2}^2 + \|\underline{s} - \underline{c}\|_{l^2}^2 . \tag{5.2}$$

□

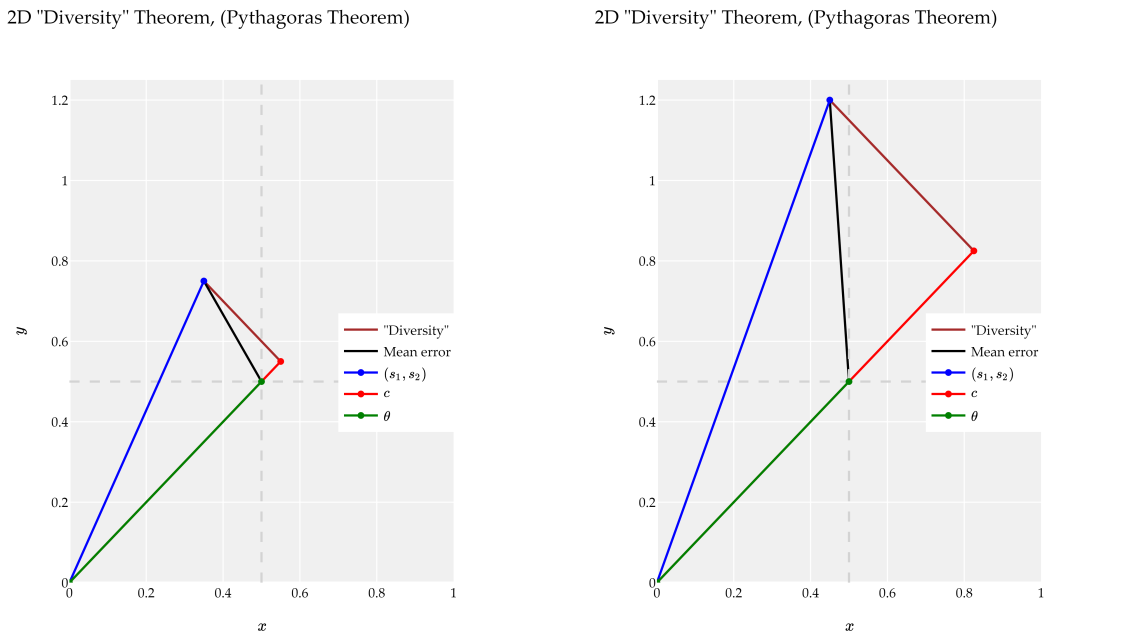
Corollary 5.2 (Crowd Beats Averages Law)**.** *The squared error of the collective's prediction is less than or equal to the averaged squared error of the individuals that make up the crowd.*

$$SE\left(c\left(\underline{s}\right)\right) \leq MSE(\underline{s}) .$$

5.2. **The asymmetric role of "ability" and "diversity".** Before we proceed, let's note two simple mathematical observations:

Error 1. MSE **and** $\hat{\sigma}$ **cannot be treated as independent** as both depend on $\underline{s}$. That is, altering one will generally change the other (it is not fixed), with the effect on the prediction error being, in principle, undetermined.

Error 2. Therefore, it would be a significant mathematical error to consider that for the prediction error, SE, to be small is enough to make "diversity", $\hat{\sigma}$ large.



(A) Initial result, prediction error is relatively "small": $(c-\theta)^2 \approx 0.0025$.

(B) Result after increasing diversity, prediction error is relatively "large": $(c-\theta)^2 \approx 0.1056$.

FIGURE 1. Increasing diversity does not always improve predictions and can sometimes significantly worsen them. The prediction error is represented by the red line, with the red dot indicating the prediction. The black line corresponds to MSE, and the brown line to $\hat{\sigma}$. The true value, θ, is $\frac{1}{2}$ (represented by the green dot). As diversity increases more than threefold (from 0.04 to 0.14), the squared error becomes more than forty times larger.

These observations are mathematically trivial. Also, they can be graphically demonstrated when we consider the case of $n = 2$, which brings us back to the standard Pythagoras theorem, see Figure 1. Knowing either $\mathrm{MSE}(\underline{s})$ or $\hat{\sigma}$ alone is not sufficient to determine the value of the prediction error. In fact, according to the Crowd Beats Averages Law, we can see:

$$\mathrm{SE}\left(\mathrm{MSE}, \hat{\sigma}\right) \in \left[0, \mathrm{SE}^{\max}(\underline{s})\right]. \tag{5.3}$$

This bound is sharp, with $\mathrm{SE}^{\max} := \mathrm{MSE}$. Since SE is not solely determined by either "ability" or "diversity", these variables can be observed in the context of the maximum prediction error, i.e., $\mathrm{SE}^{\max}$. More precisely:

Proposition 5.3. *Let $SE^{\max}$ represent the maximum prediction error. Then,*

- *If $\Delta MSE < 0$, then $\Delta SE^{\max} < 0$. In other words, if "ability" increases, the maximum prediction error decreases. Particularly, if the increase in ability is large enough, the prediction error will decrease.*

- *If $\Delta\hat{\sigma} > 0$, then $\Delta SE^{\max} \geq 0$. This implies that if "diversity" increases, the maximum prediction error also increases. In particular, an increase in diversity alone does not guarantee a reduction in the prediction error. Furthermore, if the increase in diversity is substantial enough, the maximum prediction error will also increase.*

Proof. This is a trivial consequence of $\mathrm{MSE} = \mathrm{SE}^{\max}$ and the twin inequality of the Crowd Beats Averages Law: $\hat{\sigma} \leq \mathrm{SE}^{\max}$. □

Using the Crowd Beats Averages Law (and other trivial results), we arrive at a seemingly contradictory result: increasing "ability" eventually reduces the prediction error, but increasing diversity ultimately increases the maximum prediction error. Consequently, the Diversity Prediction Theorem and the Crowd Beats Averages Law provide limited insight into how diversity impacts the prediction error in a general setting without controlling for ability.

5.3. **A basic mathematical error in advocating for diversity.** Scott Page has argued that large diversity implies small prediction error. However, this conclusion, while favorable to the hypothesis that diversity reduces prediction error, constitutes a significant mathematical mistake. Indeed[24], in a lecture (University of Michigan), Page states:

> And you might also ask, where does the madness of crowds originate? How could it be that a crowd could get something completely wrong? Well, that's not difficult to understand either, because crowd error equals average error minus diversity. If I want this to be large, if I want large collective error, then I need large average error, meaning that I need people to be getting things wrong, on average. **Additionally, I need diversity to be small**. So, **the madness of crowds comes from like-minded individuals** who are all incorrect, and once again, the equation provides us with this result. *(emphasis added)*

This mathematical misunderstanding involves a basic arithmetic error that we mentioned in Error 2. From the "Diversity Prediction Theorem" (with $\underline{s}$ term omitted for simplicity),

$$\mathrm{SE} = \mathrm{MSE} - \hat{\sigma}\,,$$

we **cannot** deduce that a large SE implies a small $\hat{\sigma}$. Rather, it implies that MSE must be much larger than $\hat{\sigma}$, where $\hat{\sigma}$ could be as large as desired. See, for instance, Figure 1b for an illustration, where the prediction error is large, but diversity is larger (so it cannot be "small").

Remark 5.4. It is worth mentioning that Scott Page has made this point clear elsewhere. For instance, in his book "The Diversity Bonus", page 75, he says:

> Our first intuition might be to select the three most accurate students. That group would have the lowest average error. Recall though that the group's error depends in equal measure on their diversity. By choosing the group with the smallest average error, we ignore diver- sity. We ignore half of the equation. If the three most accurate people all think the same way, then we have no diversity. **Selecting the group of three people with the largest diversity makes even less sense**. To have high diversity, the group must have an even higher average error. The best approach is to search among all 570 possible groups for the one whose diversity is closest to its accuracy.

◇

[24] Note that if one states that a large prediction error implies small diversity, by logical necessity, it is also stating that large diversity implies small prediction error (as a large prediction error would imply large diversity, a contradiction).

6. Simulations in the Hong-Page framework

In the original Hong and Page paper, they present a series of simulations to both illustrate and motivate their main result. Other authors have also used the simulation part to develop the idea of the original authors, e.g., [Gri+19] (which is a worthwhile read as it has a far deeper analysis than the original paper. Additionally, the authors already warned about the misapplications and limitations of the theorem[25]) and, for instance, [Hol+18; HMS23; Gri+24]. In this section, partially based on the results of the previous sections, we analyze the simulations presented by the authors. First, we will explain the framework and show the lack of realism of the simulations. Secondly, we will show that the theorem cannot be applied to the simulations. Thirdly, we will introduce a mathematical framework to analyze the simulations. Fourthly, through the main part of the section, we will see that the general structure of the simulations is the following:

Pseudo-Theorem 6.1. *Design*: Restrict the simulation framework and parameters such that, for the vast majority of cases, the following holds:

- the random group is composed of able agents (not far from experts),
- the best-performing agents are similar to each other,
- the function V is such that it favors "diverse" heuristics, i.e., the random group,
- choose the group size such that the random group has enough elements to avoid getting stuck in local optima, while being small enough so that the best group has agents which are similar to each other.

Result: Consequently, this "random" group will outperform the best group.

The result is quite simple, the best group will not perform much better than the single best agent by construction while the random group will perform better as we are adding enough agents, which, by construction, never degrade the solution, to avoid getting stuck in local optima. This will be the problem for the best-performing agents in the kind of V used by Hong and Page. This is not a surprising result, but it is all that the simulations show. It is a similar idea as in the theorem, see Section 2.6. Thus, we will see that by changing some (arbitrary) parameters in the original simulations, the "Diversity Trumps Ability" result *no longer holds*. These are trivial changes with respect to the original paper (that could have been included), but it shows that the results are not robust to the parameters. Additionally, we will modify the framework of the simulations to include a more realistic setting to properly compare ability and diversity, which will show, again, the lack of robustness of the results; the "Diversity Trumps Ability" result does not hold. The changes will be motivated by our previous analysis of, especially, Section 4.

[25] We will explore more limitations than those mentioned in the paper, but the conclusion of some of the authors is clear, [GS20]:

> The Hong-Page result is therefore very sensitive to the "smoothness" of the epistemic landscape modeled. As hinted in section 3.2.2, this is an indication from within the modeling tradition itself of the danger of restricted and over-simple abstractions regarding epistemic landscapes. Moreover, the model's sensitivity is not limited to landscape smoothness: social epistemic success depends on the pool of numbers from which heuristics are drawn as well, with "diversity" showing strength on smoother landscapes if the pool of heuristics is expanded. Results also depend on whether social interaction is modeled using of Hong-Page's "relay" or an alternative dynamics in which individuals collectively (rather than sequentially) announce their results, with all moving to the highest point announced by any. Different landscape smoothnesses, different heuristic pool sizes, and different interactive dynamics will favor the epistemic advantages of different compositions of groups, with different proportions of random and best-performing individuals (Grim et al. 2019).

Thus, we will conclude that the simulations support the Hong-Page result only if we tune the simulation parameters in a certain way, or better said, if we start from the original setup and change the parameters, the "Diversity Trumps Ability" result no longer holds.

6.1. **Simulation framework: a far from realistic setting.** Even if all the previous assumptions of the theorem were true, this is not enough for the simulation part. The simulation framework is even more limited than the theorem, which we criticize in Sections 2, 3, or 4 for its lack of realism as it includes, among other things:

- Unrealistic individual performance, e.g., Assumption 4.
- Unrealistic group and individual dynamics, Section 3.3.
- Always end up in unanimous agreement, and with enough agents, it will always find the solution, Section 2.6.
- It does not allow disagreement in values or erroneous values while searching for the maximum, Section 4.
- It does not allow for knowledge of the search space, Section 4.

The theorem, while restrictive, at least allows for a general class of functions V and heuristics ϕ. In contrast, the simulation framework imposes a very specific structure:

- The search space is artificially constrained to a circular array of integers or similar structures [HP98; HP01].
- The heuristics are limited to fixed-length tuples of numbers used as step sizes with a "strange" algorithm to find the maximum. The connection with a real-world setting is dubious.
- The possible kind of V is limited to a random function, or, at best, a linear interpolation of random points [Gri+19].

Therefore, even if the simulation results supported the "Diversity Trumps Ability" claim (which, as we will see, they do not), generalizing these findings to actual organizational decision-making would be highly questionable.

In the following sections, we will demonstrate that even within this highly constrained simulation framework (or slight variations of it), the claim that "Diversity Trumps Ability" does not hold under the authors' own definitions of ability and diversity. The simulations will show that small changes to the framework can quickly break the result, suggesting it lacks robustness and requires parameter selection to produce the desired outcome. However, it's important to note that the more fundamental question—whether this artificial framework has any meaningful connection to real-world group problem-solving—remains entirely open. The gap between these simulations and actual organizational decision-making is vast and largely unaddressed. The small changes that, in my humble opinion, we introduce to make the simulations more realistic are, by no means, all that is needed to make the simulations a useful tool to understand real-world group problem-solving; see also Section 7.

6.1.1. *The theorem does not apply.* In [HP04], the authors claim that the theorem explains the simulations. This is not the case. First, the theorem assumes an infinite size of the pool of agents and the number of agents N and N_1 would vary for each simulation, but this does not happen in the actual simulations. Recall that as we said in Section 3.3, the number of agents N is not fixed but a random stopping point and the same happens for the number of agents N_1, the size of the group. In the actual simulations, these numbers are *fixed*, say $N_1 = 9, 10$, or 20 and N is the number of possible heuristics, for instance, if $l = 12$ and $k = 3$, then $N = 12 \times 11 \times 10 = 1320$. But, in the theorem, first N_1 is chosen so that the random group has enough agents to surpass the best-performing agent, and then N is chosen so that the best group has only clones of the best-performing agent. These numbers in the theorem vary for each simulation, but in the actual simulations, they are fixed. Second, the simulations do not necessarily satisfy the hypotheses. For instance, the important diversity assumption, Assumption 7, is not satisfied, see [Tho14] for more details. Third, the predictions (conclusions) of the theorem are not the same as the simulations. For instance, the theorem predicts with probability one that the random group will outperform the best group, which is not the case even in the original simulations, as we will see in Section 6.3. In particular, the form of V will be important but this is ignored by the theorem which works for all functions according to the assumptions. The problem is deeper, as we will analyze, the theorem misses essential aspects of the simulations, like the overshooting of the random group or its stickiness in local maxima, Section 6.3.4, which explain why the DTA sometimes does not hold.

In any case, the ideas behind the theorem and the simulations are similar, compare the arguments of "Pseudo-theorems" 2.1 and 6.1, but that does not mean that the theorem explains the simulations in the sense of establishing when the parameters are such that the result of the simulations holds and when it does not (as we will see, in a large number of situations, the result does not hold). Finally, the ideas behind both arguments are not surprising. The hypotheses are doing all the work. All in all, the theorem seems an artifact based on a simple insight to try to give validity to a particular set of unrealistic simulations, while in reality it fails completely in its purpose.

6.2. **General setting and original example.** As we have said, Hong and Page [HP04] present a series of simulations where, allegedly, the "Diversity Trumps Ability" result holds. The original example is explained in [HP04] in natural language and in [HP98] more formally, although with some confusion[26]. Let us present a rigorous setting here using pseudocode that will be useful for the latter analysis. As before, let $X = \{1, 2, 3, \dots, n\}$ be the set of the first n positive integers, and let $V : X \to [0, 1]$ be a function. As above, the objective is to find the maximum value of V on X. We assume V has a unique global maximum at some point $x^* \in X$. Recall that, in general, an agent is a function $\phi : X \to X$ that transforms an initial state x into another state.

6.2.1. *Heuristics definition.* Fix two integers l and k such that $1 \leq k < l < n$. We define an *agent* α as an ordered k-tuple of distinct integers:

$$\alpha = (a_1, a_2, \dots, a_k), \quad \text{where } a_i \in \{1, 2, \dots, l\} \text{ for all } i, \text{ and } a_i \neq a_j \text{ for } i \neq j.$$

Each agent α defines a search procedure on X, ϕ_α, that can be formalized with the following pseudocode:

[26]For instance, strictly speaking, the statement $t = t + 1$ is incorrect as an equality, since no value can be equal to itself plus one (that is, $0 \neq 1$). In the context of algorithms, we want to convey an assignment rather than an equality, so we write $t \leftarrow t + 1$. The left arrow ($\leftarrow$) makes clear that t is being updated to a new value, rather than declaring the two sides to be identical.

Algorithm 6.1: Heuristics algorithm

Input: Starting point $i \in X = \{1, 2, \ldots, n\}$, function $V : X \to [0, 1]$
Heuristic list $\alpha = (a_1, a_2, \ldots, a_k)$ with $a_j \in \{1, 2, \ldots, l\}$
Output: Stopping point $\phi_\alpha(i)$

Initialize $x \leftarrow i$
while *True* **do**
  for $j \leftarrow 1$ **to** k **do**
    Compute $x_j \leftarrow (x + a_j) \mod n$
    if $V(x) < V(x_j)$ **then**
      Rearrange $\alpha \leftarrow (a_{j+1}, a_{j+2}, \ldots, a_k, a_1, a_2, \ldots, a_j)$
      Update $x \leftarrow x_j$
      break
  if no j **found such that** $V(x) < V(x_j)$ **then**
    return x *as the stopping point* $\phi_\alpha(i)$

In other words, this algorithm is a particular method for finding a peak in a set of numbers based on a given function that assigns a value to each number. Starting from an initial number, it uses a list of steps (like directions or moves) to explore nearby numbers. At each step, it checks if moving to a new number yields a better (higher) value according to the function. If it finds a better number, it moves there and adjusts its list of steps to avoid repeating the previous ones. This process repeats until it goes through all its steps without finding a better number, suggesting that it has reached a peak where the value cannot be improved by moving to adjacent numbers using its available steps, "heuristics".

Therefore, we can define the stopping point $\phi_\alpha(i)$ as:

$$\phi_\alpha(i) = x \text{ where } x \text{ satisfies: } V(x + a_j) \leq V(x) \quad \forall j \in \{1, 2, \ldots, k\}\,. \tag{6.1}$$

Some observations are in order:

(1) The stopping point $\phi_\alpha(i)$ is uniquely determined by α and the starting point i.
(2) While $V(\phi_\alpha(i))$ is a local optimum in the sense of (6.1); the algorithm usually fails to find the maximum.
(3) The efficiency of this method depends on the choice of α and the structure of V, as we will see.

