Humility, ability and group inquiry: A counsel of moderation

Abstract

Research communities differ greatly in their beliefs about the relevance of innate ability to success within the field. Recent work suggests there is a robustly negative correlation between a field's emphasis on innate ability and the representation of racial and gender minorities in the field. This creates the appearance of a tradeoff between the epistemic benefits of cognitive ability and the costs to minoritized investigators. I challenge this apparent tradeoff using recent findings suggesting a negative correlation between emphasis on innate ability and expressions of epistemic humility. Drawing on tools of network epistemology, I show that groups with moderate levels of ability emphasis often epistemically outperform groups with higher levels of ability emphasis by reaping the benefits of increased epistemic humility. This finding softens the tradeoff between the epistemic benefits of cognitive ability and the costs to minoritized investigators. It also yields a novel argument against epistocracy and lessons for the structure of academic philosophy.

1   Introduction

Research communities differ greatly in their beliefs about the relevance of innate ability to success within the field. Recent studies have revealed two facts of note to academic philosophy (Bian et al. 2017; Leslie et al. 2015; Storage et al. 2016). First, academic philosophy emphasizes innate ability more than almost any other field in the humanities or social sciences does. Second, there is a robustly negative correlation between a field's emphasis on innate ability and the representation of racial and gender minorities in the field. In particular, academic philosophy grants a smaller percentage of doctoral degrees to women than almost any other field in the humanities or social sciences does (Figure 1).

My aim in this paper is to build on the case against a high emphasis on the importance of cognitive ability within research communities. I do this by drawing on recent findings linking increased ability emphasis to diminished expression of intellectual humility. Investigators whose fields place a strong value on intellectual ability find it more difficult to admit doubts or mistakes, seek advice, collaborate with others, avoid dogmatism, and remain open-minded or curious towards alternative perspectives (Porter et al. 2022b; Porter

and Cimpian 2023; Porter et al. forthcoming; Vial et al. 2022). These consequences can weigh against the epistemic benefits of cognitive ability.

The underlying lesson of this paper will be a counsel of moderation. It may well be right that moderate levels of ability emphasis are important to the success of inquiring communities. However, it does not follow that more ability emphasis is better – to the contrary, I argue that under many conditions, communities that place a moderate emphasis on intellectual ability outperform those that place lower or higher emphases on intellectual ability.

I make the case for moderation by drawing on tools from network epistemology (Seselja 2022; Zollman forthcoming) to simulate the effects of ability emphasis on the speed and probability of group learning. After some preliminaries (Section 2), I present a Base Model in which inquiring communities use their beliefs about the importance of cognitive ability to determine the investigators allowed into the field (Section 3). Because this Base Model does not incorporate any of the epistemic downsides of ability emphasis, I find in this section that higher degrees of ability emphasis are robustly good for inquiring communities.

The next four sections ask how this result changes once we incorporate the costs of ability emphasis. First, I build on results by Jingyi Wu (2023) and Sarah-Jane Leslie and colleagues (2015) in modeling the costs of ability emphasis in a world where the abilities of minoritized investigators are systematically underestimated (Section 4). I find that the cost to inquiring groups, while severe, may not always be enough to offset the benefits of ability emphasis, motivating a search for further factors telling against high levels of ability emphasis.

The next two sections incorporate two further consequences of ability emphasis that have been associated with decreased expression of intellectual humility: overconfidence (Section 5) and a reduced willingness to seek advice from others (Section 6). I find that under these conditions, communities which place a moderate emphasis on innate ability often outperform those which place lower or higher emphasis on cognitive ability.
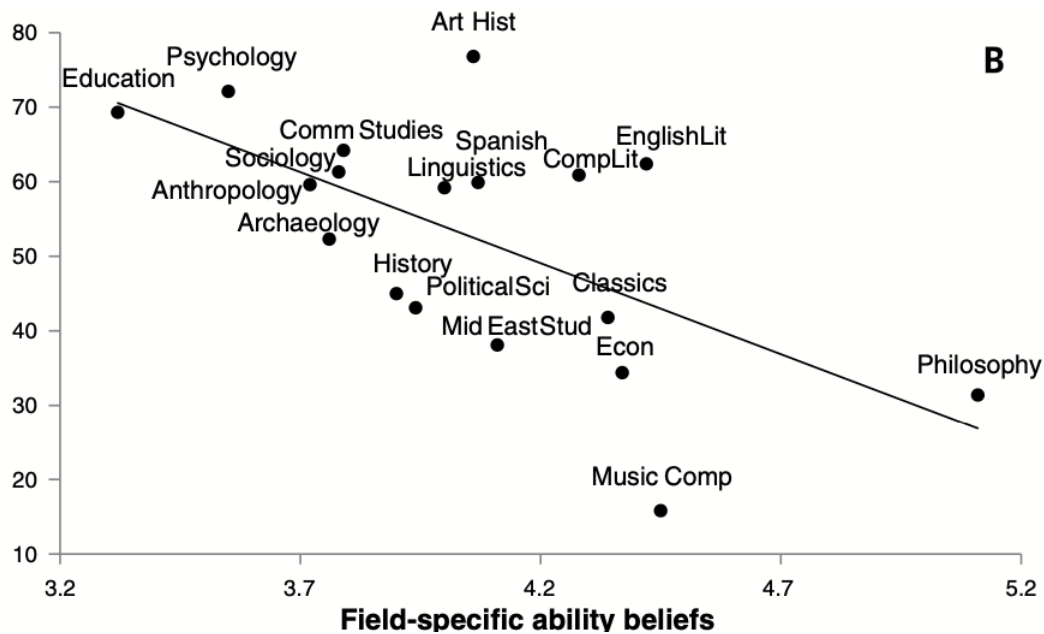
Figure 1: Percentage of doctorates awarded to women against emphasis on brilliance (field-specific ability beliefs), from Leslie et al. (2015)

In particular, when these latter consequences are combined with discrimination against minoritized investigators, I find a robust advantage to communities with moderate levels of ability emphasis (Section 7).

Section 8 concludes by discussing lessons for the structure of philosophy (Section 8.1) and the epistemology of democracy (Section 8.2) as well as potential extensions of the models in this paper (Section 8.3).

2 Theoretical background

2.1 Field-specific ability beliefs

Field-specific ability beliefs reflect the degree to which members of a field take innate ability to be relevant to success within the field (Bian et al. 2017; Leslie et al. 2015). Field-specific ability beliefs are measured by asking investigators to rate their agreement with statements such as the following: "Being a top scholar of [my discipline] requires a special aptitude that cannot be taught" (Leslie et al. 2015).