6.2.2. *Group dynamics.* The group's performance, comprising the set of problem solvers $\Phi' := \{\phi_1, \ldots, \phi_{N_1}\}$, is described in Assumption 9. Here, i_k^x denotes the next agent, following the sequential order. That is, we follow[27] "relay dynamics":

$$i_1^x = 1\,,\ i_k^x = i_{k-1}^x + 1 \mod N_1\,.$$

Alternatively, we could define it as the first agent, following the sequential order, who returns a different point

$$i_k^x := \min\{i \in \{i_{k-1}^x + 1, \ldots, i_{k-1}^x + N_1\} \bmod N_1 \text{ such that } \phi_i(x_{k-1}) \neq x_{k-1}\}\,,$$

[27] An alternative dynamics has also been studied in [Gri+19]. Indeed, in "tournament dynamics", within the framework of Assumption 9, i_k^x is the agent who provides the best improvement:

$$i_k^x = \arg\max_{i \in \{1,2,\ldots,N_1\}} V(\phi_i(x_{k-1}))\,.$$

We will come back to these dynamics in Section 7.

where we are assuming that x is not a fixed point for all agents. We will use the former because, in some cases, we will keep track of the number of agents that give a solution until the final solution is reached. See next section for more details.

6.2.3. *Random V and random group.* Consider the case where V is a random function. At a general level, we would have

$$\begin{aligned} V : X \times \Omega_{\mathrm{V}} &\to [0,1] \\ (x,\omega) &\mapsto V(x,\omega)\,, \end{aligned} \tag{6.2}$$

where Ω_{V} is the sample space for the possible V functions. Similarly, we consider a random group of agents Φ from a pool of agents that will be determined for each case. Again, at a general level, we would have

$$\Phi(\omega_{\mathrm{G}}) \subset \Phi \quad \text{where } \omega_{\mathrm{G}} \in \Omega_{\mathrm{G}}\,, \tag{6.3}$$

where Ω_{G} is the sample space for the possible groups of agents. Therefore, if ν is the uniform distribution over X, then the expected value of the performance of the group is

$$\mathbb{E}_{\nu}(V \circ \phi^{\Phi})(\omega_{\mathrm{V}},\omega_{\mathrm{G}}) = \frac{1}{n}\sum_{i=1}^{n} V\left(\phi^{\Phi(\omega_{\mathrm{G}})}(i),\omega_{\mathrm{V}}\right)\,. \tag{6.4}$$

Note that the expected value is random, since it depends on ω_{V} and ω_{G}. Let us assume that on the space $\Omega_{\mathrm{V}} \times \Omega_{\mathrm{G}}$, we have defined a measure $\tilde{\mu}$ that will be specified for each case. Then, we can define the expected value of the performance of the group as

$$\mathbb{E}_{\tilde{\mu}}\left(\mathbb{E}_{\nu}(V \circ \phi^{\Phi})\right) = \int_{\Omega_{\mathrm{V}}\times\Omega_{\mathrm{G}}} \mathbb{E}_{\nu}\left(V(\omega_{\mathrm{V}}) \circ \phi^{\Phi(\omega_{\mathrm{G}})}\right) d\tilde{\mu}(\omega_{\mathrm{V}},\omega_{\mathrm{G}})\,. \tag{6.5}$$

In general, as standard, we will estimate the expected value of the performance of the group by sampling the random variables ω_{V} and ω_{G} from the measure $\tilde{\mu}$, not always independently[28]. That is,

$$\hat{\mathbb{E}}_{\tilde{\mu}}(\mathbb{E}_{\nu}(V \circ \phi^{\Phi})) := \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_{\nu}(V(\omega_{\mathrm{V}}^{m}) \circ \phi^{\Phi(\omega_{\mathrm{G}}^{m})})\,, \tag{6.6}$$

where ω_{V}^{m} and ω_{G}^{m} are sampled from the measure $\tilde{\mu}$ for $m = 1,2,\ldots,M$. To simplify the notation, we will denote

$$\hat{\mathbb{E}}_{\tilde{\mu}}(\mathbb{E}_{\nu}(V \circ \phi^{\Phi})) =: \hat{\mathcal{A}}(\Phi) \equiv \hat{\mathcal{A}}_{M}(\Phi)\,, \tag{6.7}$$

representing the estimated ability of the group.

Similarly, we can estimate other quantities of interest, such as the value of the stopping point defined in Assumption 9, which measures the number of steps taken by the group to find the maximum. This will be denoted as $\hat{\mathbb{E}}_{\tilde{\mu}}(\tau)$. We will denote this by $\hat{\tau}$, i.e.,

$$\hat{\tau} := \hat{\mathbb{E}}_{\tilde{\mu}}(\tau)\,. \tag{6.8}$$

Also, we will be interested in the number of times one group outperforms another group. That is, for groups Φ_1 and Φ_2, we will be interested in if $\mathbb{E}_{\nu}(V \circ \phi^{\Phi_1}) > \mathbb{E}_{\nu}(V \circ \phi^{\Phi_2})$ for a given realization of ω_{V}, $\omega_{\mathrm{G},1}$ and[29] $\omega_{\mathrm{G},2}$. So we will estimate

$$\mathbb{E}_{\tilde{\mu}}\left(\mathbb{1}_{\mathbb{E}_{\nu}(V\circ\phi^{\Phi_1})>\mathbb{E}_{\nu}(V\circ\phi^{\Phi_2})}\right)\,, \tag{6.9}$$

[28]That is, the measure $\tilde{\mu}$ is not necessarily a product measure. But, in any case, the sampling of the sequence is obviously independent so the law of large numbers holds.

[29]We will use the notation $\omega_{\mathrm{G},1}$ and $\omega_{\mathrm{G},2}$ to denote the group ω_{G} for the groups Φ_1 and Φ_2 respectively. In our cases one of the groups will depend entirely on ω_{V}, the group of best-performing agents, see below.

using the standard estimator as above. This, obviously, is the estimation of the probability that the group Φ_1 outperforms the group Φ_2. We will denote this by $\hat{\mathcal{O}}_{\Phi_1,\Phi_2}$, i.e.,

$$\hat{\mathcal{O}}_{\Phi_1,\Phi_2} := \hat{\mathbb{E}}_{\tilde{\mu}}\left(\mathbb{1}_{\mathbb{E}_\nu(V\circ\phi^{\Phi_1})>\mathbb{E}_\nu(V\circ\phi^{\Phi_2})}\right). \tag{6.10}$$

Remark 6.1. Note that, in the following, the convergence of these Monte Carlo estimates should be analyzed in more detail. The number of samples might not be enough to ensure convergence, but it is enough for us; we are performing the same analysis as in [HP04] (with far larger M) or [Gri+19] and we will go further and show some convergence plots. In any case, in a real-world setting, $M = 1$, so, for instance, showing that Φ_B outperforms Φ_R 95% for 2000 samples is enough to reject the "Diversity Trumps Ability" result. ◇

In the subsections, we will set ν to be the uniform distribution over X. We will, in general, compare two groups, i.e., define two different $\tilde{\mu}$ measures. For what we will call the "random group", ω_G is sampled independently from ω_V. For what we will call the "ability" group, ω_G depends on ω_V. That is, the group will be chosen depending on the function V, so these events are not independent.

6.2.4. *Hong-Page original simulations.* Let us first replicate the simulations of [HP04]. For this, we set $n = 2000$ and $k = 3$. Importantly, V is a random function, and the number of random V functions is $M = 1000$. More specifically, V is of the form

$$V(x,\omega_V) = U_x(\omega_V), \tag{6.11}$$

where $\{U_x\}_{x\in X}$ is a set of independent uniform random variables over $[0,1]$, and ω_V is as in (6.2).

Let us replicate the results of [HP04], Table 1. This is done in Tables 1, 2, and 3. First, if $l = 12$, Table 1 shows that the random group outperforms the best group. Increasing the number of elements in each group to 20, Table 2 shows, again, that the random group outperforms the best group, now with a smaller advantage. Set $l = 20$. Table 3 shows that the random group outperforms the best group, with a slightly larger advantage.

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$
0.946 (0.006)	13.717 (0.422)	0.926 (0.012)	12.618 (0.587)	0.000 (0.000)

TABLE 1. Results of the expected values and the number of iterations with n=2000, $l = 12$, M=50 number of trials, N_1=10 elements in the random group. In parenthesis, the standard deviation of the expected values.

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$
0.948 (0.005)	24.092 (0.436)	0.936 (0.010)	23.658 (0.961)	0.000 (0.000)

TABLE 2. Results of the expected values and the number of iterations with n=2000, $l = 12$, M=50 number of trials, N_1=20 elements in the random group. In parenthesis, the standard deviation of the expected values.

We can also replicate the results of [Gri+19, Section 2]. This is done in Table 4. Furthermore, we can analyze the convergence of the expected value of the performance of the group as a function of the number of V functions. This is done in Figure 2.

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$
0.962 (0.005)	15.194 (0.433)	0.938 (0.013)	13.183 (0.761)	0.020 (0.141)

TABLE 3. Results of the expected values and the number of iterations with n=2000, $l = 20$, M=50 number of trials, N_1=10 elements in the random group. In parenthesis, the standard deviation of the expected values.

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$
0.944 (0.006)	12.526 (0.360)	0.922 (0.014)	11.397 (0.606)	0.020 (0.140)

TABLE 4. Results of the expected values and the number of iterations with n=2000, M=1000 number of trials, N_1=9 elements in the random group. In parenthesis, the standard deviation of the expected values.

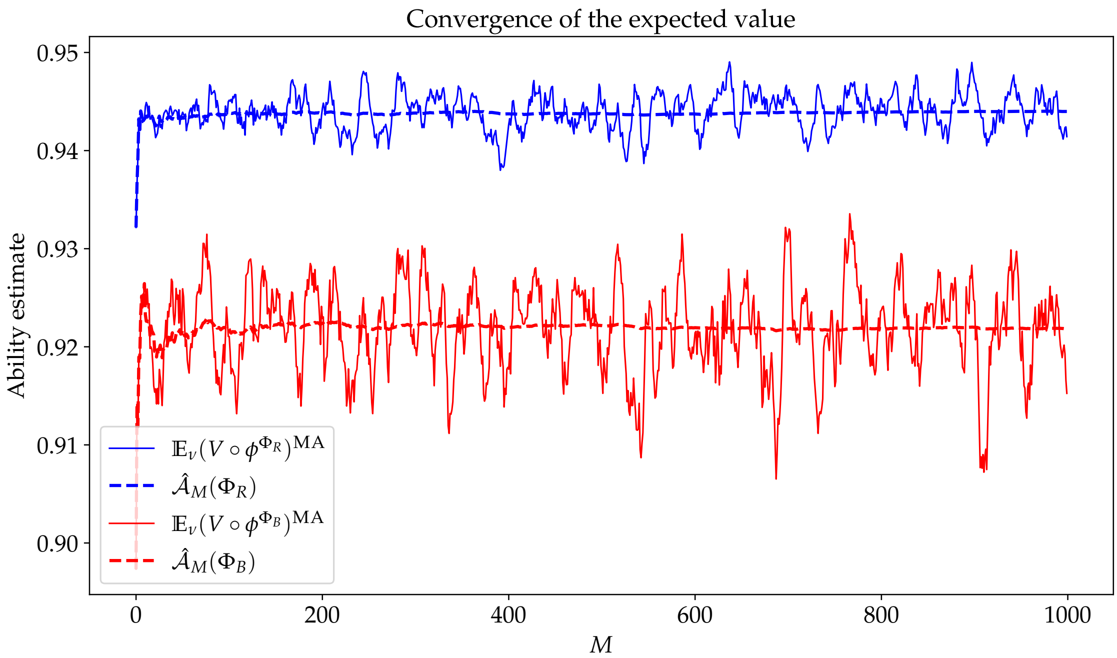


FIGURE 2. Convergence of the expected value of the performance of the group as a function of the number of V functions. The values $\mathbb{E}_\nu(V \circ \phi^{\Phi})$ are computed for $M = 1000$ with a moving average window of 10.

6.3. **Hong-Page framework but ability trumps diversity.** While maintaining the same framework, we are going to change the parameters so that ability trumps diversity. As we said, we do not aim to perform a comprehensive analysis[30], but to show that the results are not robust to the parameters, showing the weakness of the Hong-Page conclusion. That is, we will show that there are a great number of cases where ability trumps diversity even in the original framework presented by Hong and Page, although these cases were not included in the original paper.

6.3.1. *Reducing the domain.* First, and trivially, we can see what happens if we reduce the domain of V, whose length is n. This is done in Tables 5 and 6. This can make sense in a real-world setting, where the domain is not infinite, but finite, and we cannot always assume that all the agents can evaluate 2000 different points. As we can see, the best group outperforms in almost half of the cases, and the estimation of the expected ability is slightly greater than that of the random group.

To ensure that the results were not generated by chance, we show the evolution of the expected

$\hat{A}(\Phi_R)$	$\hat{\tau}^R$	$\hat{A}(\Phi_B)$	$\hat{\tau}^B$	$\hat{O}_{\Phi_B,\Phi_R}$
0.981 (0.026)	23.886 (2.177)	0.983 (0.019)	22.139 (1.622)	0.426 (0.495)

TABLE 5. Results of the expected values and the number of iterations with n=25, l=12, M=1000 number of trials, N_1=20 elements in the random group. In parenthesis, the standard deviation of the expected values.

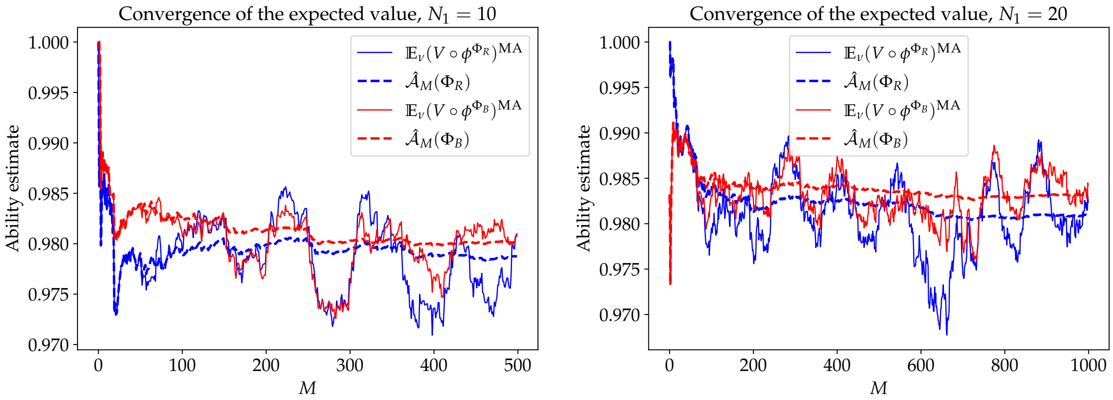
$\hat{A}(\Phi_R)$	$\hat{\tau}^R$	$\hat{A}(\Phi_B)$	$\hat{\tau}^B$	$\hat{O}_{\Phi_B,\Phi_R}$
0.979 (0.026)	13.499 (1.660)	0.981 (0.019)	11.454 (0.745)	0.430 (0.496)

TABLE 6. Results of the expected values and the number of iterations with n=25, l=12, M=500 number of trials, N_1=10 elements in the random group. In parenthesis, the standard deviation of the expected values.

value of the performance of the group as a function of the number of V functions. This is done in Figure 3. There we can see that the convergence is "slower", but the expected ability of the best group is still slightly greater than that of the random group.

The results are clearer if we increase the ratio n/l, so the problem is not too easy. This is done, for instance, in Tables 7 and 8. We can see that the best group outperforms the random group more often than the other way around and that the estimated ability of the best group is significantly greater than that of the random group. In the following sections, we will see what can be causing this result. But here, note that in the original paper, Hong and Page only include the case $n = 2000$, instead of more realistic cases where the motto "diversity trumps ability" might be false. Recall this choice implies agents can evaluate and rank exactly 2000 solutions to a problem and, consequently, can determine the superior option among nearly 2 million pairs of solutions, with all agents agreeing on this ordering.

[30] As we will see in Section 7, the framework has fundamental problems that go beyond the simulations and its lack of robustness.



(A) $N_1 = 10$. (B) $N_1 = 20$.

FIGURE 3. Convergence of the expected value of the performance of the group as a function of the number of V functions. The values $\mathbb{E}_v(V \circ \phi^{\Phi})$ are computed with a moving average window of 50.

$\hat{A}(\Phi_R)$	$\hat{\tau}^R$	$\hat{A}(\Phi_B)$	$\hat{\tau}^B$	$\hat{O}_{\Phi_B,\Phi_R}$
0.928 (0.044)	11.213 (0.377)	0.938 (0.038)	11.046 (0.383)	0.754 (0.431)

TABLE 7. Results of the expected values and the number of iterations with n=25, l=5, M=2000 number of trials, N_1=10 elements in the random group. In parenthesis, the standard deviation of the expected values.

$\hat{A}(\Phi_R)$	$\hat{\tau}^R$	$\hat{A}(\Phi_B)$	$\hat{\tau}^B$	$\hat{O}_{\Phi_B,\Phi_R}$
0.947 (0.036)	21.893 (0.747)	0.953 (0.031)	21.645 (1.062)	0.655 (0.475)

TABLE 8. Results of the expected values and the number of iterations with n=30, l=7, M=2000 number of trials, N_1=20 elements in the random group. In parenthesis, the standard deviation of the expected values.

6.3.2. *Larger group size: analysis of the best-performing agents.* [HP04] argue that "Diversity Trumps Ability" because the elements of the best group tend to be similar. As we saw in the analysis of the theorem, Section 2, this was an assumption, not a consequence of reasonable hypotheses. In the simulations, something similar happens. By Hong and Page's construction of the agents, given a V function, the best group consists of the best individually performing agents. However, in this setting, the best-performing agents, given certain V functions, are those with heuristics containing similar numbers, as described in Algorithm 1. For instance, when $k = 3$, the best-performing agents share similar numbers $\alpha = (a_1, a_2, a_3)$, while the random group has more diverse numbers. As the number of elements is large enough and adding more elements does not worsen the result in this setting, the diversity group performs better. Again, there is no surprise in this result. The same critique as in Section 2.6 applies. In the following sections, we will deepen the critique, but first, let us go into more detail on the best-performing agents and why the setting provided by Hong and Page creates this result.

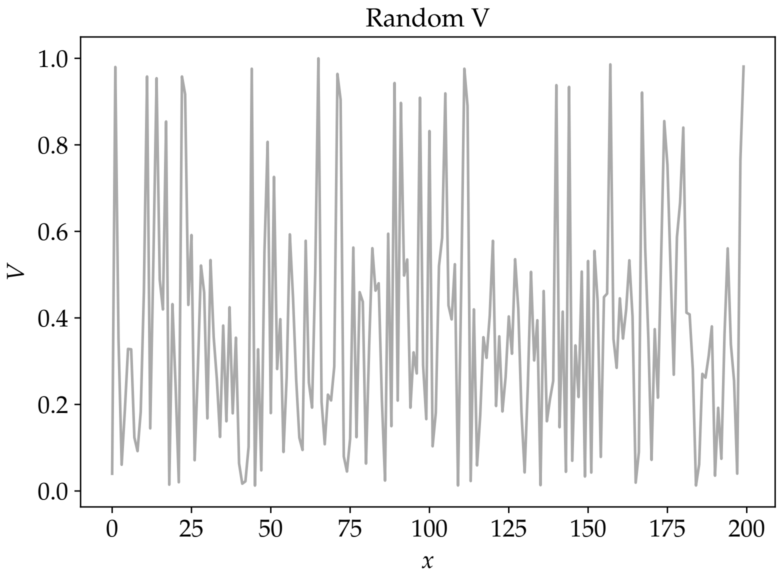


FIGURE 4. Plot of the function V over X.

For instance, in Table 9, we can see the heuristics of a random and best groups for a random V function, Figure 4. The best group has heuristics containing similar numbers a_i, while the random group has more diverse heuristics. In particular, the random group has heuristics with all the numbers from 1 to 12 except 11, while the best group has heuristics that miss the numbers $\{1, 3, 5, 6, 9, 10, 11\}$ and the agents are just permutations of the others[31]. For instance, up to permutations, the best group has only 2 agents (the first two in the table). Let us call N_B the number of agents in the best group without considering permutations. In this case, $N_B = 2$. In Section 6.4.1, we will see more statistics on the number of agents in the best group.

The best agents, individually, can perform well on certain regions of the landscape, but as a group, they are sometimes unable to move out of certain "basins of attraction", especially when starting from points where their limited heuristics do not provide particular jumps to reach higher peaks. This can be seen in Figure 5.

For instance, consider the point $x = 175$. The random group will move to $x = 180$, a better point, while the best group's agent will stay at $x = 175$. This is because the random group has a heuristic that contains the number 5, while the best group has no agent with a heuristic that contains the number 5. Similarly, for the point $x = 198$ (penultimate point), the random group will move to $x = 1$, a better point, while the best group's agent will stay at $x = 198$. This is because the random group has a heuristic that contains the number 3, while the best group has no agent with a heuristic that contains that number. There is no surprise in this result; the random group is better because, by construction of the simulations, it has more diverse heuristics that allow it to move to nearby better

[31]A permutation of a heuristic is a heuristic that contains the same numbers, but in different orders. This agent is not the same as the original one; the order is important for moving to the next point, but facing a stopping point, the order is not important.

Random	Best
[8,9,4]	[2,7,12]
[3,10,1]	[4,7,8]
[8,7,5]	[8,4,7]
[6,1,2]	[4,8,7]
[5,12,1]	[8,7,4]
[10,7,2]	[2,12,7]
[6,7,8]	[7,8,4]
[10,12,7]	[7,12,2]
[7,1,8]	[7,4,9]
[6,5,10]	[12,7,2]

TABLE 9. Comparison of random and best heuristics for a particular V function and random sample.

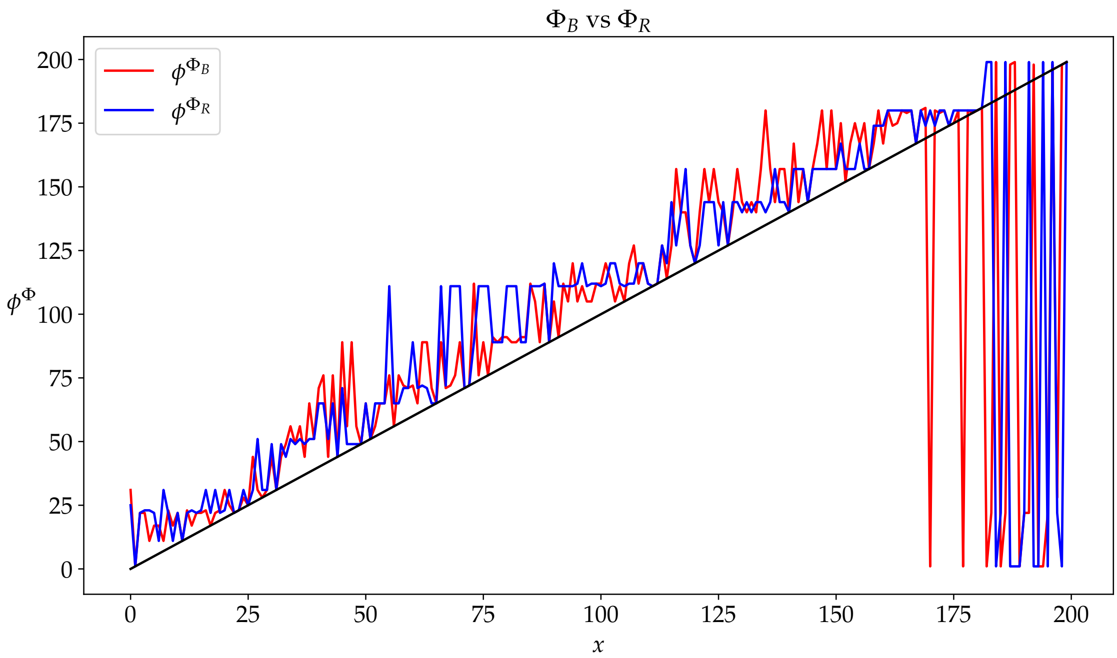


FIGURE 5. Plot of the functions $\phi^{\Phi} : X \to X$. In black the identity function.

points while the best group gets stuck in a local maximum more often[32]. This will be explored in more detail below, where we will see V functions that do not have this property, for instance, there might not be close enough good points to move to, so adding more agents will not help. In these cases, Hong-Page's explanation and theorem fail completely.

Having said this, an immediate question is whether this holds as the group size increases; if we increase the group size, will the best group still consist of similar elements? This is analyzed

[32]It is therefore the number of distinct α_i in the heuristic, and not randomness per se (as claimed in [Tho14]) or diversity in the sense of [HP04] what explains the original result. For an statistical analysis of this point, see [Sin19]

in Figure 6. There is a small, but significant, difference in the expected value of the performance of the group, and the best group outperforms the random group more often than the other way around. The logic of this result will be understood better in the following sections, but by now, we can already see that if we increase the number of elements in the group, the best group will have more agents with heuristics that contain new numbers to move to better points and not get stuck in a local maximum more often. Note that in the original paper, Hong and Page noted in their Table 1 that the outperformance of the random group was higher when $N_1 = 10$ than when $N_1 = 20$, and, in the *Concluding remarks*, they said if the group size is large, the best group will be more diverse and perform relatively better, but they did not include the case of larger N_1 in their numerical analysis that would have shown that "diversity trumps ability" is false there.

A similar result is obtained for $l = 15$ and $l = 20$; when the group size is large enough, there is a slight advantage for the best group and it is more likely to outperform the random group. See Figure 7 for more details.

6.3.3. *The broader implications of this result.* Thus, based on the previous simulations and fixing the number of agents, N, there are three possible outcomes[33].

(1) N_1 is "small" enough. Then, "ability trumps diversity" is *true*.
(2) N_1 is "large" enough. Then, "ability trumps diversity" is *true*.
(3) N_1 is "medium". Only then "diversity trumps ability" is *true*.

So, the "diversity trumps ability" result is only true for a "medium" number of agents. As we have seen Section 6.1.1, the theorem is of no use to determine the number of agents, as it needs clones, and thus, N can be as large as we want, but here it is fixed. Thus, it seems to be the case that we have these inequalities, at least for the parameters used in [HP04],

$$\mathcal{A}(\Phi_R^{N_1^S}) < \mathcal{A}(\Phi_B^{N_1^S}) < \mathcal{A}(\Phi_B^{N_1}) < \mathcal{A}(\Phi_R^{N_1}) < \mathcal{A}(\Phi_R^{N_1^L}) < \mathcal{A}(\Phi_B^{N_1^L}), \tag{6.12}$$

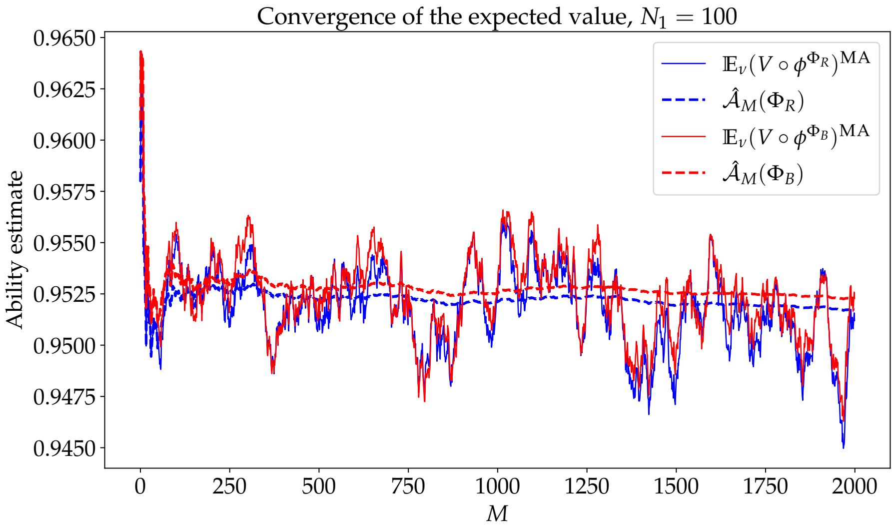
where N_1^S is the number of agents in the small group, N_1 is the number of agents in the medium group, and N_1^L is the number of agents in the large group. The conclusion drawn from here cannot be that "diversity trumps ability", as the result is not robust to the number of agents. Actually, the conclusion should be that "enough able agents trump any other consideration, in particular, diversity", so the best solution for any problem should be to have as many able agents as possible, i.e., a form of epistocracy[34]. It would only make sense, in the Hong-Page setting, to have to reject the group of the best agents if, somehow, we knew that we cannot increase N_1 and that N_1 is not small enough. That is, choosing a random group or "Diversity Trumps Ability" is, in their setting, a second-best solution.

Thus, if one were willing to apply this result in a real-life scenario (for instance, considering it reasonable to take $n = 2000$ rather than $n = 25$ as above), why assume that we are in the middle region rather than in one of the other cases? Furthermore, as we will see, incorporating modifications that bring the framework closer to a real-world setting tends to reduce the benefits of diversity—so much so that the slight advantage found above could become considerably larger. This is why these results refute the already disputed (for its naive flaws in its derivation[35]; see [Rom23; Rom25]) claim known as the "Numbers Trumps Ability" Theorem (sic). In essence, Landemore argues that

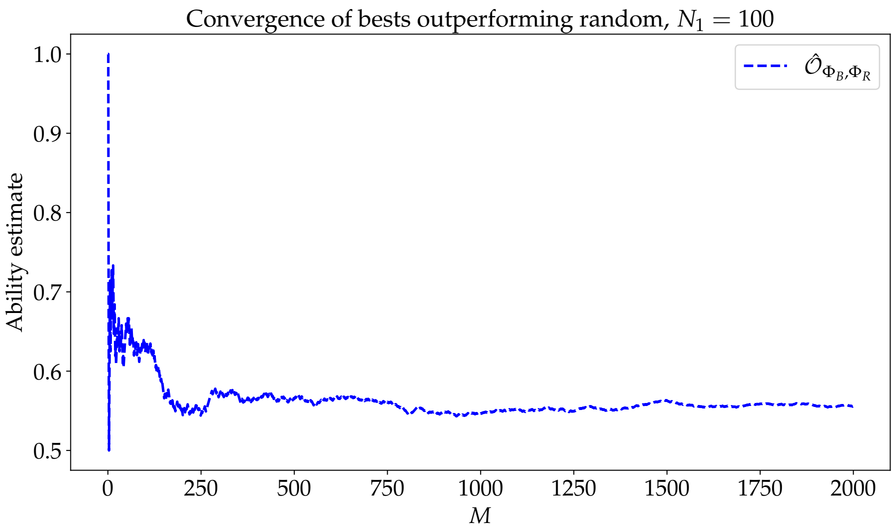
[33]The result for small N_1 follows from the fact that if $N_1 = 1$, by definition, the best group outperforms the random group. Therefore, it is natural to conclude that for a neighborhood (perhaps with just one element), the result holds. This is indeed the case; see [Gri+19, Footnote 9]. However, note that the third point here contradicts their conclusion that "The larger the group, all else being equal, the greater the advantage for groups of random heuristics." They only tested $N_1 \leq 9$.

[34]This general conclusion aligns with [Rom22], see also Section 1.

[35] Despite that in [Rom23] a great deal of effort was devoted to analyzing Landemore's misapplication of the theorem in Section 5, was thoroughly analyzed in Section 5.5, and was announced in the title, abstract, and introduction section, the



(A) Convergence of the expected value of the performance of the group as a function of the number of V functions. The values $\mathbb{E}_\nu(V \circ \phi^{\Phi})$ are computed for $M = 2000$ with a moving average window of 50.



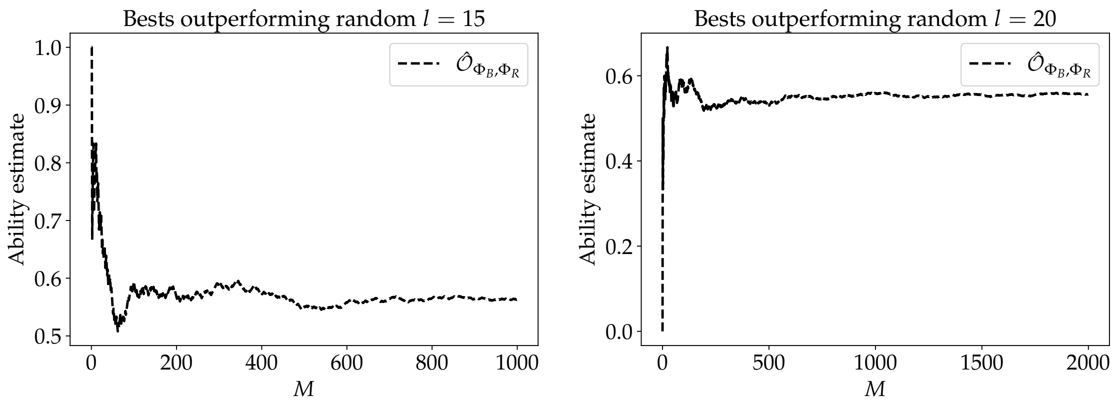
(B) Convergence of the probability that the best group outperforms the random group as a function of the number of V functions.

FIGURE 6. Convergence of expected value and best outperforming random for $N_1 = 100$.

---

authors of [LS24, Table 3] consider that the theorem is not treated in the paper, as the column "Numbers trumps ability?" says "Not reported". With this omission, Landemore and Sakai safely conclude:

> Landemore never formalized or tested any of these claims. Several studies (S3, S4, S5, S7 [S9=[Rom23] is missing]), however, have seemingly found support for the numbers trump ability theorem.



(A) Convergence of expected value and best outperforming random for $l = 15$.

(B) Convergence of expected value and best outperforming random for $l = 20$.

FIGURE 7. Convergence of expected value and best outperforming random for $l = 15$ and $l = 20$.

this "theorem" supports a strong epistemic case for democracy by emphasizing the instrumental—specifically epistemic—benefits of inclusiveness; namely, that under the right conditions, including everyone in the decision-making process makes the group more likely to arrive at the correct (or, at least, better) answers. However, for large-sized pools ("numbers") the result becomes "ability trumps diversity" (even within the original framework). That is, "Enlightened Numbers Trump Numbers" as we already discussed, [Rom23, Theorem 5.3]. If the Hong–Page framework were applicable, this outcome would not support democracy at all but rather an epistocracy composed of enough experts. In any case, note the unrealistic consequence of the framework—increasing the number of elements guarantees that the group reaches the correct solution in nearly all situations with unanimous agreement (cf. Proposition 2.12). This is a bug of the model, not a feature. Using this bug to derive conclusions, rather than to recognize its unsoundness and reject the hypotheses of the theorem, is, in any case, incorrect.

6.3.4. *Parametric V.* Also, we can consider a different function V, which is random but not necessarily of the form (6.11). Without the pretense of being general, we define the function[36]

$$V(x) := \frac{\tilde{V}\big((x+\delta) \bmod n\big)}{\max\limits_{0 \le j < n} \tilde{V}(j)}, \tag{6.13}$$

where

$$\tilde{V}(x) := \left|x - x_0\right|^{\gamma} \left(1 + \sin(\beta\, x)\, \left(\tfrac{x}{n}\right)^{\alpha}\right).$$

[36] That the DTA fails for different V functions was already shown in [Gri+19, Section 3]. There, they "smooth" (although they mean "slow variation," not smooth in a mathematical sense) a random V like the ones defined above; that is, they interpolate the values of the function at the points of the domain and show small differences in estimated ability. Here, the argument is slightly different: we generate a random V function by sampling the parameters α, β, γ, x_0, and δ from a uniform distribution over some interval. The differences in performance are much larger, as we will see. Additionally, depending on the parameters, the landscape is not necessarily of "slow variation" and can be very oscillatory (although generated by a smooth function on the real line).

Here, the parameters satisfy

$$\alpha,\ \beta,\ \gamma\ \in\ \mathbb{R}_{\geq 0},\ \text{and}\ x_0,\ \delta \in \{0, 1, \ldots, n-1\}.$$

Thus, if these parameters are random, V is a random function of the form (6.2). The interpretation of the parameters is straightforward. The parameter α controls the amplitude of the oscillations, β controls the frequency of the oscillations, γ controls the global amplitude as we move away from the point x_0, and δ is the shift of the function, which should not affect the results, but was included for visualization and testing purposes.

We can see what happens when we consider this random-parametric V function in Tables 10 and 11, and the convergence of the expected value of the performance of the group as a function of the number of V functions in Figure 8. The results show that when using this random-parametric V function, the best group (Φ_B) consistently outperforms the random group (Φ_R), both in expected ability and in outperformance probability. Additionally, the best group requires fewer iterations ($\hat{\tau}^B \approx 10.9$) to reach its solutions compared to the random group ($\hat{\tau}^R \approx 12.4$).

The convergence plot in Figure 8 further supports these findings, showing that after an initial period of volatility, both groups' performance metrics stabilize, with the best group maintaining a consistent advantage over the random group. This demonstrates that these results are robust across the 2000 trials.

So, these findings stand in contrast to the original Hong-Page simulations, suggesting that when using a more structured parametric V function, ability can indeed trump diversity. This provides another example where slight modifications to the original framework can lead to significantly different conclusions about the relative importance of ability versus diversity in group performance.

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$
0.652 (0.187)	12.429 (2.271)	0.678 (0.184)	10.901 (0.227)	0.861 (0.346)

TABLE 10. Results of the expected values and the number of iterations with n=200, l=20, M=2000 number of trials, N_1=10 elements in the random group, V functions are parametrically generated. In parenthesis, the standard deviation of the expected values.

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$
0.623 (0.205)	12.260 (3.238)	0.641 (0.200)	10.871 (0.224)	0.796 (0.403)

TABLE 11. Results of the expected values and the number of iterations with n=200, l=12, M=2000 number of trials, N_1=10 elements in the random group, V functions are parametrically generated. In parenthesis, the standard deviation of the expected values.