Recent work suggests that field-specific ability beliefs emerge early in childhood (Bian et al. 2017; Jenifer et al. 2023; Zhao et al. 2022), are shaped by racial and gender stereotypes (Muradoglu et al. 2023; Storage et al. 2020), and lead to decreased representation of minoritized investigators within a research field (Leslie et al. 2015; Storage et al. 2016). We will see below that field-specific ability beliefs have also been connected with decreased expression of intellectual humility.

2.2 Intellectual humility

Intellectual humility is often thought to be a central epistemic virtue. In this paper, I work with a leading conception of intellectual humility as owning one's limitations (Haggard et al. 2018; Whitcomb et al. 2017). For example, investigators may exhibit appropriate humility in avoiding overconfidence by recognizing the limitations of their ability to make predictions based on data, or in seeking the advice of others to complement their own expertise. Intellectual humility is causally associated with a number of desirable traits including reduced dogmatism and increased curiosity, perspective-taking and prosociality (Leary et al. 2017; Porter et al. 2022a,b).

Recent results have suggested that increased levels of field-specific ability beliefs may lead to decreased expressions of intellectual humility not only in the individuals who hold these beliefs, but also across entire academic fields. Investigators eager to demonstrate their own abilities to themselves and others avoid behaviors such as expressing doubt or soliciting opinions from others in order to conform internally and externally to the value placed on ability within the field (Porter and Cimpian 2023; Porter et al. forthcoming). It is reasonable to expect that these humility-reducing consequences of field-specific ability beliefs may weigh against the role those beliefs play in attracting high-ability investigators to the field. This paper asks whether, and under what conditions, the humility-reducing consequences of field-specific ability beliefs combine with their diversity-reducing consequences to promote negative epistemic outcomes for inquiring groups.

2.3 Network epistemology

Network epistemology studies the consequences of epistemic norms and practices in inquiring groups (Seselja 2022; Zollman forthcoming). Drawing on a combination of mathematical theorems and simulation modeling, network epistemologists seek to guide social epistemologists, zetetic epistemologists, philosophers of science, and others interested in understanding how groups may be structured to promote better epistemic outcomes. Recent topics of interest include the epistemic benefits cognitive diversity (Weisberg and Muldoon 2009; Wu 2023); the credit economy of science (Kitcher 1990; Zollman 2018); misinformation (O'Connor and Weatherall 2019; Skyrms 2010); and the relationship between speed and probability of successful learning (Zollman 2007, 2010), among many others.

There are several ways to interpret the explanations provided by network epistemology. Conor Mayo-Wilson and Kevin Zollman (2021) argue that models in network epistemology play many of the same functions as philosophical thought experiments, including justifying counterfactual claims, exploring logical relationships among philosophical theses, illustrating conceptual possibilities and impossibilities, distinguishing explanatory reasons and identifying the causes that explain a phenomenon, and exploring the dynamics of social and physical systems. And at a minimum, computational models provide *proof of concept* (Arnold 2008; Gelfert 2016) or *how-possibly explanations* (Frey and Seselja 2018; Rosenstock et al. 2017), showing how a phenomenon can come about under given conditions and what mechanisms might drive it. These insights can then guide future empirical research.

Some authors have suggested that results in network epistemology become more reliable when they are grounded in empirical data (Lux and Zwinkels 2018) and agree broadly with the results of a family of related models (Aydinonat et al. 2021). The models in this paper will be empirically grounded insofar as they derive the modeled mechanisms from the scientific literature on field-specific ability beliefs. The models add to a family of related results, some of which are described in Section 2.4, and should be considered together existing and future models.

2.4 Related work

I want to close by comparing the work in this paper with two related results in network epistemology. First, the *diversity trumps ability theorem* of Lu Hong and Scott Page (2004) finds that groups of randomly-selected problem solvers may outperform groups of high-ability problem solvers. Roughly, they argue, this happens because randomly-selected groups tend to be more diverse, and the benefits of cognitive ability are outweighed by the benefits of diversity.

This paper builds on Hong and Page's result in three ways. First, while my findings will support Hong and Page's contention that a maximal emphasis on cognitive ability is sometimes worse than no ability emphasis at all, the present paper also asks whether groups that place a moderate emphasis on cognitive ability may outperform those that place maximal emphasis or no emphasis at all. Under many conditions, I find that moderate ability emphasis outperforms higher or lower levels of ability emphasis. Second, many authors have charged that Hong and Page's result relies on a nonstandard conception of intellectual ability (Grim et al. 2019; Reijula and Kuorikosi 2022; Thompson 2014). My results avoid this charge by modeling cognitive ability as an increased ability to conduct scientific experiments, which the Base Model (Section 3) shows is directly relevant to epistemic success. Third, some authors have expressed concern that Hong and Page's result concerns the diversity of cognitive strategies, which may come apart from the demographic diversity of investigators (Huang 2024). My results avoid this concern by explicitly distinguishing an investigator's cognitive strategy from their minoritized status.

A second result is Jingi Wu's (2023) network standpoint epistemology. Building on the same modeling tradition used in this paper (Bala and Goyal 1998), Wu explores the impact of *one-sided testimonial ignorance*, in which dominant groups ignore the testimony of marginalized groups, but not vice-versa. In particular, Wu explores the possibility of using one-sided testimonial ignorance to ground the *inversion thesis* of standpoint epistemology, that marginalized knowers come to be more knowledgeable about many

matters than their dominant counterparts. Wu finds that marginalized communities often do learn more quickly under one-sided testimonial ignorance, lending support to the inversion thesis.

This paper builds on Wu's result in several ways. First, I explore a complementary mechanism by which discrimination against marginalized investigators may hamper epistemic success. Where Wu studies how prejudice may lead the testimony of marginalized investigators to be ignored, I study how prejudice may lead marginalized investigators to be pushed to leave a research community entirely. Second, I combine a model of discrimination (Section 4) with two further harms of ability emphasis (Sections 5-6) to show how the epistemic harms of discrimination combine with other harms of ability emphasis. I find that even under many conditions where the harms of discrimination alone are not enough to counteract the epistemic benefits of ability emphasis, discrimination combines with other harms to make high degrees of ability emphasis epistemically unattractive (Section 7).

3 The Base Model

In this section, I present the Base Model used to study the effects of ability emphasis on epistemic communities. I show that in the Base Model, which does not incorporate the costs of ability emphasis, ability emphasis is robustly good for the success of epistemic communities. Extensions of the model (Sections 4-7) will then ask how this result changes once key harms of ability emphasis are modeled.

3.1 Starting points

The model in this paper builds on a model first introduced by Venkatesh Bala and Sanjeev Goyal (1998). The Bala-Goyal model has been used extensively throughout philosophy to model phenomena such as scientific polarization (O'Connor and Weatherall 2017), confirmation bias (Gabriel and O'Connor 2024), and the epistemic consequences of limited

communication (Zollman 2007),

In the model, a team of investigators must decide which of two acts is best. On one natural interpretation, the acts are competing research strategies, and the best act is the one with the highest chance of success. The success chances, which are unknown to the investigators, are set so that one act is slightly better than the other. Specifically, act A succeeds with probability 0.5, and act B succeeds with probability $0.5 + \epsilon$, with ϵ varied to assess the impact of changes in problem difficulty.

The investigators set out to determine which of A or B is the better strategy. They proceed in rounds. In each round, the investigators do three things. First, they experiment using one of the strategies. Second, they communicate, sharing their results with other investigators. Finally, they update their beliefs on what they have learned.

More specifically, investigators experiment by using the strategy which they currently take to have the highest chance of success. They observe the results of experimentation, either success or failure. Investigators then communicate the strategy that they used together with the result of experimentation to each other investigator. Finally, investigators update their beliefs about both strategies using Bayesian conditionalization, incorporating both their own findings and the findings communicated to them.[1]

Each phase is repeated in subsequent rounds until one of two end conditions is met. First, the investigators may *succeed* by converging on the correct answer. This happens when the average group member's estimate of the success probability of the better act, B, is within 0.001 of the true value. Second, the investigators may *fail* by not converging on the correct answer. This happens when investigators fail to converge for a sufficiently long stretch of time (here, 1,000 rounds) that future convergence is unlikely.

---

[1]Investigators' beliefs are initialized to a beta distribution, with initial observed successes (alphas) and failures (betas) for both acts selected uniformly from the integers between in [1, 10]. Agents who have observed α successes and β failures of an act assign chance $\alpha/(\alpha + \beta)$ to the act succeeding again. At experimentation time, agents then act to maximize $\alpha/(\alpha + \beta)$, and at update time they update their beliefs by adjusting the values of α and β for each act to take account of newly observed successes and failures.

3.2 Incorporating ability

The Base Model builds on the traditional Bala-Goyal model by incorporating two ways in which investigator abilities affect learning communities. First, higher-ability investigators are better experimenters. This captures a sense in which ability is positively relevant to epistemic success. Second, communities exhibit some tendency to favor higher-ability investigators in selecting community members. This captures a sense in which an emphasis on cognitive ability structures the composition of epistemic communities.

To incorporate these ideas, each investigator is assigned an ability level.[2] To incorporate the idea that higher-ability investigators are better experimenters, in each round an agent conducts a number of experiments equal to their ability level. This means that higher-ability investigators can gather and share more information each round.

To incorporate the role of ability in structuring community membership, a fourth phase is introduced after updating. During this phase, one investigator leaves the community and is replaced with another. The investigator to leave is chosen in an ability-weighted lottery.

In this lottery, each investigator is assigned a score.[3] The investigator's score is a combination of their own ability together with the average ability of the group. The weight placed on an investigator's own ability in scoring is determined by the level of emphasis placed by the community on cognitive ability. With maximal ability emphasis, investigators' scores are entirely determined by their own abilities, so that chances of leaving will be proportional to ability. With minimal ability emphasis, investigators' scores are entirely determined by the average group ability, so that chances of leaving will be equal for all investigators. One of the investigators is selected to leave with probability inversely proportional to their score. That investigator is replaced by a new investigator, and the next round begins.[4]

---

[2] Abilities are selected uniformly from the integers in [1,10].

[3] Each investigator's score is α times their own ability plus $(1-\alpha)$ times the average ability of investigators within the group, where α is the group's level of ability emphasis.

[4] The new investigator's ability is randomized, but they inherit the beliefs of the departing investigator. This is done to disentangle the target phenomenon, group ability structure, from the phenomenon of

3.3 Research questions and preview of findings

Throughout this paper, I investigate four research questions. The first is the *categorical question*: under what conditions is some emphasis on cognitive ability beneficial? Most models in this paper will suggest that some emphasis on cognitive ability is often beneficial.

This motivates a second *degree question*: when some degree of ability emphasis is beneficial, how much ability emphasis is optimal? It might seem natural to suggest that if some ability emphasis is beneficial, then more ability emphasis should be better still. However, most models in this paper will suggest that moderate levels of ability emphasis are often best.

There are at least two natural ways to measure learning outcomes: the probability and speed of successful learning. Many epistemic communities desire to learn quickly, but also desire to reduce their chance of not learning at all. Under many conditions, there is an observable 'Zollman effect' in which the probability and speed of successful learning trade off against one another (Zollman 2007, 2010). My third research question is the *Zollman question*: under what conditions does ability emphasis introduce a Zollman effect? Most models in this paper will not produce a substantial Zollman effect, suggesting that the results may have similar implications for communities that value probability or speed of convergence.

Finally, I pose the *cost question*: beyond learning outcomes, what costs do epistemic communities incur by emphasizing ability? As a case study, I consider the cost to minoritized investigators whose abilities may be wrongly underestimated. I find that ability emphasis comes at a steep cost to the representation of minoritized investigators within a field.

In the Base Model, ability emphasis can only benefit epistemic communities. This is because agents with higher ability can gather more information, and ability emphasis helps communities retain higher-ability investigators. The Base Model is therefore useful

---

collective forgetting.

to study the epistemic benefits of ability emphasis, considered in isolation from potential harms.

Below, I report the results of a computational experiment on the Base Model. The remaining models will ask how the results change once the Base Model is enriched to incorporate potential harms of ability emphasis.

3.4   Results

A computational experiment was performed on the Base Model, examining the effects of varied levels of ability emphasis on learning speed and rates of successful learning (Figure 2).[5]

Because the Base Model considers only the learning benefits of ability emphasis, it answers our research questions in the expected way. On the categorical question, ability emphasis is always beneficial. On the degree question, more ability emphasis is always better. On the Zollman question, there is a qualitatively similar effect of ability emphasis on learning speed and failure rates, suggesting that learning speed and failure rates do not substantially trade off in the Base Model. And on the cost question, there are no modeled costs of ability emphasis. We can also verify that ability emphasis had the intended effect, correlating strongly with the average ability of group members at the end of each model run (Figure 2).[6]

The remaining three models will ask what happens to these results as additional consequences of ability emphasis are incorporated into the Base Model.