We can focus on some particular cases to gain more insight, for instance, in Figure 9. There, two cases are shown, with $(\alpha, \beta, \gamma, x_0, \delta) = (0.657, 0.397, 1.5, 12, 0)$ and $(\alpha, \beta, \gamma, x_0, \delta) = (0.386, 1.577, 1.5, 6, 189)$. Note now that the best group outperforms the random group by far more than in the previous cases where the best group was not able to outperform the random group. The differences are of 10 or 20 percentage points, not just a few as in the Hong-Page cases or the ones studied in [Gri+19]. Clearly, ability trumps diversity for these particular landscapes.

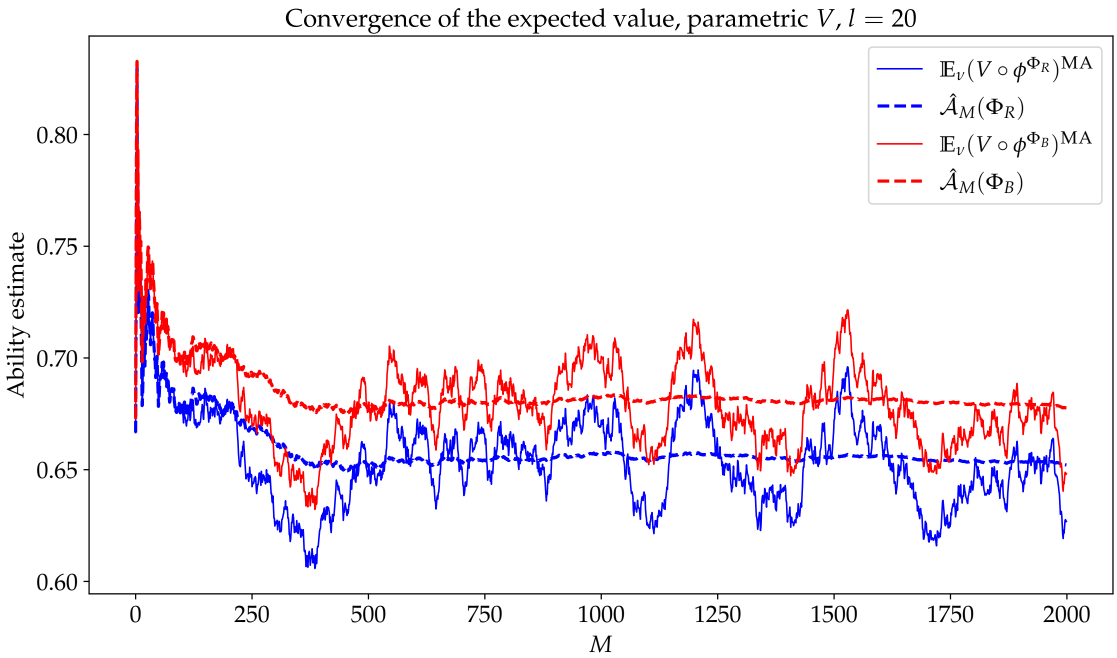


FIGURE 8. Convergence of the expected value of the performance of the group as a function of the number of V functions. The values $\mathbb{E}_\nu(V \circ \phi^{\Phi})$ are computed for $M = 2000$ with a moving average window of 50.

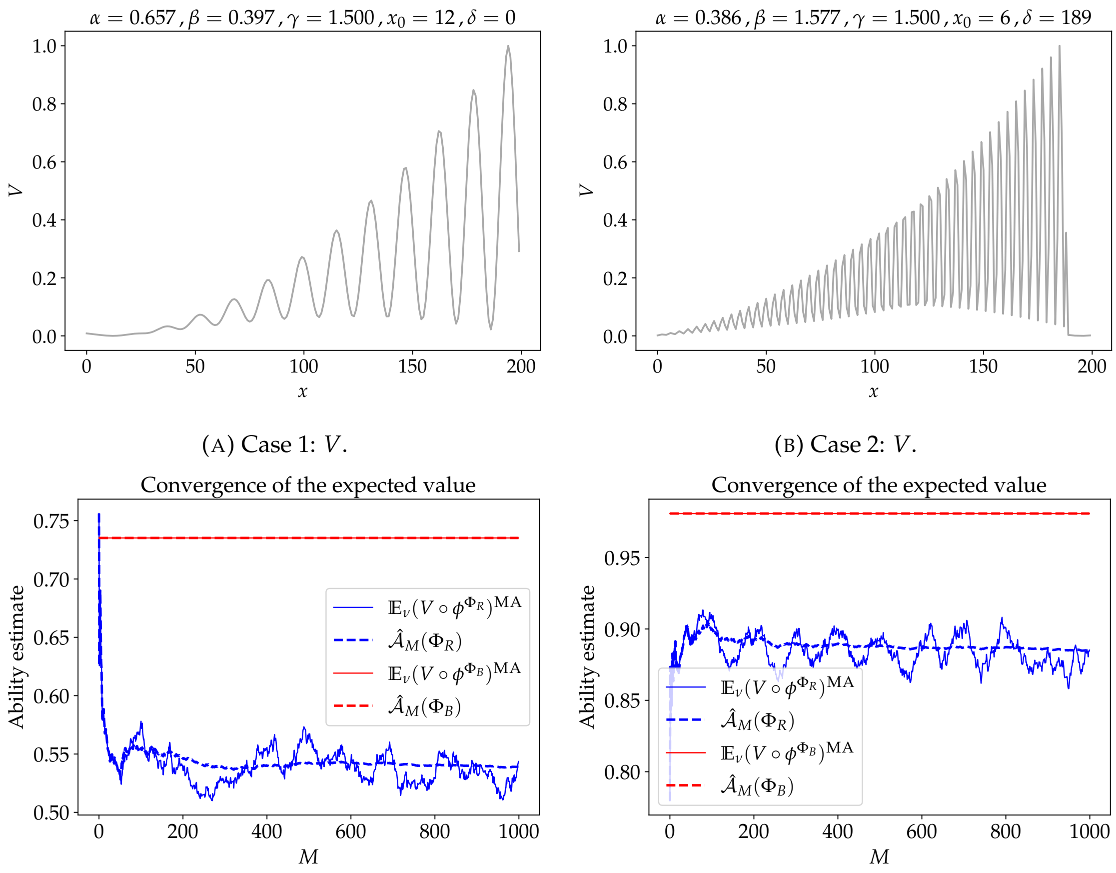
Tables 12 and 13 show the results for the two cases of Figure 9. Note that the problem is harder in the sense that the estimated ability is lower than in the original simulations.

$\hat{A}(\Phi_R)$	$\hat{\tau}^R$	$\hat{A}(\Phi_B)$	$\hat{\tau}^B$	$\hat{O}_{\Phi_B,\Phi_R}$
0.539 (0.101)	14.794 (5.471)	0.735 (0.000)	10.920 (0.000)	0.927 (0.260)

TABLE 12. Results of the expected values and the number of iterations with n=200, l=12, M=1000 number of trials, N_1=10 elements in the random group, V function is parametrically generated with $\alpha, \beta, \gamma, x_0, \delta = 0.657, 0.397, 1.5, 12, 0$. In parenthesis, the standard deviation of the expected values.

$\hat{A}(\Phi_R)$	$\hat{\tau}^R$	$\hat{A}(\Phi_B)$	$\hat{\tau}^B$	$\hat{O}_{\Phi_B,\Phi_R}$
0.885 (0.087)	11.115 (0.297)	0.981 (0.000)	10.980 (0.000)	0.938 (0.241)

TABLE 13. Results of the expected values and the number of iterations with n=200, l=12, M=1000 number of trials, N_1=10 elements in the random group, V function is parametrically generated with $\alpha, \beta, \gamma, x_0, \delta = 0.386, 1.577, 1.5, 6, 189$. In parenthesis, the standard deviation of the expected values.



(A) Case 1: V.

(B) Case 2: V.

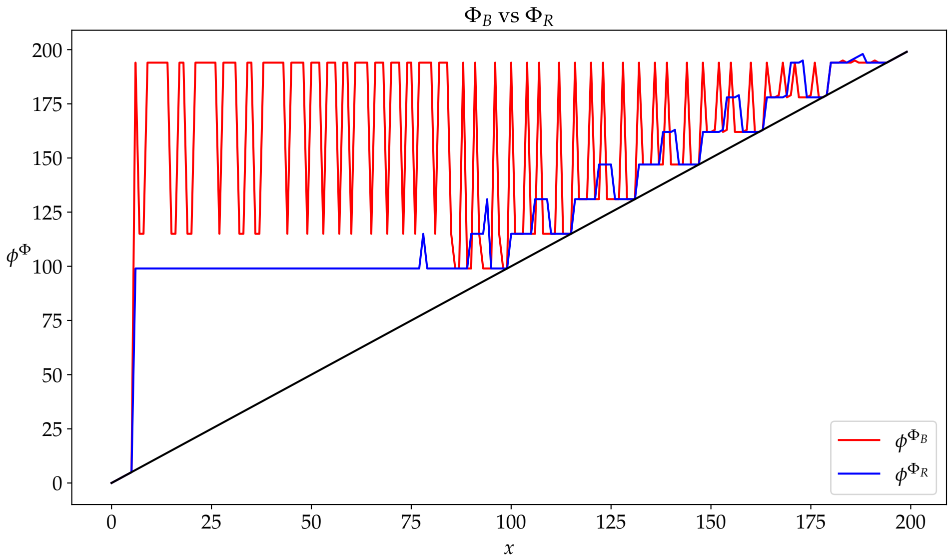
(C) Case 1: convergence.

(D) Case 2: convergence.

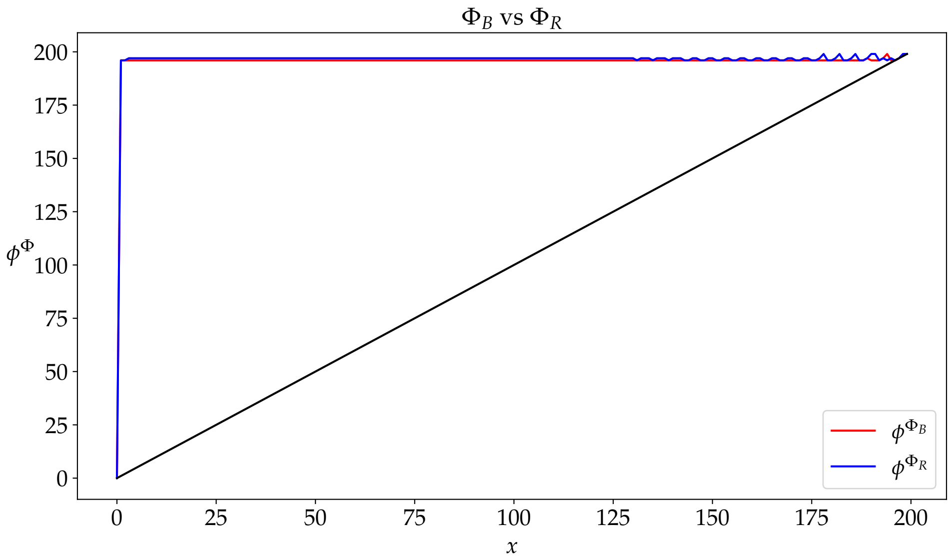
FIGURE 9. Convergence of the expected value of the performance of the group as a function of the number of trials for two particular cases of the parametric V. The values $\mathbb{E}_\nu(V \circ \phi^{\Phi})$ are computed for $M = 1000$ with a moving average window of 50.

The explanation of these results can be found in Figure 10a. For the same case as in Figure 9a, we can see the performance of the random and best groups. As we see in this figure, the maximum is at x near 200, and there are several local maxima starting from $x = 50$ with a period of approximately 15 (> 12). Figure 10a shows that the random group gets stuck in these local maxima far more often than the best group, only achieving the maximum at a few points where x is near the maximizer. In this landscape, in contrast to the random ones where there are closer enough good points to any starting point, the best group is able to escape these local maxima, but the diversity of the random group is not useful for this task. The basic reason is that the best agents, by definition, are better in avoiding these local maxima. Here, adding more agents with the different heuristics does not help, as this will only add numbers from 1 to 12 to the heuristics of the best group, which might not be useful to move to a further maximum. The best idea is to avoid these local maxima in the first place. As we stated earlier, for these harder problems (the expected ability is now lower and not between

0.9 and 1 as before), the magnitude of the difference in performance can be orders of magnitude larger than the previous advantage of the random group. That is, for these difficult problems, the random group is not the better choice. Although [HP04] (or [Pag07]) claim to be interested in hard/difficult problems, they did not include these types of landscapes in their analysis.



(A) Case 1



(B) Case 2

FIGURE 10. Comparison of a random and the best groups for the cases of Figure 9.

We can perform the analysis for Case 2 in Figure 9b, where the reasoning is different. Now, as we can see in that figure, the maximum is around $x = 185$, and after that a sudden drop in the value of the function with a much worse local maximum. So one problem could be that we go "too far" and jump to that maximum. This can happen, for instance, if we are in one of the local minima just before the global maximum. The best problem solvers will be able to avoid this problem, but the random group with random steps can fall into this trap. Actually, that is what we see in Figure 10b. The best group is able to avoid this problem in most cases, obtaining an almost perfect score, but this particular random group is not able to do so, and for almost all initial points, it ends up overshooting into a worse local maximum.

The conclusion is that ability trumps diversity for these particular landscapes. For these V functions the difference is substantial, ability is much important than diversity, which can be a disadvantage. Therefore, the results of [HP04] are not robust to the parameters, and the "Diversity Trumps Ability" result does not hold. While maintaining exactly the same framework (which is highly specific), the results change dramatically. Therefore, just testing for different landscapes in [HP04] would have been enough to show that the slogan "Diversity Trumps Ability" is not correct. Nevertheless, the authors of [HP04] did not include these types of landscapes in their analysis.

6.4. **More realistic framework: ability trumps diversity.** As we have seen in Proposition 2.12, Hong-Page's framework is not realistic, as deliberations always end up in unanimous agreement and, with enough elements, the random group always reaches the correct solution for every V function and for every initial point. As we did in Section 4, we can consider a more realistic framework also in the simulations. This is the content of this section. When a more realistic framework is considered, the results change dramatically, showing that ability trumps diversity. We will introduce these elements following the lines of Section 4.

6.4.1. *Heuristics model is arbitrary.* As we have seen in Section 6.3.2, one property of the setting used by Hong and Page is that the best group consists of similar elements that are outperformed by the random group, if we tune the parameters properly. But this fact is just a property or deficiency of the model, not a fact of the real world. Actually, the algorithm, Algorithm 1, that leads to this property is quite a "peculiar" way to find the maximum, and it is not clear what the real-world counterpart of this procedure is. In contrast, as is standard, the best algorithm one can devise is to check every element in the list and select the one with the highest value. This can be a computationally expensive operation, so a natural heuristic is to check a random subset[37] of the list and select the one with the highest value. Furthermore, this actually has a clear interpretation in the landscape setting: these agents are experts in a particular subset of the domain, and they are able to find the maximum in that subset when compared with the starting point. More precisely, the agents are as in Algorithm 6.2.

Algorithm 6.2: Subset Heuristic

Input: Current element $i \in X = \{1, 2, \ldots, n\}$, subset $\alpha \subseteq X$, value function V
Output: Maximizer j^*
```
// Form the region to search by adding the current element to the subset
```
$\mathcal{R} \leftarrow \alpha \cup \{i\}$
```
// Find the maximizer in R
```
$j^* \leftarrow \arg\max_{j \in \mathcal{R}} V(j)$
return j^*

Considering more algorithms makes the model more realistic, as it reflects the fact that there are several approaches to solving a problem and we do not need to be restricted to a single algorithm. But it is enough to introduce a small number of these agents in the pool to select the best ones, of the thousands of possible heuristics, to debunk the DTA result. For instance, in Table 14, we can see the results of a random and thebest group for a random V function, where we have added 5 additional agents with these heuristics and a (random) subset of 5 elements (of 200). This is enough to show that the best group outperforms the random group. Let us call N_I the number of agents in the best group that are using this new heuristic. In this case, $N_I \leq 5$. The logic is clear: these agents are experts in a particular subset of the domain, so they are able to move to better points and not get stuck in a local maximum more often. These agents, complemented with the best heuristics, are able to outperform the random group. The fact that the best group is composed of similar elements is a consequence of the arbitrary nature of the model hypothesis and disappears when we consider a more realistic framework.

Something even more intense happens in the parametric case. As we can see in Table 15, the best group outperforms the random group by far more than in the previous cases, around 35 percentage points, compared with 2 percentage points (which were enough to say in [HP04] that ability trumps diversity).

[37] For an even fairer comparison, we have excluded the case where the random subset includes the global maximum.

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$	N_B	N_I
0.95 (0.01)	13.70 (0.75)	0.97 (0.02)	12.27 (2.66)	0.82 (0.38)	4.42 (1.29)	2.38 (1.14)

TABLE 14. Results of the expected values and the number of iterations with n=200, l=12, M=2000 number of trials, N_1=10 elements in the random group, 5 additional ϕ functions. In parenthesis, the standard deviation of the expected values.

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$
0.534 (0.101)	14.531 (5.268)	0.881 (0.102)	13.103 (2.524)	0.977 (0.152)

TABLE 15. Results of the expected values and the number of iterations with n=100, l=12, M=2000 number of trials, N_1=10 elements in the random group, V function is parametrically generated with $\alpha, \beta, \gamma, x_0, \delta = (0.657, 0.397, 1.5, 12, 0)$, 5 additional ϕ functions. In parenthesis, the standard deviation of the expected values.

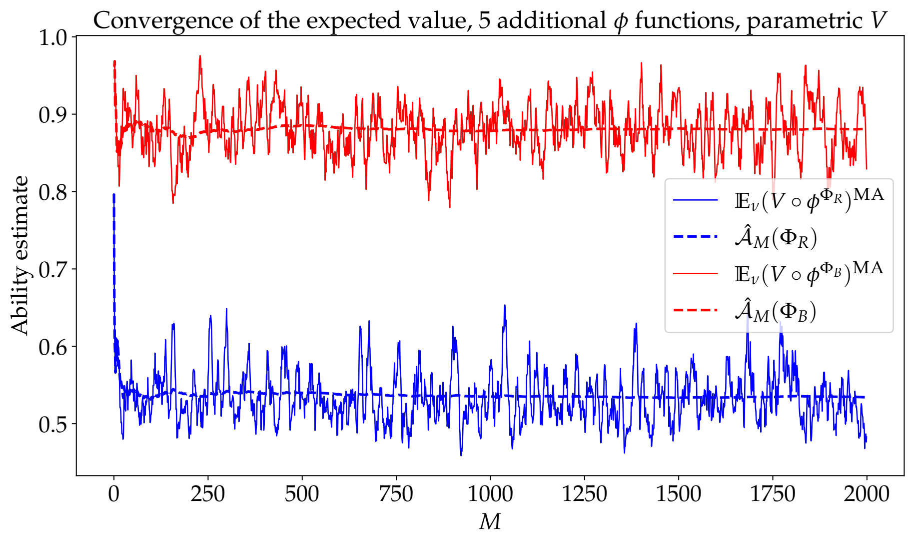


FIGURE 11. Convergence of the expected value of the performance of the group as a function of trials for a parametric V function. The values $\mathbb{E}_\nu(V \circ \phi^{\Phi})$ are computed for $M = 2000$ with a moving average window of 10. Note that the best group outperforms the random group by far more than in the previous cases, around 35 percentage points, compared with 2 percentage points in [HP04] that the random group outperformed the best group.