[5]50,000 model runs were performed for each condition. Ability emphasis was varied among $\{0, 0.25, 0.5, 0.75, 1\}$. Sensitivity analysis revealed that results were not qualitatively sensitive to the number of investigators or problem difficulty. I therefore fixed the number of investigators at 6 and difficulty ϵ at 0.01 throughout all experiments in this paper.

[6]This finding continues to hold across models, and is therefore omitted from future analyses except when relevant, as in the Discrimination Model.
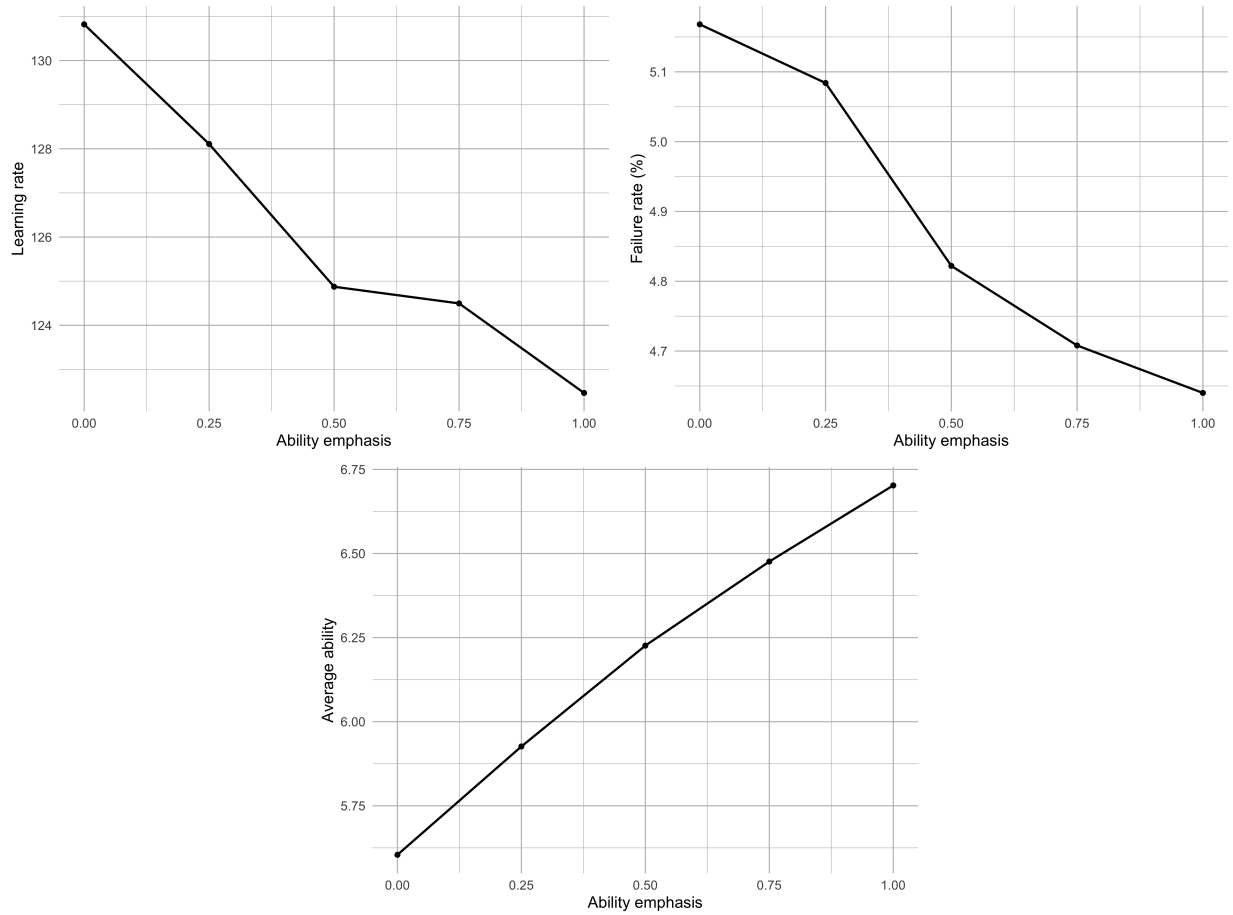
Figure 2: Base Model: Effect of ability emphasis on learning speed (rounds), failure rate (% of trials), and average ability of group members

4 Extension 1: Discrimination

There is good evidence that ability emphasis causally contributes to a decreased representation of minoritized investigators within research communities (Hannak et al. 2023; Leslie et al. 2015). This may happen in two ways. First, a cultural belief that dominant individuals have higher intellectual abilities (Muradoglu et al. 2023; Zhao et al. 2022), emerging in childhood (Bian et al. 2017; Jenifer et al. 2023; Zhao et al. 2022), may lead the abilities of minoritized individuals to be underestimated by others. In fields which value ability, this leads to external pressure against minoritized investigators within the field.

Second, beliefs about the importance of intellectual ability and the distribution of ability among groups may be internalized by marginalized investigators (Muradoglu et al. 2023). Indeed, there is strong evidence that field-specific ability beliefs contribute to impostor syndrome among minoritized investigators (Muradoglu et al. 2022; Vial et al. 2022). Impostor syndrome and other felt deficits in belonging may lead minoritized investigators to leave fields even in the absence of external pressures.

Below, I present a model of discrimination in which the abilities of minoritized investigators are wrongly underestimated. This causes them to exit the field at a greater rate due to a combination of the internal and external mechanisms described above. I ask how the results of the Base Model change once discrimination is allowed to shape the composition of research communities in these ways.

4.1 Model

In the Discrimination Model, investigators are either minoritized or non-minoritized. The initial proportion of minoritized investigators is fixed across trials. Subsequent investigators entering the group are drawn from a population with the same proportion of minoritized individuals as the initial investigators.

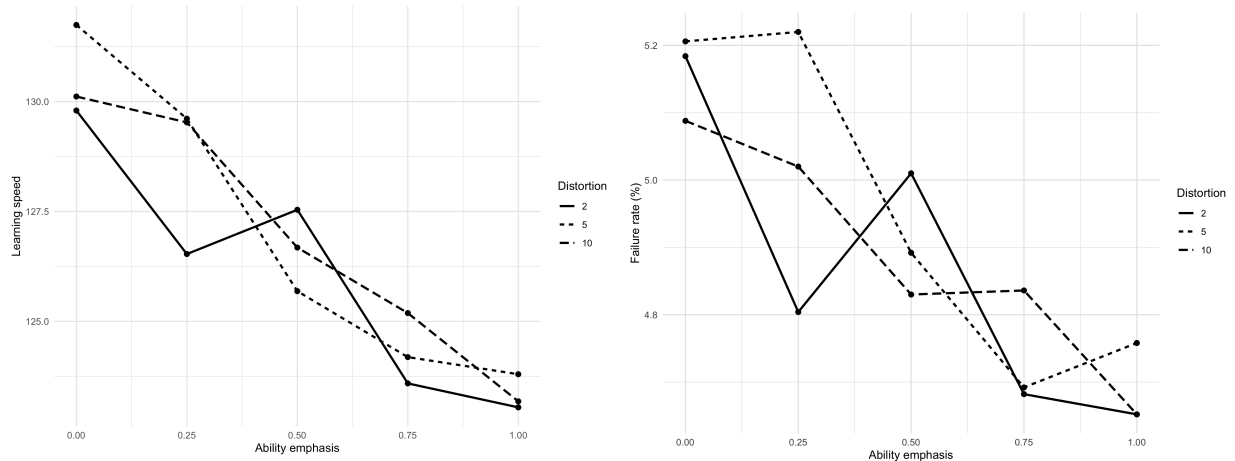In this model, the ability of minoritized individuals is wrongly mis-perceived to be

Figure 3: Discrimination Model: Effect of ability emphasis on learning speed (rounds) and failure rate (% of trials)

lower than its true value.[7] This may be interpreted as reflecting biased community assessments of minoritized investigators' abilities, or else as interpreting minoritized investigators' self-doubts under conditions of stereotype threat. When an investigator is selected to leave the group, the scores assigned to investigators reflect their perceived rather than actual abilities. Because the abilities of minoritized investigators are wrongly perceived to be lower than they are, minoritized investigators have a heightened chance of leaving, and this chance increases in groups with greater degrees of ability emphasis.

I report the results of a computational experiment on the Discrimination Model below.

4.2   Results

A computational experiment was performed on the Discrimination Model. This experiment examined the effects of ability emphasis across varying levels of distortion in how the abilities of minoritized individuals are perceived (Figure 3).[8]

On the categorical question, even in the presence of severe stereotypes, such that

---

[7]This is operationalized by dividing the actual abilities of minoritized investigators by a fixed 'distortion', varied below among $\{2, 5, 10\}$.

[8]As before, ability emphasis was varied among $\{0, 0.25, 0.5, 0.75, 0.1\}$. Distortion was varied among $\{2, 5, 10\}$, with these values selected to illustrate the effects of 'low', 'moderate' and 'severe' distortion, respectively. The proportion of marginalized investigators in the population was fixed at 50%. 50,000 model runs were performed for each condition. Other parameters were fixed as before.
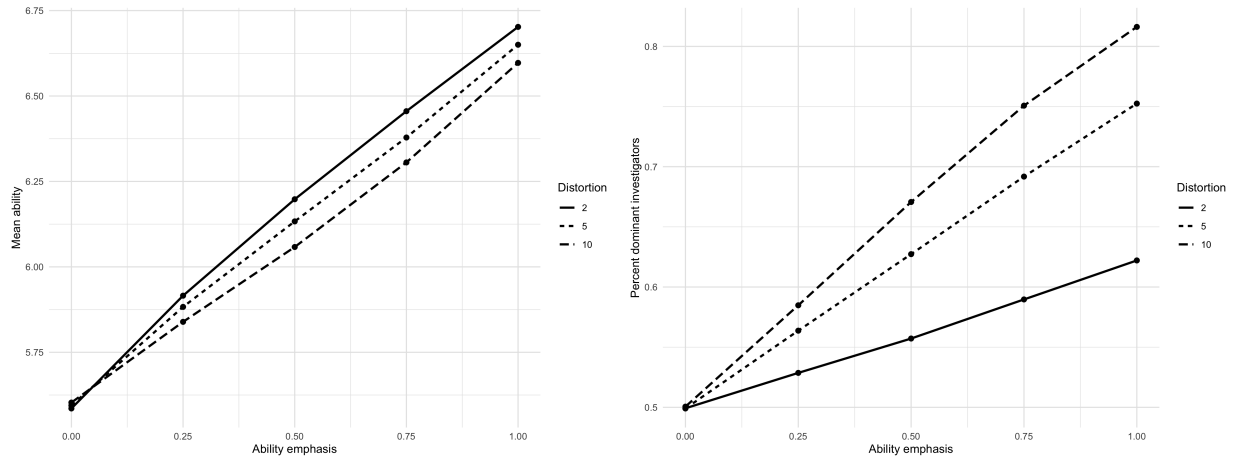
Figure 4: Discrimination Model: Effect of ability emphasis on mean investigator ability and percentage of dominant investigators

minoritized investigators' perceived abilities are one-tenth of their actual abilities, some degree of ability emphasis is beneficial. Similarly, on the degree question, with isolated exceptions, more ability emphasis tends to be better. On the Zollman question there is no clear categorical difference between results expressed in terms of learning speed and failure rates.

One explanation for these findings is that even given severe stereotypes, higher levels of ability emphasis tend to lead to higher-ability groups of investigators (Figure 4). Stereotypes do reduce average investigator ability, because high-ability minoritized investigators are wrongly driven from the field. However, in this model the effects of discrimination on the average ability of investigators are largely offset by the potential to recruit large numbers of high-ability dominant investigators.[9]

This brings us to the cost question. In Figure 4, we see clearly how groups recruit high-ability investigators in the presence of stereotypes: they do this by preferentially recruiting dominant investigators. The proportion of dominant investigators increases strongly in the level of ability emphasis and the severity of stereotypes, so that in the persistence of

---

[9]This is not to say that other unmodeled effects of stereotypes on earlier stages of group inquiry such as experimentation (Schmader et al. 2008; Spencer et al. 2016) and testimony (Fazelpour and Steel 2022; Fricker 2007; Wu 2023) may not combine to put pressure on the benefits of ability emphasis. It is simply to say that the effects of ability emphasis on group composition may not alone be enough to counteract the learning benefits of ability emphasis if the cost to minoritized investigators is ignored.

severe stereotypes, full ability emphasis leads over eighty percent of investigators to be drawn from dominant groups. This is a severe cost to minoritized investigators which may be objectionable in itself (Basu 2019; Dotson 2014; Fricker 2007), or because of downstream epistemic costs to group inquiry (Gendler 2011; Schmader et al. 2008; Wu 2023).