6.4.2. *Comparing the computational cost.* We can compare the average number of evaluations of the costly function V required for different types of ϕ functions across two distinct V function cases: Parametric V with parameters as in Case 1 of Figure 9 and the random V function used in the previous section, Figure 4. Note that evaluating V must be computationally expensive; otherwise, computing the maximum would be trivial.

First, the subset heuristic needs less than 1 evaluation per initial point for both types of V functions. Indeed, these agents are experts in a particular subset of the domain, so they have cached the value of the function in that subset and do not need to evaluate it again[38]. Therefore, given a point x, the subset heuristic will only evaluate the function at x if x is not in the subset. More precisely, we have the following modification (equivalent in terms of output) of the Algorithm 6.2:

Algorithm 6.3: Subset Heuristic (with precached values)

Input: Current element $i \in X = \{1, 2, \dots, n\}$, function V
Subset $\alpha \subseteq X$ (with precomputed $V(j)$ for each $j \in \alpha$)

Output: j^*, the element in $\alpha \cup \{i\}$ maximizing V

if $i \notin \alpha$ **then** `// Compute only if needed`
  Compute $V(i)$
Let $j^*_\alpha = \arg\max_{j \in \alpha} V(j)$ (cached)
if $V(i) \geq V(j^*_\alpha)$ **then** `// Check if` i `beats or ties the best in` α
  $j^* \leftarrow i$
else
  $j^* \leftarrow j^*_\alpha$
return j^*

In stark contrast, heuristics-based best ϕ functions necessitate significantly more evaluations, averaging 10.06 for the parametric V and 5.49 for the random function per initial point. This substantial difference highlights the limitations of Hong-Page's heuristics. Clearly, if there is a less computationally expensive alternative achieving similar results, there will be at least a percentage of agents using it. That is, it makes sense from a computational point of view to use the subset heuristic, and not only the heuristics-based best ϕ functions proposed by Hong-Page. With this, the random group no longer outperforms the best group. The explanation is that in this more realistic framework, the best group is not composed of similar agents, contrary to what Hong-Page claimed, but of agents with different heuristics in a general sense. Therefore, "ability trumps diversity" is, again, the correct result.

6.4.3. *Adding a stopping point.* Continuing with the aim of introducing a more realistic framework, we can consider the case where the deliberations stop after a certain number of steps, as we suggested in Section 4. The logic is clear: deliberation is a costly operation, so it cannot be performed indefinitely and it does not seem realistic that deliberations always end in a unanimous agreement, as in the original framework [HP04]. The mathematical implementation of this is straightforward; we substitute the τ defined in Assumption 9 with a new stopping point $\underline{\tau}$

$$\underline{\tau} := \min(\tau, \tau_{\max}) .$$

where $\tau_{\max}$ is the maximum number of steps that can be performed. To illustrate this, we can consider the case of the "easy" parametric V function, where we can set $\tau_{\max} = 12$ steps. In Table 16, we can see the results without stopping point, although the expected ability of the random

[38]The results do not change if we take into account the cost of the evaluation of these cached values, as we have set that the number of points is relatively small compared to n.

group is a little smaller than that of the best group, but the random group outperforms the best group for around 90% of the trials. For a given trial, we can see that the random group is the perfect algorithm, as shown in Figure 12a.

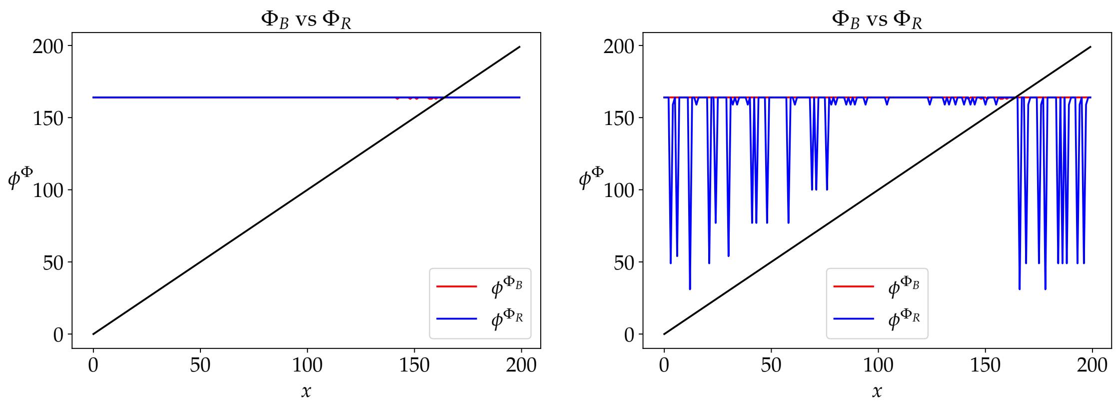
$\hat{A}(\Phi_R)$	$\hat{\tau}^R$	$\hat{A}(\Phi_B)$	$\hat{\tau}^B$	$\hat{O}_{\Phi_B,\Phi_R}$
0.975 (0.100)	20.170 (10.301)	0.994 (0.000)	11.000 (0.000)	0.121 (0.326)

TABLE 16. Results of the expected values and the number of iterations with n=200, l=12, M=1000 number of trials, N_1=10 elements in the random group, V function is parametrically generated with $\alpha, \beta, \gamma, x_0, \delta = 0.328, 1.367, 1.39, 3, 165$. In parenthesis, the standard deviation of the expected values.

In Table 17, we can see the results with stopping point, where now the best group outperforms the random group for most cases and the expected ability of the random group is much smaller than in the previous case. For a given trial, we can see that the best group is almost the perfect algorithm, as shown in Figure 12b, while the random group is not, as it fails to achieve the maximum in a significant number of initial points.

$\hat{A}(\Phi_R)$	$\hat{\tau}^R$	$\hat{A}(\Phi_B)$	$\hat{\tau}^B$	$\hat{O}_{\Phi_B,\Phi_R}$
0.889 (0.145)	11.504 (0.389)	0.994 (0.000)	10.995 (0.000)	0.598 (0.491)

TABLE 17. Results of the expected values and the number of iterations with n=200, l=12, M=1000 number of trials, N_1=10 elements in the random group, V function is parametrically generated with $\alpha, \beta, \gamma, x_0, \delta = 0.328, 1.367, 1.39, 3, 165$, stop set to 12. In parenthesis, the standard deviation of the expected values.



(A) Without maximum number of steps. (B) With maximum number of steps.

FIGURE 12. Comparison of the random and best groups with and without stopping point, $\tau_{\max} = 12$. The global maximum is around $x = 165$.

Note that, although this was done for the "easy" parametric V function, this result is more general. As we can see in the previous tables, the random group tends to take longer to reach the final solution ($\hat{\tau}^R > \hat{\tau}^B$), so limiting the number of steps will penalize the random group to a greater extent than the best group.

6.4.4. *Adding disagreement.* As we have seen in Proposition 2.12, Hong-Page's framework is not realistic, as deliberations always end up in unanimous agreement and, with enough elements, the random group always reaches the correct solution for every V function and for every initial point. This is a significant limitation, as real-world group decision-making frequently involves disagreement between participants. In contrast, our Ability Trumps Diversity (ATD) framework explicitly allows for disagreement through a simple mechanism: when one agent proposes a solution and another agent reverts back to the previous state, this creates a deadlock that ends the deliberation. Even this minimal incorporation of disagreement is sufficient to demonstrate that the "Diversity Trumps Ability" result does not hold in more realistic settings. As we did in Section 4, we can also consider a more realistic framework in the simulations. This is the content of this section. And again, when a more realistic framework is considered, the results change dramatically, showing that ability trumps diversity.

There are several ways to introduce disagreement in the simulations. Following the lines of Section 4, aiming to make a fair comparison between ability and diversity, we are going to consider the case of the best group being composed of agents with ability, so they all have the same function V. In contrast, the random group is composed of diverse agents, but to reflect a minimal loss in ability, we are going to consider that a small percentage of the agents (say 20%) are guided by a function V that is different from the other agents in a small percentage of the points, 0.5%, where it is incorrect. As before, this can introduce a tiny loss in ability, but it is enough to show that ability trumps diversity. When an agent proposes a solution and another agent reverts back to the previous state or the new proposed solution is worse for the next agent, this creates a deadlock that ends the deliberation.

The key point here, as in the previous sections, is not only that introducing changes to the framework changes the results (which in this case is expected), *but that the results change dramatically with small changes to the framework*. That is, the DTA result is not robust to the framework, and one must cherry-pick the framework to make the result hold.

We can see the results in Table 18 and Figure 13. The best group outperforms the random group for most cases, and the expected ability of the random group is smaller than in the previous cases. Again, a minimal loss in ability is enough to show that ability trumps diversity. In particular, the random group has 4 out of 20 elements that assign a higher value to points that have a middle value. This only happens for 0.5% of the points, so it is not a significant loss in ability, but enough to contradict the results in [HP04].

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$
0.933 (0.017)	18.977 (1.846)	0.945 (0.015)	23.394 (1.120)	0.859 (0.348)

TABLE 18. Results of the expected values and the number of iterations with n=200, l=12, M=1000 number of trials, N_1=20 elements in the random group, less ability activated, 4 number of less ability elements, stop when a disagreement cycle is detected. In parenthesis, the standard deviation of the expected values.

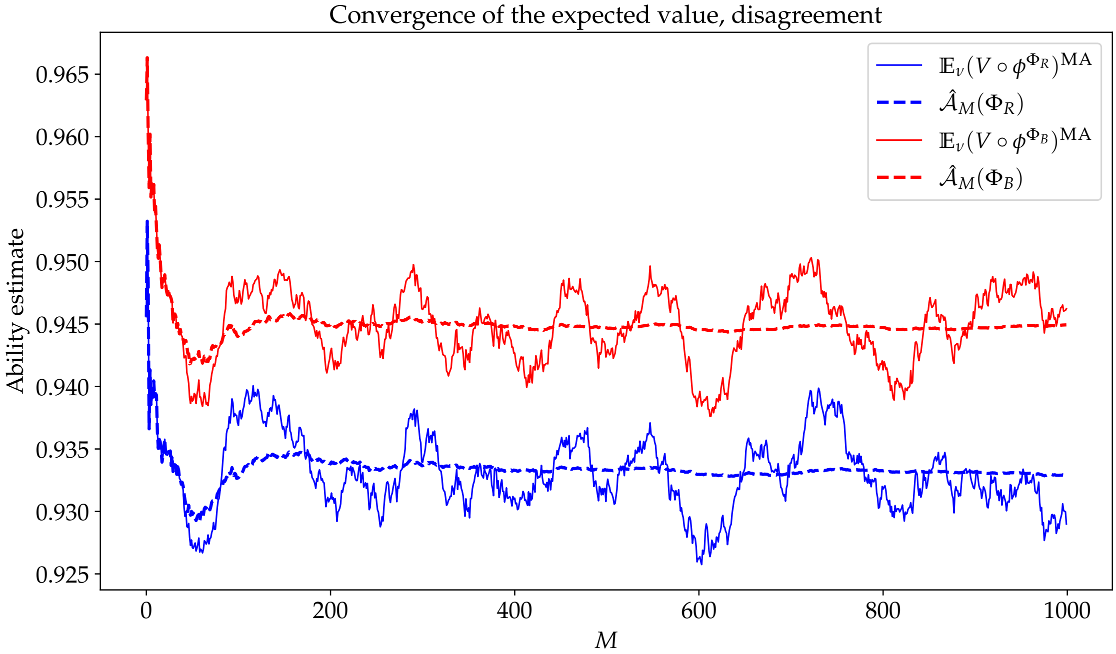


FIGURE 13. Convergence of the expected value of the performance of the group as a function of the number of trials for a disagreement case. The values $\mathbb{E}_\nu(V \circ \phi^{\Phi})$ are computed for $M = 1000$ with a moving average window of 40.

6.5. **Delta-rho framework.** As we have already noted, and as Thompson [Tho14] also observed, Hong-Page's heuristics represent a rather peculiar algorithm for searching for a solution. In Section 6.4.1, we proposed a more realistic heuristic model, and in this section we delve further by considering the "delta-rho" framework—an alternative[39] framework that allows us to incorporate an aspect of realism—namely, differences in ability among agents—that is less evident in the Hong-Page framework.

6.5.1. *Delta-rho Algorithm.* The *delta-rho* ($\delta - \rho$) framework provides an approach to modeling the diversity and ability of agents within a group. Each agent in the delta-rho algorithm is characterized by four parameters, $\rho_-, \delta_-, \rho_+, \delta_+$. These values specify how far to search to the left (ρ_-) and right (ρ_+) of the current point i, along with the corresponding step sizes (δ_- and δ_+) for each side. Concretely, the algorithm first constructs two partial search regions: one for "negative offsets," stepping by δ_- up to ρ_- units left of i, and one for "positive offsets," stepping by δ_+ up to ρ_+ units right of i. Each offset is taken modulo n, so the search wraps around if it goes below 0 or above $n-1$. The current position i itself is then added to this combined search region. Finally, the agent evaluates the value function $V(\cdot)$ at each point in this region and selects the point with the highest value.

Algorithm 6.4: Delta-rho Algorithm

Input: Current point $i \in \{0, 1, \ldots, n-1\}$, value function V
Agent parameters $\alpha = (\rho_-, \delta_-, \rho_+, \delta_+)$

Output: A new point $j \in \{0, 1, \ldots, n-1\}$ maximizing V over the agent's search region.

`// Define left-side search region (if applicable)`
$\mathcal{R}_- \leftarrow \{ j \in \mathbb{Z} \mid j = i - m\,\delta_-,\ m \geq 0,\ j \geq i - \rho_- \}$
`// Define right-side search region (if applicable)`
$\mathcal{R}_+ \leftarrow \{ j \in \mathbb{Z} \mid j = i + m\,\delta_+,\ m \geq 0,\ j \leq i + \rho_+ \}$
`// Combine and reduce modulo` n
$\mathcal{R} \leftarrow \Big(\mathcal{R}_- \cup \mathcal{R}_+\Big) \bmod n$
`// Compute the maximizer in the search region`
$j \leftarrow \arg\max_{x \in \mathcal{R}} V(x)$
return j

6.5.2. *Group of agents.* Consider a group of N agents, each characterized by a set of heuristic parameters denoted as $\phi_i \to [\delta^i_-, \rho^i_-, \delta^i_+, \rho^i_+]$ for $i = 1, 2, \ldots, N$. The parameters ρ_- and ρ_+ represent fundamental abilities of agent i, while δ_- and δ_+ capture auxiliary attributes that modulate these abilities. To regulate the diversity of abilities across the group, the $\delta - \rho$ framework imposes constraints on the ratios of these parameters. Specifically,

$$\frac{\delta^i_-}{\rho^i_-} \leq A \quad \text{and} \quad \frac{\delta^i_+}{\rho^i_+} \leq A,$$

where $A > 0$ is a predefined ratio parameter that controls the extent of variability allowed among the agents' heuristics. The parameter A serves as a threshold to prevent excessive disparity in agent capabilities.

[39]This is to reflect the arbitrariness of the Hong-Page framework via a reductio ad absurdum argument: we can design an alternative framework that introduces an aspect of realism less present in the Hong-Page framework, thereby yielding entirely different results. However, this is by no means a realistic setting; see Section 7.

6.5.3. *Results.* If we want to properly model ability versus diversity, we need to impose different ability parameters for each group of agents; otherwise, both groups will be similarly able[40]. In fact, this is what happens in the original framework [HP04], *where the diverse agents are, basically, as able as the best group*. We can see this in Table 19.

$\widehat{\text{mean}}_{\Phi_B}(\mathcal{A}(\phi))$	$\widehat{\min}_{\Phi_B}(\mathcal{A}(\phi))$	$\widehat{\max}_{\Phi_B}(\mathcal{A}(\phi))$	$\widehat{\text{mean}}_{\Phi_R}(\mathcal{A}(\phi))$	$\widehat{\min}_{\Phi_R}(\mathcal{A}(\phi))$	$\widehat{\max}_{\Phi_R}(\mathcal{A}(\phi))$
0.89 (0.02)	0.88 (0.02)	0.89 (0.02)	0.85 (0.02)	0.83 (0.02)	0.87 (0.02)

TABLE 19. Statistics of the ability difference between the best and random heuristics with n=200, l=12, M=1000 number of trials, N_1=10 elements in the random group. In parenthesis, the standard deviation of the expected values.

The results indicate that the differences in ability between the groups are not significant, as the best agents in the random group perform almost as well as the best agents overall. This is also reflected in the convergence of the estimator of ability, Figure 14a. This suggests that maybe it is ability, rather than diversity, the main reason for the advantage of the random group in the original framework [HP04]. Therefore, the random group's best agents are nearly on par with the best agents, which are more similar by design. So it might be the case that they were not testing diversity versus ability, but two able groups and one of similar agents, which are a clear disadvantage for the best group given the imposed simulation framework.

The delta-rho framework allows us to properly model ability versus diversity. In Figure 14b, cf. Tables 21 and 20, we can see that the individual ability of the best group is larger than the random group, now differences of 10 percentage points in mean. Furthermore, the random groups are more diverse in the sense of variability of each component (a metric used by Hong and Page, (5.1)). With this, (diversity of random group and ability difference increased), diversity no longer trumps ability. We are faced, again, with the same problem as in the previous sections: the "Diversity Trumps Ability" result is not robust and small changes in the framework can make the result quickly break. One must cherry-pick the framework to make the result hold.

Note also that in Table 21, the best group is not composed of similar agents, in the sense that N_B is closer to N_1 than before. Thus, *the repeated claim of Hong-Page that the best group is composed of similar agents was just an artifact of their framework*, and not a property of the real world or of other frameworks.

$\widehat{\text{mean}}_{\Phi_B}(\mathcal{A}(\phi))$	$\widehat{\min}_{\Phi_B}(\mathcal{A}(\phi))$	$\widehat{\max}_{\Phi_B}(\mathcal{A}(\phi))$	$\widehat{\text{mean}}_{\Phi_R}(\mathcal{A}(\phi))$	$\widehat{\min}_{\Phi_R}(\mathcal{A}(\phi))$	$\widehat{\max}_{\Phi_R}(\mathcal{A}(\phi))$
0.93 (0.01)	0.92 (0.01)	0.93 (0.01)	0.83 (0.02)	0.77 (0.03)	0.87 (0.02)

TABLE 20. Statistics of the ability difference between the best and random heuristics with n=200, l=12, M=1000 number of trials, N_1=10 elements in the random group, delta-rho ratios A_R=3 and A_B=5. In parenthesis, the standard deviation of the expected values.