At the same time, these results suggest that we should look to supplement the costs of discrimination with further costs of ability emphasis. The remaining models incorporate humility-related costs of ability emphasis (Sections 5-6), which will combine with discrimination to reverse the learning benefits of high levels of ability emphasis (Section 7).

5 Extension 2: Overconfidence

The remaining models ask how ability emphasis can harm inquiring communities by decreasing expressions of intellectual humility within those communities. One recent finding is that investigators whose communities place a strong value on ability may be less willing to own their limitations by admitting mistakes or confusion (Porter and Cimpian 2023; Porter et al. forthcoming).[10] This happens because investigators are concerned that admitting mistakes or confusion may indicate to others that they have low ability, a trait highly valued within their field. The model in this section shows how an overconfident failure to admit mistakes can lead to suboptimal epistemic outcomes in inquiring groups.

5.1 Model

In this model, overconfidence is reflected as a chance that individuals will believe their results to be more favorable than they are. More specifically, during each round, a given investigator may make a mistake. The chance that each investigator makes a mistake increases in the group's degree of ability emphasis together with an offset parameter con-

---

[10]This unwillingness is linked, in part, to the emergence of a 'masculinity contest culture' in which admitting mistakes is seen as an unacceptable sign of weakness (Porter and Cimpian 2023; Vial et al. 2022).
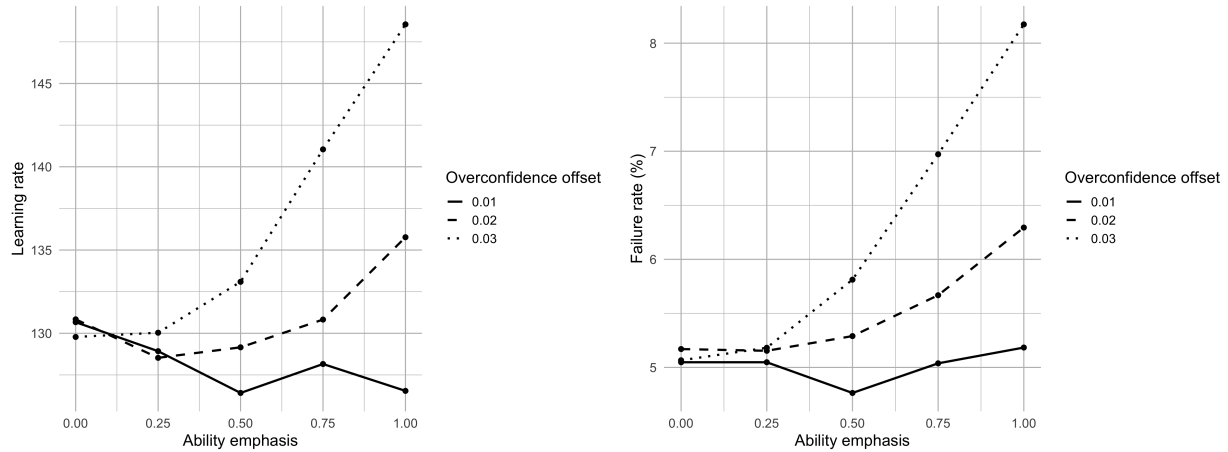
Figure 5: Overconfidence Model: Effect of ability emphasis on learning speed (rounds) and failure rate (% of trials)

trolling the rate at which increases in ability emphasis lead to increased overconfidence.[11] If a mistake is made, the investigator reports that the most common result among their experiments was the only result. For example, an investigator who observed five successes and two failures will, when overconfident, report seven successes and zero failures. I report the results of a computational experiment on the Overconfidence Model below.

5.2 Results

A computational experiment was performed on the Overconfidence Model (Figure 5). This experiment examined the effects of ability emphasis across varying hypotheses about the rate at which ability emphasis leads to overconfidence.[12] With no offsetting (*overconfidence offset* equal to one), the probability that each investigator is overconfident on a given round is equal to the group's ability emphasis. This probability decreases proportionally with the offset, so that for example an offset of 0.1 reduces the chance of overconfidence to one tenth of the group's ability emphasis.

---

[11]Specifically, the chance of a mistake being made is the product of the ability emphasis and the offset parameter, and hence is bounded above by ability emphasis.

[12]Overconfidence offset was varied among {0.01, 0.02, 0.03}. These parameters were chosen to illustrate low and moderate rates of overconfidence at which ability emphasis may be beneficial, together with a higher rate of overconfidence at which ability emphasis ceases to be beneficial. 50,000 simulations were conducted for each combination of parameter values.

On the categorical question, this analysis suggests that any sizable risk of overconfidence may overwhelm the learning benefits of ability emphasis. Even at an offset of 0.03, on which full ability emphasis produces only a 3% risk of overconfidence, increased ability emphasis robustly reduces learning speed and increases the chance of group learning failure. However, at lower offsets, it remains true that some level of ability emphasis improves learning outcomes. This suggests that in communities which are less disposed towards overconfidence, or which have reliable methods for correcting overconfidence, ability emphasis can be beneficial.

On the degree question, even under low offsets it remains true that moderate levels of ability emphasis outperform high levels of ability emphasis. After a point, the risk of overconfidence begins to swamp the learning benefits of bringing higher-ability investigators into the community through increased ability emphasis. On the Zollman question, there is again little evidence for a Zollman effect. And on the cost question, there are no additional modeled costs of ability emphasis.

6   Extension 3: Reduced questioning

Another way that ability emphasis may harm inquiring communities is by decreasing the willingness of investigators to own their limitations by seeking advice when needed. Indeed, there is robust evidence that humility correlates with willingness to seek advice and collaboration (Porter et al. 2022a,b). This finding is replicated in communities where decreased ability emphasis leads to a reduction in intellectual humility and a corresponding willingness to seek advice from others (Porter and Cimpian 2023; Porter et al. forthcoming). In this model, I show how a decreased willingness to seek advice from others may lead to suboptimal epistemic outcomes in inquiring groups.
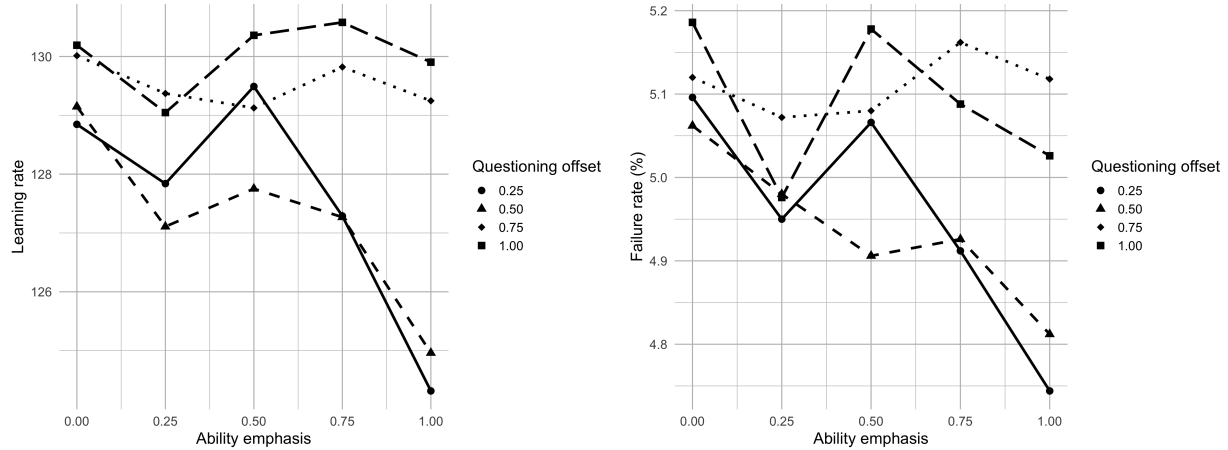
Figure 6: Reduced Questioning Model: Effect of ability emphasis on learning speed (rounds) and failure rate (% of trials)

6.1 Model

In this model, each investigator's chance of asking each other investigator for advice is decreased by the community's ability-emphasis. The rate of decrease is governed by a parameter, the *questioning offset*, governing how strongly community rates of ability emphasis affect investigators' willingness to ask others for advice.[13] The model is otherwise identical to the Base Model.

6.2 Results

A computational experiment was performed on the Reduced Questioning Model. This experiment examined the effects of ability emphasis across varying levels of questioning offset.[14] Results are displayed in Figure 6.

On the categorical question, it remains true even in the highest cost scenarios that some level of ability emphasis is beneficial. On the degree question, we observe an interesting reversal. For each level of questioning offset, moderate levels of ability emphasis perform

---

[13]By analogy to the Overconfidence Model, an investigator's chance of failing to ask an arbitrary colleague to report their findings on a given round is the product of the group's ability emphasis and the questioning offset.

[14]Questioning offset was varied in the range {0.25, 0.5, 0.75, 1}, chosen to span the space of potential offsets. 50,000 simulations were conducted for each combination of parameter values.

well, but the highest levels of ability emphasis perform best. This finding could be explained in two ways. First, as is sometimes found, it could be the case that reduced questioning benefits epistemic communities, for example by preventing the emergence of spurious consensus (Zollman 2007, 2010). Second, it could be that reduced questioning is indeed harmful, but not harmful enough to outweigh the benefits of ability emphasis.

Two facts speak in favor of the second hypothesis, on which the benefits of ability emphasis crowd out the costs of reduced questioning. First, for any fixed level of ability emphasis, increasing the questioning offset produces higher failure rates and longer learning times. This suggests that when the benefits of ability emphasis are held fixed, reduced questioning is epistemically harmful. Second, comparing the results of this experiment to the Base Model (Figure 2) shows that reduced questioning led to longer learning times and higher failure rates than in an identical experiment on the Base Model, where reduced questioning was not present. This suggests again that reduced questioning is genuinely costly. It is just not always costly enough on its own to counteract the benefits of ability emphasis.

On the Zollman question, there remains a strong qualitative match between the results for learning speed and learning chance. No new data was gathered for the cost question, since the Combined Model has the same effect on minoritized investigators as the Discrimination Model does.

So far, we have seen that discrimination and reduced questioning, while costly, are not always individually sufficient to counteract the benefits of high ability emphasis. Overconfidence is individually sufficient to counteract the benefits of ability emphasis on some, but not all assumptions about the relationship between ability emphasis and overconfidence. Will these three phenomena together be enough to tell the desired qualitative story, on which moderate levels of ability emphasis robustly outperform higher and lower levels? I take up this question in the next section.

7 Combined Model

So far, we have seen that moderate degrees of ability emphasis are often but not always best for learning speed and chance of successful learning, though they come at a substantial cost to the representation of minoritized investigators within the field. However, these results consider the harms of ability emphasis in isolation. The Combined Model asks how these findings change when all three negative effects of ability emphasis considered in this paper are modeled together.

The Combined Model incorporates all three model extensions into the Base Model, exactly as they were described above. A computational experiment was conducted across three scenarios. In the 'low cost' scenario, levels of distortion, questioning offset and overconfidence offset are set to their lowest nonzero values considered in the previous sections.[15] In the 'medium cost' scenario, each parameter takes a moderate value, and in the 'high cost' scenario, parameters take nearly their highest studied values. The experiment examined the effects of varying ability emphasis across each scenario (Figure 7).[16]

On the categorical question, it remains true even in the highest cost scenarios that some level of ability emphasis is beneficial. On the degree question, it is now robustly true across conditions that moderate levels of ability emphasis are best for group learning speed and failure rates. No new data was gathered for the cost question, since the Combined Model has the same effect on minoritized investigators as the Discrimination Model does. And on the Zollman question, there remains a strong qualitative match between the results for learning speed and learning chance.

The lessons of the Combined Model therefore strengthen and reinforce the findings from the previous models. Moderate levels of ability emphasis lead to improved speed and probability of successful group learning, but higher levels of ability emphasis lead to

---

[15]More specifically, the values of distortion are 2, 5, and 10 across the low, medium and high-cost scenarios. Questioning offset takes the values 0.25, 0.5 and 0.75 and overconfidence offset takes the values 0.01, 0.02 and 0.03.