The situation is even clearer when we consider the parametric V function. The results presented in Table 22 and Figure 15 illustrate the performance of the two groups, Φ_R and Φ_B, under the delta-rho framework with parametric V functions.

[40]This is equivalent to assuming that the pool of less able agents is large enough (perhaps through repetition) to make the random selection of top-performing agents highly unlikely.

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$	N_B
0.934 (0.016)	12.366 (0.506)	0.978 (0.010)	13.238 (0.935)	1.000 (0.000)	9.234 (0.790)

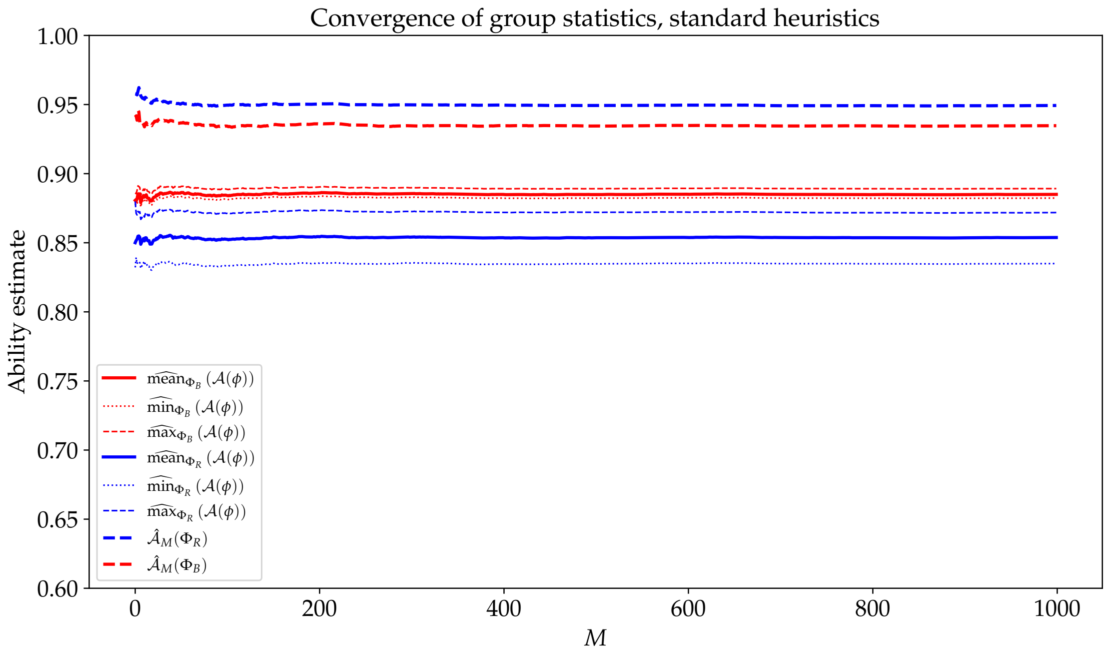
TABLE 21. Results of the expected values and the number of iterations with n=200, l=12, M=1000 number of trials, N_1=10 elements in the random group, delta-rho ratios A_R=3 and A_B=5. In parenthesis, the standard deviation of the expected values.

Table 22 summarizes the expected values and the number of iterations for both groups, highlighting the differences in their performance metrics. The estimates of ability $\hat{\mathcal{A}}(\Phi_R)$ and $\hat{\mathcal{A}}(\Phi_B)$ indicate that the random group Φ_R has a much lower ability estimate (0.66) compared to the best group Φ_B (0.92). Recall that the original Hong-Page framework claimed that "Diversity Trumps Ability" because there was a difference, approximately 0.02, between the ability of the best group and the random group. Now, the difference is much larger, more than ten times with respect to the original. So, following their logic, we can say that "ability absolutely dominates diversity" when we consider hard problems (not the easy and less interesting ones of [HP04]). In fact, the results show that ability not only trumps but completely overwhelms diversity in these more realistic scenarios, where the problem is hard. The outperformance probability $\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$ is also noteworthy, with a value of 0.94, suggesting that the best group outperforms the random group in a majority of trials.

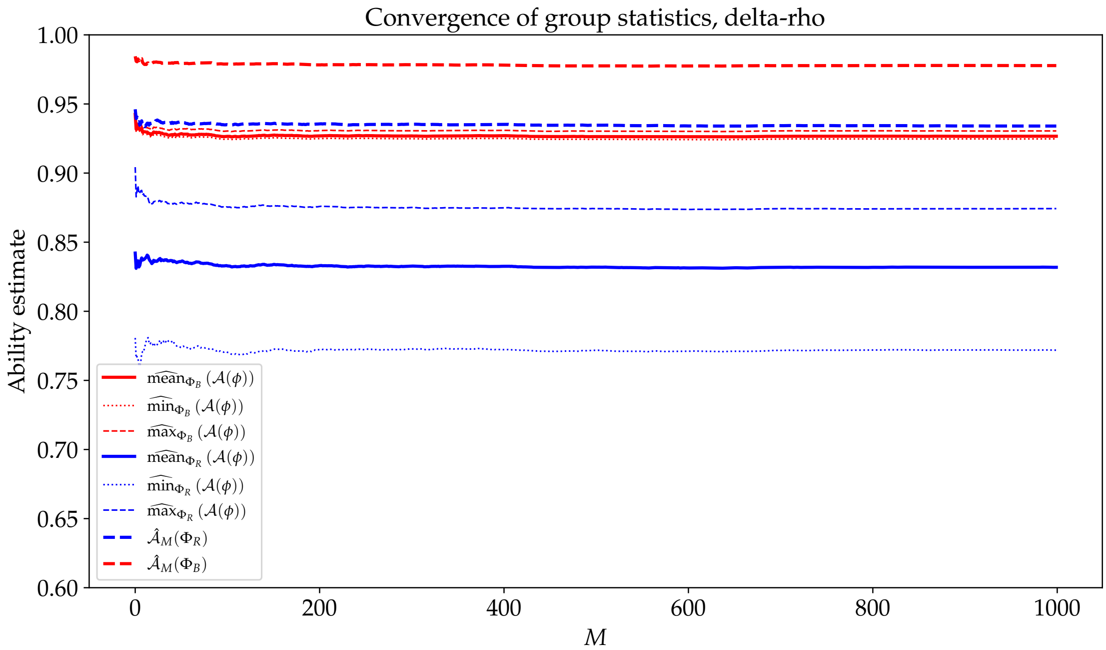
Figure 15 further elucidates the convergence of group statistics over the number of trials M. The ability estimates for both groups stabilize as M increases, with the mean ability estimate for the best group $\hat{\mathcal{A}}(\Phi_B)$ consistently hovering around 0.92, while the random group's mean ability estimate $\hat{\mathcal{A}}(\Phi_R)$ stabilizes around 0.66. The dotted lines representing the minimum and maximum individual ability estimates for both groups illustrate the variability within each group, with the random group showing a wider range of ability estimates. With our framework, the random group is more diverse in the sense of variability of each component (a metric also used by Hong-Page, (5.1)), and the ability of the random group is larger than the ability of the best group. Thus, again, *we are properly modeling ability versus diversity and not ability and diversity versus ability.*

$\hat{\mathcal{A}}(\Phi_R)$	$\hat{\tau}^R$	$\hat{\mathcal{A}}(\Phi_B)$	$\hat{\tau}^B$	$\hat{\mathcal{O}}_{\Phi_B,\Phi_R}$	N_B
0.66 (0.17)	17.12 (7.43)	0.92 (0.10)	15.75 (3.19)	0.94 (0.24)	6.96 (1.29)

TABLE 22. Results of the expected values and the number of iterations with n=200, l=12, M=1000 number of trials, N_1=10 elements in the random group, V functions are parametrically generated, stop when a disagreement cycle is detected, delta-rho ratios A_R=3 and A_B=5. In parenthesis, the standard deviation of the expected values.



(A) Standard Hong-Page heuristics.



(B) Delta-rho case with ability differences.

FIGURE 14. Convergence of expected value and group statistics for the standard heuristics and the delta-rho case with ability differences.

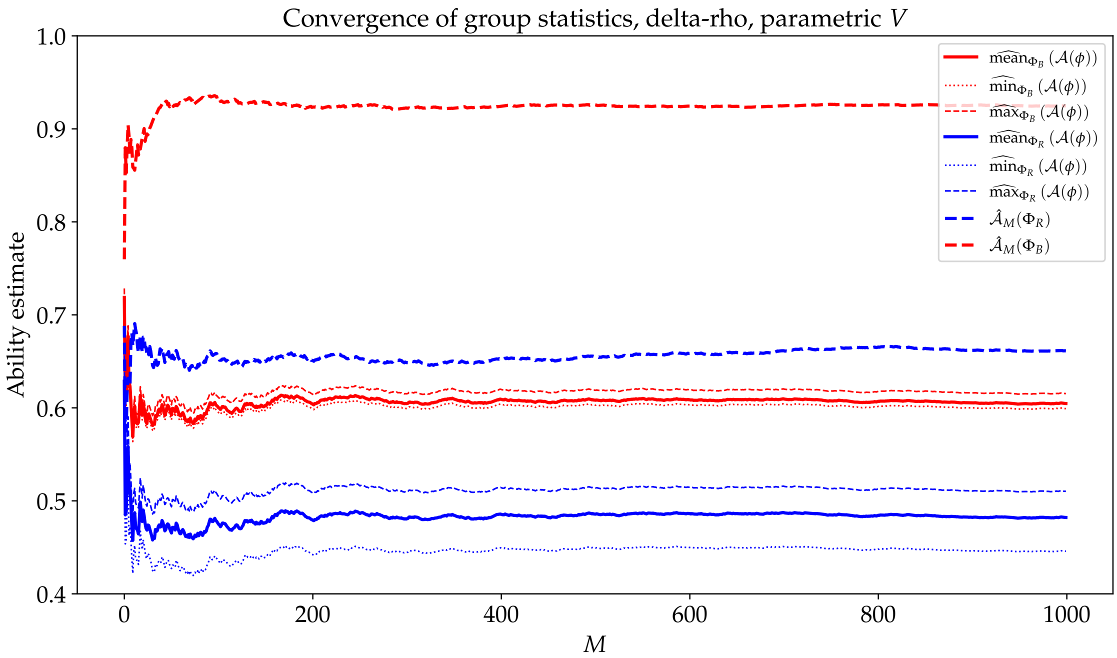


FIGURE 15. Convergence of group statistics for the delta-rho case with ability differences and parametric V function.

7. Fundamental problems with the Hong-Page framework

After delving into our new theorem and the analysis of the simulations, let us examine the fundamental limitations of the Hong-Page framework that, while not directly addressed in our theorem or our modification of the simulation framework, merit careful consideration. While our primary aim in the theorem and our simulation framework was to elucidate certain, but fatal, weaknesses in their framework and motto "diversity trumps ability", we acknowledge that fully resolving these particular issues would necessitate an almost complete reconstruction of the theoretical foundation—an endeavor that would require essentially rebuilding the framework from scratch, which extends beyond the scope of this paper.

This highlights a deeper issue with the Hong-Page framework: it is not merely a simplified model with imperfect assumptions (as all models are), but rather a framework that fundamentally fails to capture the essential relationships among diversity, ability, and effective group problem-solving. That is, it would be analogous to analyzing the motion of objects with a model that assumes there is no gravity, not just a simplified model like assuming there is no friction. The conclusions derived from a no-gravity model would be of little relevance to the real world, but those from a no-friction model would be a good starting point. Hong-Page's framework is more like a no-gravity model. The framework's rigid structure and unrealistic assumptions about how agents interact with problems make it unsuitable for understanding how to optimally compose groups to tackle real-world challenges. The framework misses crucial aspects of human expertise and group dynamics that are vital for actual decision-making and problem-solving. Furthermore, it leads to unrealistic conclusions that are not supported by empirical evidence or a minimal amount of realism. All in all, this model seems to be a good example of a model that is not effective and should be avoided.

7.1. **Incorrect focus on the starting point.** A key issue with the Hong-Page framework is that the ability of agents is measured against the starting point, x_0, and not the problem or "landscape" as they often say, V. Let us mathematically develop this point. For Hong-Page, the agents are fixed to a problem V and evaluated on their performance across different starting points. That is

$$\phi : X \to X\,.$$

Although some algorithms need a starting point, for instance some numerical minimization methods, applying this analogy to real human problem solvers is dubious. And, as we will see below, even these algorithms are not evaluated on their performance across different starting points.

Remark 7.1. In [HP98; HP01], and also mentioned in [HP04], the authors develop a framework of perspectives, $M : X' \subset X \to \Gamma$, where Γ is an internal language (modeled, e.g., by strings of bits in [HP01]), and $H = \{h_1, \ldots, h_m\}$ is the set of heuristics ($h : \Gamma \to \Gamma$). Consequently, ϕ can be seen as:

$$\phi(x) = x^* \text{ such that } V(x^*) = \max_{h \in H} V\left(M^{-1}\left(h(M(x))\right)\right)$$

where M^{-1} should be well-defined by construction. These heuristics are not the same as the ones used in the simulations, but the latter are a particular case of the former ($M =$ identity and H consisting of a one-element set). Furthermore, it is clear from the above equation that M is avoidable. One could instead define heuristics as $\tilde{h} : X \to X$ such that $\tilde{h}(x) = M^{-1}(h(M(x)))$. This approach encapsulates the perspective M within the heuristics. Although they attempt to defend this perspective in, for instance, the introduction of [HP98], whether it is useful or constitutes unnecessary complexity depends on the context. Defining the internal language and perspectives without using them adds unnecessary complexity, this was the case of [HP04]. We will see below that if M was not the identity, but depended on the agent, the slogan "diversity trumps ability" would be less likely to hold. When it is used elsewhere, its purpose is not clear. For instance, in [HP01, Section 3], M points to strings of bits where, supposedly, the agents are representing the possible solutions. Agents represent solutions as different strings of bits. A basic heuristic for all agents is changing the i-th bit of the string from 0 to 1 or from 1 to 0, and agents are characterized by which subset of positions can change. Thus, moving to an unrealistic internal language like this, when one could simply remain in X, again seems to add unnecessary complexity to an otherwise simple (and, in any case, incorrect) framework. Further discussion on this topic is provided later. ⋄

In reality, agents are evaluated on their performance across different problem landscapes, which will constitute expertise. That is, an agent should be a function as

$$\begin{aligned} F : \mathcal{V} &\to X \\ V &\mapsto F(V)\,, \end{aligned}$$

where $\mathcal{V}$ is the set of possible problem landscapes of our interest. Let us denote by x^{V} the unique solution to the problem V, i.e., $V(x^{\mathrm{V}}) = 1$. It is important to note that, in general,

$$V : \mathcal{D}^{\mathrm{V}} \to [0,1]\,,$$

where $\mathcal{D}^{\mathrm{V}}$ is the domain of the problem V.

Returning to the analogy of the numerical minimization methods, although they depend on the starting point, they can be seen as a function F that maps the problem V to the solution x^{V}. That is, either one usually has an additional function $\mathcal{X}_0 : \mathcal{V} \to X$ that maps the problem V to a reasonable starting point x_0, for instance, obtained through some analytical approximation, one considers a grid of points in the domain and the solution is the one that minimizes the function in the grid, or they do not need a starting point, as, for instance, simulated annealing where this starting point is an optional parameter. In any case, we form the function F as

$$F : \mathcal{V} \to X$$

where $F(V) = \phi(\mathcal{X}_0(V))$ or $F(V) = \arg\max_{x\in X_V} V(\phi(x))$ being X_V the previous grid, or, in general, a proper subset of the domain. This implies that ν is an atomic measure with a single atom, not the uniform distribution. Thus, one would hardly measure the performance of an algorithm as Hong-Page does, (2.1), but, for instance, as an expectation across different problems,

$$\mathbb{E}_{\mu^{\mathcal{V}}}\left[V(F(V))\right]\,,$$

where $\mu^{\mathcal{V}}$ is a probability measure on the set of problems $\mathcal{V}$, where a suitable σ-algebra should be defined.

7.1.1. *An example.* Consider the problem of determining economic measures to deal with climate change[41], V. The domain of this problem is the set of possible policies that can be implemented, $\mathcal{D}$. Let us consider a group of top economists in that field or related fields, Φ_B, and a random group of, say, a random assembly of university graduates (which by assumption are able in the Hong-Page sense) from a diverse set of fields, Φ_R. Would it make sense to model a particular economist as a function of the possible solutions to a given problem? What does it mean to "start from the particular point" x? What kind of expert would give a different solution if he/she first thinks x instead

[41] These examples are typical of those proposed in the literature. For instance, Hélène Landemore (a collaborator of Page on these issues) advocates for "open democracy," a system where legislative power becomes more directly accessible to citizens through randomly selected assemblies, based on Hong-Page's work (see [Rom23; Rom25] and references therein). These bodies, such as France's climate conventions, involve citizens (with no exclusion from the pool) in extended deliberations on complex policy issues, aiming to produce recommendations that inform government decision-making. Even setting aside the fact that the theorem cannot be applied to problems with value disagreements (see Remark 2.3), her application contains important dubious logical steps (see [Rom23; Rom25]).

of $x' \neq x$? That is, if they are primed to first think about increasing carbon taxes, they will reach the conclusion that the solution is ESG criteria, ϕ_B(increasing carbon taxes) $=$ ESG criteria but, if they are primed to first think about nuclear plus renewables, they will reach the conclusion that the solution is a small increase in carbon taxes, ϕ_B(nuclear + renewables) $=$ small increase in carbon taxes. The assumed level of irrationality, they can be manipulated to give different solutions based on a priming effect, contrasts with the fact that they will always be able to find the solution, say a set of reforms

$$x^{\mathrm{V}} = \{r_i^* \text{ in the } i\text{-th sector}\}_{i=1}^m$$

and order the solutions from better to worse. Following their numbers, agents can evaluate and rank exactly 2000 solutions to a problem and, consequently, can determine the superior option among nearly 2 million pairs of solutions, with all agents agreeing on this ordering. That is, ϕ_B can recognize and order by increasing value different $(r_1, \dots, r_m)$, where reforms in each sector have different costs and benefits and interaction effects, but it is not able to avoid the priming effect even when they are made to think about it, which contradicts the empirical evidence [SWT16, p.152].

Indeed, in the Hong-Page framework, thinking 100 times about the problem V will always lead to the same solution as thinking once, Assumption 5, $\phi_B \circ \phi_B = \phi_B$ even though the expert is aware of the framing effects and knows that it should be independent of the starting point. Also, consider two points x_1, x_2. If an expert first thinks about x_1, the solution is $\phi_B(x_1) = x_1^*$. If the expert first thinks about x_2, the solution is $\phi_B(x_2) = x_2^* = x^{\mathrm{V}}$, the solution to the problem V. Then, if the expert were asked to start from x_1, then x_2, and again from x_1, the solution is x_1^* although he already knows that the solution is $x_2^* = x^{\mathrm{V}}$.

So the agents, including experts, in the Hong-Page framework are subject to framing effects even when given infinite time to deliberate, yet they can precisely determine whether SMR reactors plus renewables with soft ESG criteria under the ECB's mandate are superior or inferior to fourth-generation reactors plus renewables with hard ESG criteria under the ECB's mandate—a contradiction that constitutes a wildly incongruous model with an internally dissonant structure.

7.2. **Group dynamics that always find the solution.** One might think that going from ϕ to F^ϕ is a simple matter of, for instance, given a problem V and an agent ϕ, $F^\phi(V)$ is the best solution that F^ϕ can find or that experts are rational enough to not be subject to framing effects in their area of expertise, i.e., $\phi_B(x)$ is the same for all $x \in X$. However, this is absurd in the Hong-Page framework, as this implies from their assumptions that $F^\phi(V) = x^{\mathrm{V}}$ for all $V \in \mathcal{V}$; recall that $\arg\max_{x \in X} V(x) = x^{\mathrm{V}} = \phi(x^{\mathrm{V}})$ for all $\phi \in \Phi$. That is, all possible agents always find the solution to the problem V. Also, it violates their own assumptions, see Assumption 6.

Actually, this reflects a deeper issue with the Hong-Page framework: we can design the following group dynamics that always finds the solution to the problem V. This is a variation of the tournament dynamics, see Footnote 27, used in the literature in which, in general, "diverse" groups perform better, see [Gri+19, Section 5]. Recall that, in Assumption 9, the group dynamics remained somehow open, so let us consider the following dynamics in a more general setting:

(1) Let us divide the initial points[42] into agents, at least $\lfloor |X|/N_1 \rfloor$ for each agent until all the initial points are covered. So each agent $\phi \in \Phi$ has assigned a set of initial points $X_\phi \subset X$. With large groups, as claimed in the Hong-Page framework, the number of points will be

[42] Or the union of their local optima. This can be done as "the local optima of every problem solver can be written down in a list" [Pag07, Chapter 6]. See Section 5.3 of [Rom23]. Alternatively, divide the groups so that they start at different points, collect the points, and repeat until all points are covered. In some problems, such as a jury or a company solving a problem with a limited amount of options, the number of possible solutions is not large and known. In other problems, like the set of possible reforms, one could categorize the solutions and assign them to different agents.

small. That is,
$$X = \bigsqcup_{\phi \in \Phi} X_\phi \,, \tag{7.1}$$
where $\sqcup$ denotes the disjoint union.

(2) By hypothesis, (7.1), there is a unique solution x^{V} to the problem V and this will belong to some set X_ϕ for some agent $\phi \in \Phi$.
(3) We choose "tournament" dynamics, the next best solution is the best one that improves the current solution. The difference with respect to the standard tournament dynamics is that they do not start at the same point.
(4) Each agent returns the best solution in the set X_ϕ[43].

With this dynamics, the group will always find the solution to the problem V in just one iteration with unanimous agreement. Note that this works even for one agent, just take $|N_1| = 1$ and give it, in general and on average, 10 or 20 times the time given to a group (of 10 or 20 agents) to find the solution. This shows the incapacity of the Hong-Page framework to capture realistic features of group problem-solving as any other conclusion is polluted by these absurd features.

7.2.1. *Continuation of the example.* Say we have the group of university graduates, Φ_R and all the possible reforms, $\{(r_1, r_2, \ldots, r_m)\}$, with solution $(r_1^*, r_2^*, \ldots, r_m^*)$. In the Hong-Page framework, take a group of 20 agents and divide the possible reforms (say 100) into 20 groups of 5. Then, the group will always find the solution to the problem V in just one iteration with unanimous agreement. The solution $(r_1^*, r_2^*, \ldots, r_m^*)$ will be given to an element of Φ_R, and this agent will recognize that this is the solution to the problem V. He/she will communicate this solution to the rest of the group and the group will agree and declare that this is the solution to the problem V. That easy.

7.3. **The domain $\mathcal{D}^{\mathrm{V}}$.** Another major limitation of the Hong-Page framework is that the domain, in the simulations, is constrained to be the set $\{1, \ldots, n\}$ or similar structures. This is problematic for several reasons:

First, with the heuristics algorithm used in the literature (see Algorithm 1), it means that problem solvers are not invariant under domain shifts or reorderings. For instance, consider $\sigma : X = \{1, \ldots, n\} \to \{1, \ldots, n\}$ a permutation of the possible solutions. Then, define
$$V^\sigma := V \circ \sigma$$
as the problem V with the solutions reordered according to σ. This is a new function, but it is the same problem and the solution remains the same, $x^{\mathrm{V}^\sigma} = \sigma^{-1}(x^{\mathrm{V}})$. However, the solution found by the agents is not invariant under this transformation, i.e., in general,
$$\sigma \circ \phi_\alpha^{V^\sigma} \circ \sigma^{-1} \neq \phi_\alpha^V \,, \tag{7.2}$$
when ϕ_α^V follows Algorithm 1 according to the heuristics α and function V. That is, if we were to change the order in which the solutions are presented, the final solution of the agents would change, even though the underlying problem remains identical. Note that in their heuristics algorithm, it is not possible to always change the heuristics to produce the same solution independent of the order of the solutions, i.e., (7.2) does not necessarily hold. For instance, let $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ be the vector of heuristics of an agent and x_1 be the second best option and such that $\phi_\alpha(x_1) = x^{\mathrm{V}}$. However, once reordered, these points are separated by more than l positions, so there are no α' such that
$$\phi_{\alpha'}^{V^\sigma}\left(\sigma^{-1}(x_1)\right) = \sigma^{-1}(x^{\mathrm{V}}) \,.$$

[43]It is indifferent if $\arg\max_{x \in X_\phi} V(x)$ or $\arg\max_{x \in X_\phi} V(\phi(x))$, although being the latter more computationally demanding. But the agents, by hypothesis, are able to perform both computations

This point will be discussed in more detail in the next subsection.

Second, the interactions across the domain are artificial. First, agents always recognize the solution, i.e., $\phi(x^V) = x^V$, which is unrealistic and the source of the also unrealistic corollary that we can always find the solution to any problem with enough agents, as in the theorem, cf. Proposition 2.12, and similarly for the simulations. Furthermore, the movements across the domain are unrealistic, specially in the random landscapes used in the simulations. For instance, let us assume that x_1^V, the best solution, and x_2^V, the second best solutions are far apart (more than l), but, going from the second best solution to the best one is just a trivial step. For instance, realizing that if we remove an element from the second best solution, we get a win-win situation, or Pareto improvement. On the contrary, let us assume that the worst solutions, $x_{|X|}^V, x_{|X|-1}^V, \ldots, x_{|X|-l+k}^V$, are just before the best solution. Going from these worst solutions to the best one is a far from trivial step. Say, for instance, going from anti-economic measures to a perfectly balanced mix of economic and environmental measures based on the best available empirical evidence. But, in the Hong-Page framework, this is a trivial step; for all agents ϕ in the simulations there is a "terrible" economic measure from where they will arrive in one step to the best solution. On the contrary, no one, even the best agents, will realize that they can go from the second best solution to the best one, although, as we have said, this is a trivial economic argument based on a Pareto improvement.

One might be tempted to think that this is a matter of ordering the solutions increasingly. However, this is even more absurd as we will see in the next subsection. It would imply, again, that groups are almost always able to solve any problem. In particular, best agents always solve the problem. But this is a contradiction to the Hong-Page framework, as it violates one of the assumptions, see Assumption 6.

This combination - the arbitrary assumption of interactions across the domain, lack of invariance properties we would expect from rational agents, and the tension between assumed agent capabilities - makes the framework dubious as a model of real group problem-solving. While mathematical abstractions and simplifications are necessary in any model, these particular modeling choices seem to distort the key phenomena we are trying to understand rather than illuminate the essential features of the problem.

7.3.1. *Continuation of the example.* Agents are, again, subject to framing effects. If we change the order of the solutions, the solution found by the agents will change, even though the underlying problem structure remains identical, and the agents are able to always find the solution to any hard problem, as we saw in the previous subsection. We have a similar problem as above; this violates basic principles of rational agency - a truly rational problem solver should care, after enough reflection, about the consequences of the solutions, not their arbitrary labels. This creates a tension between two aspects of the framework: on one hand, the agents are modeled as fairly unsophisticated with limited rationality, see also the discussion above. But on the other hand, as shown in the previous subsection, these same agents are guaranteed to eventually always find the global optimum, displaying a level of performance that would seem to require much more sophisticated reasoning capabilities. Thus, agents have an inherent irrationality that cannot be corrected, but, at the same time, they are able to always find the solution to any hard problem. Again, this illustrates the erratic and artificial nature of the Hong-Page framework.

Similarly, let us assume that the second-best solution is given by the set of reforms

$$\{r_1^*, r_2^*, \ldots, r_m^*\} \cup R_0 \,,$$

where R_0 denotes a regulation that, for instance, prohibits a technology reducing emissions while channeling the associated revenues exclusively to the poorest 10% of the population while improving the welfare of everyone else. This regulation leaves the other reforms unchanged, and it is evident that an improvement would be achieved by simply removing R_0. However, within the Hong-Page framework, no agent—not even the most able—can discern that the solution can be improved by eliminating this absurd regulation.

On the other hand, suppose that a candidate "solution" is to subsidize high-emitting firms producing low-value commodities that are owned by affluent individuals connected to the government, which we denote by R_1. This option is clearly far removed from the optimal solution, which is the set of reforms

$$\{r_1^*, r_2^*, \ldots, r_m^*\},$$

derived from a critical reassessment of the best empirical evidence and economic models. In the Hong-Page framework, agents starting from this overtly deficient solution R_1 will quickly identify the optimal set. Paradoxically, if agents begin with the near-optimal solution $\{r_1^*, r_2^*, \ldots, r_m^*\}$ tainted by the regulation R_0, they fail to recognize that removing R_0 leads to a clear improvement although is trivial. In other words, the framework implies that, under the conditions mentioned above, the trivial task

$$R_0 \cup \{r_1^*, r_2^*, \ldots, r_m^*\} \to \{r_1^*, r_2^*, \ldots, r_m^*\}, \tag{7.3}$$

is unattainable for any agent, whereas the highly non-trivial task

$$R_0 \to \{r_1^*, r_2^*, \ldots, r_m^*\}, \tag{7.4}$$

is trivially achieved by all agents. This discrepancy underscores a fundamental shortcoming of the framework: agents are paradoxically effective at escaping from a blatantly flawed solution, yet they remain incapable of effecting a straightforward improvement when starting from a nearly optimal configuration. This behavior is completely unrealistic and shows the erratic and artificial nature of the Hong-Page framework.

7.4. **The shape of V.** In the simulations literature, the "shape" of the problem has been given a great deal of attention, see [Gri+19], as the conclusion is sensitive to the shape of the problem. But the shape of the problem is just an arbitrary feature that can be changed without changing the problem. As we have already seen, considering a domain $X = \{1, \ldots, n\}$ of the problem V, the same problem can be reformulated as a problem V^σ with the solutions reordered. Obviously, the shape of the "landscape" is different, but the problem is the same. In extreme cases, we can consider a function V of slow-variation, what is called a "smooth" landscape, or a landscape with a large number of "random" oscillations, which is called a "rough" landscape just by reordering the solutions. See Proposition 7.2 for more details on how we can transform the "hardest" problem into the easiest one just by reordering the solutions.

For instance, fixing a problem V, let $\sigma^{\mathrm{I}} : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a permutation of the solutions such that $V(\sigma^{\mathrm{I}}(i)) < V(\sigma^{\mathrm{I}}(i+1))$ for all $i \in \{1, \ldots, n-1\}$, see Assumption 3. For this landscape, any agent that has the number 1 as one of its α_i will always find the solution to the problem V as the maximum. In particular, any group with at least one agent with $\alpha_1 = 1$ will always find the solution to the problem V as the maximum. So its ability $\mathcal{A} = 1$. That is, with a proper ordering of the solutions (which could be done by the agents), any agent can almost always find the solution to any hard problem.

Now, suppose we shuffle the solutions, then the landscape is different, but the problem is the same. Let $\sigma^{\mathrm{R}} : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a permutation of the solutions such that $V^{\sigma^{\mathrm{R}}}$ is a "rough" landscape, see [Gri+19]. For this landscape, and maintaining the original Hong-Page assumptions, the random group will tend to perform better than the group of experts. But consider now a permutation σ^{S} such that $V^{\sigma^{\mathrm{S}}}$ is a "smooth" landscape, see [Gri+19], or that follows a similar pattern as in Figures 9a, 9b. In this case, the random group will tend to perform worse than the group of experts. In all the cases we have the same problem, but orderings of the solutions which imply radically different conclusions on whether ability trumps diversity (for instance, Figure 2 in [Gri+19]), but *exactly the same problem*. This reflects the lack of realism of the Hong-Page framework.

In fact, we can prove the following result which essentially shows that for two different problems, one can reorder the solutions to make the problems equivalent. In particular, just reordering the solutions, we can transform the most "difficult" problem into the easiest one. This is a clear drawback of the Hong-Page framework, as it is completely unrealistic.

Proposition 7.2. *Let V and V' be two (completely) different problems satisfying Assumption 3. Then there exists a permutation σ of the solutions such that, for any group of problem solvers Φ following Algorithm 1,*

$$\phi^{\Phi}_{V\circ\sigma} = \phi^{\Phi}_{V'}\,,$$

where ϕ^{Φ}_{V} is the result of the group of problem solvers Φ following Algorithm 1 applied to the problem V.

Proof. Define

$$\sigma^{\mathrm{V}}(i) := x_i^{\mathrm{V}}\,,\ \sigma := \left(\sigma^{\mathrm{V}}\right)\circ\left(\sigma^{\mathrm{V}'}\right)^{-1}.$$

Then $V\circ\sigma$ satisfies $V\circ\sigma\left(x_i^{\mathrm{V}'}\right) = V\left(x_i^{\mathrm{V}}\right)$ for all $i\in\{1,\dots,n\}$. Indeed,

$$V\circ\sigma\left(x_i^{\mathrm{V}'}\right) = V\left(\sigma^{\mathrm{V}}\left(i\right)\right) = V\left(x_i^{\mathrm{V}}\right).$$

Thus $x_i^{\mathrm{V}'} = x_i^{\mathrm{V}^{\sigma}}$ for all $i\in\{1,\dots,n\}$. Thus, the relative order of the solutions is the same in V' and $V\circ\sigma$, i.e., if, for instance, $x_1^{\mathrm{V}'} = 10$, then $x_1^{\mathrm{V}^{\sigma}} = 10$, etc. Because the algorithm's outputs depend only on the relative values of the solutions, the same group of solvers Φ will produce the same output. Indeed, let $x,y\in X$ so there exist i,j such that $x = x_i^{\mathrm{V}'}, y = x_j^{\mathrm{V}'}$, thus

$$V'(x) > V'(y) \Leftrightarrow i > j \Leftrightarrow V^{\sigma}(x_i^{\mathrm{V}}) > V^{\sigma}(x_j^{\mathrm{V}}) \Leftrightarrow V^{\sigma}(x) > V^{\sigma}(y)\,.$$

Therefore,

$$\phi^{\Phi}_{V\circ\sigma} = \phi^{\Phi}_{V'}\,.$$

□

Remark 7.3. This result applies to any algorithm that only depends on the relative values of the solutions. ◇

We could choose the shape to obtain some desired properties, but which ones? In other words, there would be an unobservable preferential order[44]. To the best of our knowledge, no analysis exists to determine which shape is the most realistic. Some partial results will be commented on below. For instance, the shapes described in (6.13) are by no means comprehensive, but they offer something less present in the literature—"difficult" problems where the group scores are much lower than in the original case. In any case, this already refutes the classification of problems as "smooth" or "rough", which translates to predictability or not, [LS24, Table 3], see below, is not complete. It is not just about rough or smooth landscapes, but landscapes with varying degrees and kinds of "roughness". These are more realistic than consistently high scores averaging over 90%. Additionally, the intuition behind the shape of the problem was to represent different heuristics of the agents, as discussed in [Pag07, Pages 139-143]. However, these heuristics should be defined for each agent in their internal language (see Remark 7.1). The consequences of this choice again depend on the modeling decisions made. Experts can be defined with more increasing internal landscapes, which would avoid the framing effects presented above—a desirable property.

[44] There can be landscapes with a clear preferential order. For instance, see Figure 9 in [GS20].

However, in that case, the slogan "Diversity Trumps Ability" is less likely to hold, see below for more details.

Also, there is a related problem already pointed out in the literature, see [Gri+19], ability (or expertise) should be translatable to different problems. That is, if an agent is able to solve a problem V, it should be able to solve a similar problem V', measured by some distance defined in $\mathcal{V} \times \mathcal{V}$. However, this is not the case in the original Hong-Page simulations. As pointed out[45] by [Gri+19], when experts are translated to different problems, the motto "Diversity Trumps Ability" does not necessarily hold as the group of experts tends to outperform the random group. That is, the best agents for a problem to be considered experts should perform well on a similar problem. But this only happens for "smooth" landscapes, not the ones originally considered by Hong and Page. And, as we have said, in that setting the claim "Diversity Trumps Ability" does not necessarily hold.

Surprisingly, some authors have regarded this "bug" in the model as a feature (see [LS24]), using it to state that experts do not exist for certain problems. They conclude:

> Therefore, Plato's Cave needs to be revisited as follows: in a pitchblack 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 questions), and problem-solving tends to be epistemically superior among diverse participants. Thus, in darkness, diversity outperforms ability.

The idea here is that, because the Hong-Page framework cannot model experts when the shape of V is not "smooth," it follows (according to these authors) that experts do not exist for such problems. However, this confuses a limitation of the model with a property of the real world. In other words, it treats a "bug" in the model as if it were an actual feature of reality. Is this property present in all models? Could it reflect a limitation of the model or of the agents we are using in the modeling framework? Could it be reflecting that modeling best-performing agents as agents with limited rationality, subject to priming effects, is not a good modeling choice? Or does it reflect a limitation in how we measure expertise?

As a toy example to answer this questions, consider the following distance[46]:

$$d(V, V') := \sum_{i=1}^{d} \omega_i \, d^T\left(x_i^{\mathrm{V}}, x_i^{\mathrm{V}'}\right) \; + \; \sum_{i=1}^{d} \omega_i \left| V\left(x_i^{\mathrm{V}}\right) - V\left(x_i^{\mathrm{V}'}\right) \right|,$$

where each $\omega_i \geq 0$ (non-increasing), and x_i^{V} is the i-th best solution to the problem V. Here, d^T is the trivial metric:

$$d^T(a, b) \;=\; \begin{cases} 1, & a \neq b, \\ 0, & a = b, \end{cases}$$

so d is indeed a distance. Under this metric, and by using "subset heuristics" (Algorithm 6.2), agents performing well in one problem will also perform well in a related problem, regardless of the problem's shape. Thus, *there are experts for all problems*.

Indeed, for an appropriately chosen set of weights $\{\omega_i\}$ and a sufficiently small ε such that $d(V, V') < \varepsilon$, if a particular agent ϕ_{X_0} performs well in V, then the set X_0 may contain points of the form $\{x_i^{\mathrm{V}}\}_{i=1}^{d}$ for sufficiently small d, with $V(x_d^{\mathrm{V}})$ close enough to $V(x_1^{\mathrm{V}}) = 1$. If ε is small or ω_i is

[45]There is a limitation on how the measure "related problems" as it only applies in the case of their "smoothing factor". Below we will consider the more general metric approach outlined here.

[46]Note that this metric is a invariant under permutations of the solutions.

large, one can show that

$$d^T\left(x_i^{\mathrm{V}}, x_i^{\mathrm{V}'}\right) < \frac{\varepsilon}{\omega_i} \quad \text{and} \quad \left|V\left(x_i^{\mathrm{V}}\right) - V'\left(x_i^{\mathrm{V}'}\right)\right| < \frac{\varepsilon}{\omega_i} \quad \Longrightarrow \quad x_i^{\mathrm{V}} = x_i^{\mathrm{V}'},$$

for $i \leq d$ with sufficiently close values in V. Consequently, the performance of ϕ_{X_0} in V' will exceed $V(x_d^{\mathrm{V}}) - \frac{\varepsilon}{\omega_d}$, which is close to 1. Hence, the agent will perform well in V', qualifying it as an expert by definition.

Another example, still within the Hong-Page heuristics framework, involves allowing different internal representations of the problem (see Remark 7.1). Suppose experts use a function

$$M \colon X \to \Gamma = X,$$

where Γ is an internal language designed so that similar problems have similarly "landscape" representations. In this setup, M effectively acts as a permutation of the solutions. By definition, experts are those for whom $V \circ M$ has a sufficiently increasing landscape. When they encounter a similar problem (under any measure d), they represent the problem in a similar manner and therefore achieve good results on both problems. Consequently, experts can solve similar problems even if the general shape of $V \colon X \to [0,1]$ is not "smooth," because the shape of $V \circ M$ is.

7.4.1. *Continuation of the example.* Again, agents are subject to framing effects. A group of experts in a deliberation will only perform better than a group of university graduates if the solutions are given in a particular order. That is, even assuming there are no time limitations or disagreements in the deliberation of experts, the order of the solutions will be important to determine the solution to the problem as they won't be able to recognize the solution if presented in a different order. What is more, the most difficult problem can be transformed into the easiest one just by reordering the solutions. That is, if the solutions are reordered properly, solving climate change problems will be as easy as solving a simple household temperature regulation problem.

Similarly, in the original Hong-Page framework, there is no guarantee that the "experts" on the problem of economic consequences of climate change will be the same as the "experts" on the problem of the economic consequences of rising temperatures.

7.5. **Concluding remarks.** The Hong-Page framework is a model that, like all models, relies on simplified assumptions. Ideally, such simplifications serve to filter out irrelevant features of reality and capture the essence of the problem at hand. This would constitute a good model. However, as we have demonstrated, the Hong-Page framework fails to capture the essence of the problem and instead distorts it. Our analysis shows that it leads to conclusions which are, in many cases, absurd—sufficient grounds to reject the model. For example, if the model predicts that one can design simple mechanisms to always locate the solution to any hard problem using random groups, that deliberations always culminate in unanimous agreement or that we can solve any problem starting from every point in the domain with enough agents, then the model's validity must be questioned. If the predictions diverge markedly from empirical observations, at least one of the underlying hypotheses must be rejected. In other words, for proponents of the Hong-Page framework, how extreme must the absurdity of a model's conclusions be in order to recognize that it has been falsified and, consequently, reject it? For instance, if a gravity model always predicted that objects would move toward Everest (or Mount Fuji), such a prediction would cast serious doubt on the model's validity. A similar critique, albeit to a different extent, applies to the Hong-Page framework; it always predict that we can manage to locate the solution to any hard problem using any group of agents and always with unanimous agreement. Other distortions in the modeling choices, coming from taylored assumptions and unnecessary mathematical apparatus—such as an overemphasis on the starting point, unrealistic movement across the problem domain, and a dependence on the problem's shape—further lead to conflicting conclusions regarding the agents' abilities, an outcome that should be avoided.

Another salient issue is the lack of robustness of the framework. It is highly sensitive to several parameters, and simple modifications can yield divergent conclusions (see Sections 3, 4, 6). Coupled with its limited empirical content and questionable real-world applicability, these factors significantly undermine the framework's utility in addressing real-world problems. To quote [Hol+18]:

> Specifically, one must ensure that both the input and the mechanism match the situation in which the model is being put to use. Moreover, because arguments pertain to general outcomes, either such work must be done across the range of possible situations or it must be shown that the set of relevant models robustly yields the same result across the parameter space. Additional empirical work is therefore required to show that both the inputs to the simulation and the mechanisms in the simulation sufficiently capture the real-world situation being modeled. A complete run down of the Hong-Page model is beyond the scope of the paper, but a couple of examples of correspondence should suffice to provide a feel for the additional warrant needed if we are to rely on models in this capacity.

Furthermore, alternative frameworks that reject the slogan "Diversity Trumps Ability" incorporate aspects of realism that are absent from the Hong-Page framework. Thus, adopting a Moorean-style argument, adherence to the Hong-Page framework is tantamount to accepting its inherent shortcomings, "biting the bullet" [Rom23]. Consequently, review analyses such as those in [LS24] offer limited utility. When confronted with two models—one entirely unrealistic and the other more realistic yet yielding contrary results—"averaging" their outcomes is methodologically inappropriate. Similarly, if the critique holds, then reviewing ten studies employing different topologies of X against those asserting that all such X are encompassed by a model with foundational issues would be misguided—comparable to conducting a meta-analysis of treatment effects in biology without accounting for sample size or weighting different experimental features appropriately. That is, even if all studies (the critique and the studies in the review) were true, if ten studies say that the Hong-Page result holds for different topologies of X and other says that focusing on the starting point, $x \in X$ is incorrect, then the latter study will be enough to reject the Hong-Page result. Moreover, if, as we have seen, introducing the possibility of disagreement (which the simulation studies consistently omit) alters the conclusions, then comparing ten studies that incorporate only minor variations of the same model with one that includes disagreement would constitute sufficient evidence to update our beliefs in favor of the alternative framework. Recall that in this paper we have shown that:

- The Hong-Page Theorem cannot explain the simulations, and its basic insight is just a reiteration of the hypotheses. The theorem is "hardcoded" to yield the conclusion that diversity trumps ability.
- The original simulations do not support the slogan "Diversity Trumps Ability" when we modify other parameters not considered in the literature.
- When modifying the framework to include realistic features, the slogan breaks down, most importantly when including disagreement.
- We can design a similar framework, with more realistic assumptions, that yields opposite results. Thus, robustness is not just a matter of parameters, but of the framework itself.

In other words, the mere fact that most studies supported (it is not even true) the slogan "Diversity Trumps Ability" would not, by itself, justify acceptance of the Hong-Page result. Thus, claims like this are unfounded, [LS24]:

> 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.

This review should incorporate weighting that reflects a (subjective) analysis of the included studies. In addition, there are specific issues with this review that mischaracterize some of the studies, as discussed elsewhere in this paper (cf. Footnotes 22, 23, 35) and also analytic errors, see Section 7.4 and Footnotes 15 and 6.

APPENDIX A. THE ROLE OF N_1

Noting that if X is finite, then Φ must also be finite, we can choose N_1 such that almost surely every member appears in the random group. We just need to set:

$$N_1 := \max_{\phi\in\Phi}\{N_\phi, \frac{2}{\mu\left(\{\phi\}\right)}\}. \tag{A.1}$$

In the original proof by Hong and Page, N_1 is set so that every member needed to outperform the best performing agent appears with probability one, cf. Remark 2.9. Thus, N_1 must be large, which virtually[47] guarantees that at least one copy of each member of Φ is included in the "random group". In any case, this N_1 as defined is sufficient to ensure the theorem holds true.

More precisely, we can consider three different N_1 so that the following groups are included almost surely:

(1) Φ, all agents from the given pool or $\Phi_0 \subset \Phi$, a subset of agents that is always sufficient to attain the solution x^*,
(2) $\tilde{\Phi}_0 \subset \Phi_0$, a group that outperforms the best performing agent, i.e, for a small ε_1, $\mathbb{E}\left(V;\tilde{\Phi}_0\right) > 1-\varepsilon_1 \geq \mathbb{E}\left(V;\{\phi^*\}\right)$,
(3) $\{\phi^*,\phi_*\} \subset \Phi$, a pair of agents that is enough to outperform the best-performing agent, ϕ^*.

For each subset, you can define, using (A.1), a $\{N_1^{(i)}\}_{i=1,2,3}$. Each one is sufficient for the theorem to hold. But, let us consider the following:

- If our goal is merely to outperform the best agent, then $N_1^{(3)}$ from the simpler proof above should suffice and is more straightforward.
- Given the same hypotheses, one could simply select $N_1^{(1)}$ and assert a stronger result, as it has been demonstrated in Proposition 2.12. It is not just that they outperform the best agent, but they also almost surely find the optimal solution. So, with the same hypotheses, one can prove a stronger result. Shouldn't one then fully explore the implications of these hypotheses?

Given that $N_1 = N_1(\omega)$ is *not fixed* but rather a "stopping time" that halts when the desired members, as per the selector's or chooser's preference, are included, one might ask why the selector should stop merely to outperform the best agent if the correct solution can always be reached when the hypotheses are met. This more comprehensive result trivially implies the one in [HP04]: if there exists an N_1 such that the correct solution is always reached, then, by monotonicity, there must exist an N_1 such that the other group is outperformed.

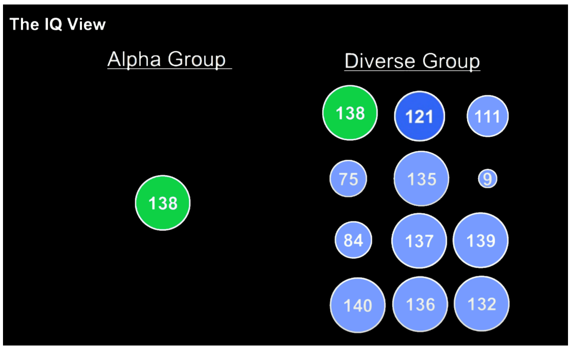
Also worth noting is that one cannot exclude the possibility of other members, different from the ones in the selected group, appearing. For instance, although we may choose $\tilde{\Phi}_0$ such that $\phi^* \notin \tilde{\Phi}_0$, the "random" group might eventually include ϕ^* because of the way these members are chosen, as detailed in Section 3.4. This is why we consistently refer to Φ_0 or Φ, as these capture the full consequences of the hypotheses. In the face of uncertainty, the only certain way to include all necessary agents is to include everyone.

In any case, the triviality of the theorem is manifest for any N_1. Indeed, its basic structure is as follows:

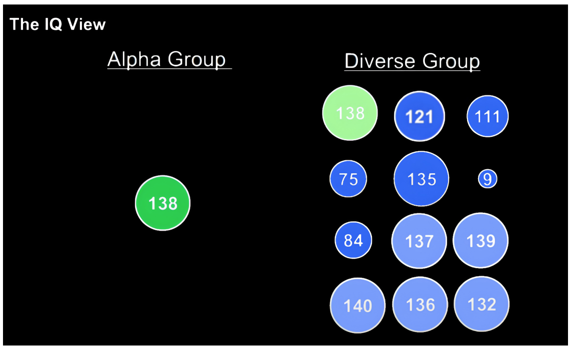
PSEUDO-THEOREM A.1. *Hypotheses*: Let there be a pool of agents such that:

- they never degrade the solution and where there is always an agent who can improve the solution,

[47]In the extreme case where ϕ_0 is not needed at all and $\mu\left(\{\phi_0\}\right)$ is close to zero, we cannot ensure that, even for a large N_1, one copy is included almost surely.



(A) A presentation slide that better encapsulates the essence of the theorem, $N_1 = N_1^{(3)}$.



(B) A presentation slide that better encapsulates the essence of the theorem, $N_1 = \tilde{N}_1^{(2)}$.

FIGURE 16. Different slides. In green, the best agent; in blue, the other agents. Light colors indicate that the probability of appearing is not one (but could be close to one). In the first figure, "121" is assumed to be complementary to the best agent; it can improve where the latter falls short. In the second figure, we are considering that the elements in dark blue are sufficient, by hypothesis, to outperform the agent "138".

- the best performing agent is imperfect by hypothesis,
- as it follows from these two points above, we can choose agents randomly until they outperform the imperfect best performing agent.

Thesis: Consequently, this "random" group will outperform the best performing agent.

A.1. **Numerical example.** For instance, taking the numbers from an example based on Page's slides of a conference at the ECB, see [Rom23; Rom25], with 12 agents, in the vast majority of cases, the best agent will be included in the random group using $N_1^{(2)}$. Additionally, the average N will have the same order of magnitude in all three cases—$N_1^{(1)}$ is not huge—with the following inequalities:

$$17 \approx \mathbb{E}\left(N_1^{(3)}\right) < 24 \approx \mathbb{E}\left(N_1^{(2)}\right) < 36 \approx \mathbb{E}\left(N_1^{(1)}\right),$$

the middle group being sufficient to outperform the best performing agent and consisting in six agents, as shown in Page's slide. For example, if we take $N_1^{(3)}$, the smallest one on average, a more equitable comparison would be as follows, see Figure 16. In the same figure we show what happens if one uses $N_1^{(2)}$, another group that outperforms the best agent. Also note that in that case, the probability of members other than the ones of this group not appearing is practically zero, so they must be represented in the figure in some way.

References

[BL21] Jason Brennan and Hélène Landemore. *Debating Democracy: Do We Need More or Less?* Oxford University Press, 2021 (cited on page 4).

[Gau10] Gerald Gaus. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge University Press, 2010 (cited on page 12).

[Gri+19] Patrick Grim et al. "Diversity, Ability, and Expertise in Epistemic Communities". In: *Philosophy of Science* 86.1 (2019), pages 98–123 (cited on pages 4, 5, 52, 54, 58, 60, 61, 70, 72, 74, 105, 110, 111, 113).

[Gri+24] Patrick Grim et al. "The Epistemic Role of Diversity in Juries: An Agent-Based Model". In: *Journal of Artificial Societies and Social Simulation* 27.1 (2024) (cited on page 52).

[GS20] Patrick Grim and Daniel Singer. "Computational philosophy". In: *Standford Encyclopedia of Philosophy* (2020) (cited on pages 3, 5, 52, 112).

[HMS23] Keith Hankins, Ryan Muldoon, and Alexander Schaefer. "Does (mis) communication mitigate the upshot of diversity?" In: *Plos one* 18.3 (2023), e0283248 (cited on pages 5, 52).

[Hol+18] Bennett Holman et al. "Diversity and democracy: Agent-based modeling in political philosophy". In: *Historical Social Research/Historische Sozialforschung* 43.1 (163 (2018), pages 259–284 (cited on pages 5, 52, 117).

[HP01] Lu Hong and Scott E Page. "Problem solving by heterogeneous agents". In: *Journal of economic theory* 97.1 (2001), pages 123–163 (cited on pages 54, 101, 102).

[HP04] Lu Hong and Scott E. Page. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers". In: *PNAS* (2004) (cited on pages 7, 9, 16, 22, 23, 27, 31, 35, 55, 56, 60, 61, 66, 69, 71, 76, 83, 85, 87, 90, 94, 96, 101, 121).

[HP98] Lu Hong and Scott E. Page. "Diversity and Optimality". In: (1998) (cited on pages 31, 35, 36, 42, 54, 56, 101).

[Kue17] Daniel Kuehn. "Diversity, Ability, and Democracy: A Note on Thompson's Challenge to Hong and Page". In: *Critical Review* 29.1 (2017), pages 72–87 (cited on pages 4, 31).

[LS24] Hélène Landemore and Ryota Sakai. "Revisiting Plato's Cave: On the Proper Role of Lay People versus Experts in Politics". In: (2024) (cited on pages 5, 13, 24, 41, 43, 71, 113, 114, 118, 119).

[Nie+24] Peter Niesen et al. "Does Diversity Trump Ability?" In: *Politische Vierteljahresschrift* 65.4 (2024), pages 785–805 (cited on page 5).

[Pag07] Scott E. Page. "Making the Difference: Applying a Logic of Diversity". In: *Academy of Management Perspectives* 21.4 (2007), pages 6–20. ISSN: 15589080, 19434529 (cited on pages 24, 76, 105, 113).

[Rom22] Álvaro Romaniega. "On the Probability of the Condorcet Jury Theorem or the Miracle of Aggregation". In: *Mathematical Social Sciences* 119 (2022), pages 41–55 (cited on pages 9, 10, 26, 71).

[Rom23] Álvaro Romaniega. "Fatal mathematical errors in Hong-Page Theorem and Landemore's epistemic argument". In: *arXiv preprint arXiv:2307.04709* (2023) (cited on pages 4, 5, 30, 41, 71, 72, 103, 105, 118, 122).

[Rom25] Álvaro Romaniega. "Analysis of the use of the Hong-Page diversity results in the collective intelligence literature". In: (2025). In preparation (cited on pages 4, 30, 71, 103, 122).

[Sak20] Ryota Sakai. "Mathematical Models and Robustness Analysis in Epistemic Democracy: A Systematic Review of Diversity Trumps Ability Theorem Models". In: *Philosophy of the Social Sciences* 50.3 (2020), pages 195–214 (cited on page 5).

[Sin19] Daniel J. Singer. “Diversity, Not Randomness, Trumps Ability”. In: *Philosophy of Science* 86.1 (2019), 178–191 (cited on pages 4, 5, 69).

[SWT16] Keith E Stanovich, Richard F West, and Maggie E Toplak. *The rationality quotient: Toward a test of rational thinking*. MIT press, 2016 (cited on page 104).

[Tho14] Abigail Thompson. “Does Diversity Trump Ability?” In: *Notices of the AMS* 61.9 (2014), pages 1024–1030 (cited on pages 4, 13, 23, 41, 55, 69, 92).

[Wey15] John A Weymark. “Cognitive diversity, binary decisions, and epistemic democracy”. In: *Episteme* 12.4 (2015), pages 497–511 (cited on pages 5, 16).

Email address: `alvaroromaniegadoc@gmail.com`