[16]Due to the increased number of varied parameters, 10,000 model runs were conducted for each parameter setting.
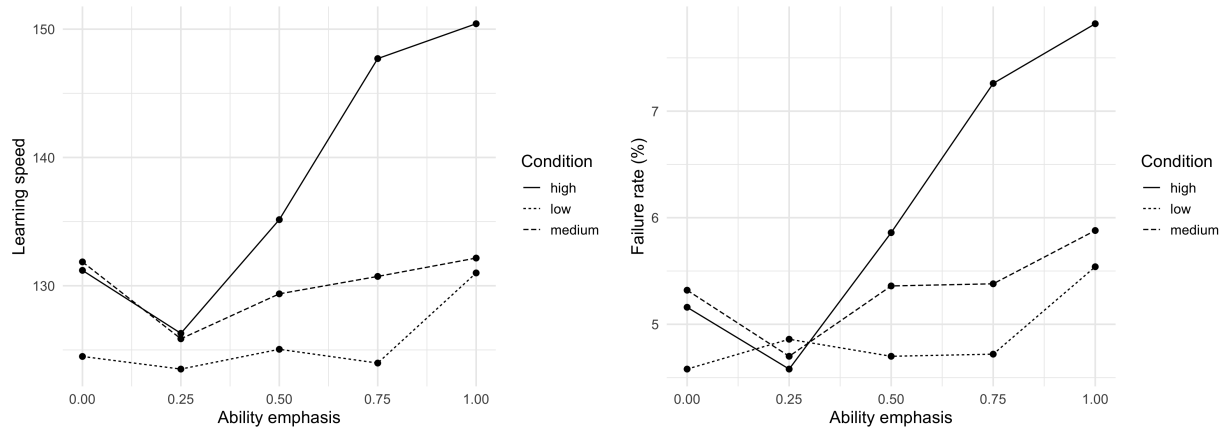
Figure 7: Combined Model: Effect of ability emphasis on learning speed (rounds) and failure rate (% of trials)

severely reduced speed and probability of successful group learning. In some conditions, the effect is strong enough to replicate Hong and Page's (2004) finding that minimal ability emphasis outperforms maximal ability emphasis. In other conditions, it supports only the more moderate contention that moderate ability emphasis outperforms both minimal and maximal ability emphasis. In no case are the lessons substantially different if we are more concerned with learning speed or more concerned with the probability of successful learning. In all cases, our findings are tempered by the result from the Discrimination Model that increased ability emphasis leads to strong decreases in the representation of minoritized investigators.

8 Discussion

The models in this paper studied the effects of ability emphasis on the speed and probability of group learning. Ability emphasis was robustly beneficial in the Base Model (Section 3), which did not incorporate costs of ability emphasis. However, the benefits of ability emphasis decreased in the Discrimination Model (Section 4), which incorporated a systematic tendency to underestimate the ability of minoritized investigators. These benefits were further reduced by incorporating the tendency of ability emphasis to reduce

expressions of epistemic humility by increasing overconfidence (Section 5) and decreasing advice-seeking (Section 6).

Combining these results into a single model (Section 7) suggested that while some degree of ability emphasis is often better than none, a moderate level of ability emphasis tends to outperform more extreme levels of ability emphasis. These models also suggested that ability emphasis comes at a high price in its tendency to exclude minoritized investigators from a research field. Below, I use these lessons to reflect on the structure of philosophy as a research field (Section 8.1) and the epistemology of democracy (Section 8.2). I also suggest productive extensions of the models explored in this paper (Section 8.3) that would increase the robustness of these results and bring them into increased dialogue with related work.

8.1   The structure of philosophy

Recent years have seen growing calls to improve the situation for minoritized students and researchers within philosophy (Antony 2012; Haslanger 2008; Hutchinson and Jenkins 2014). For concreteness, I focus in this section on the status of women in philosophy, though the lessons are broadly similar for other minoritized groups.

Tracy Bowell (2015) distinguishes between three distinct 'woman problems' in philosophy. First, women are underrepresented as students and practitioners of philosophy. Second, women are underrepresented in scholarly discourses including journal publication. Third, feminist scholarship is marginalized within many quarters of philosophy. The findings of this paper bear most strongly on the first problem.

Existing research has noted the role that field-specific ability beliefs play in perpetuating the under-representation of women as students and practitioners of philosophy (Leuschner 2019; Leslie et al. 2015). The findings in this paper build on this insight in two ways.

First, there are two mechanisms by which the influence of field-specific ability beliefs might be counteracted. One approach would be to target stereotypes which lead the abili-

ties of women to be wrongly underestimated. For example, Sally Haslanger recommends that feminists "make the schemas for gender, race, class, and philosophy explicit and defuse them" (2008, p. 219). A second and complementary approach would be to reduce levels of ability emphasis in philosophy.

This approach, while less frequently explored in the literature, is studied by the Discrimination Model (Section 4). The Discrimination Model showed that moderate reductions in ability emphasis can go a long way towards increasing the representation of minoritized investigators. Crucially, the reason for this result is not that women are assumed to have lower philosophical ability – quite the opposite, the Discrimination Model assumes that ability is equally distributed across groups. The finding is rather that moderate levels of ability emphasis may reduce the risk of women being wrongly chased from the field based on inaccurate stereotypes about women's philosophical ability.

Second, existing research has emphasized a number of epistemic benefits of interventions aimed at improving the status of women within philosophy. For example, these interventions may increase the quality and diversity of philosophical theories (Dotson 2011), promote critical reflection on other specialized fields and professions (Friedman 2013) and reduce adversarialness (Moulton 1983). The Overconfidence and Reduced Questioning models (Sections 5-6) add to this list by incorporating the role of reduced ability emphasis in promoting epistemic humility. In particular, these models suggest that moderate levels of ability emphasis may lead to increased questioning and decreased overconfidence, and that these effects may be significant enough to counteract the epistemic costs of decreased ability emphasis. This finding lends support to a tradition which holds that interventions aimed at improving the status of women within philosophy may be epistemically beneficial for the field.

8.2   The epistemology of democracy

Epistemic arguments for democracy hold that democratic systems should be adopted because they are most effective at discovering the truth (Estlund 2007; Landemore 2012).

A number of formal results suggest that the verdicts delivered by large groups of independent deliberators should be especially reliable (Hong and Page 2004; List and Goodin 2001; Surowlecki 2004).

However, some theorists have noted that the same arguments might better support *epistocracy*, a system in which judgments made from a better epistemic position are given additional weight in decisionmaking (Brennan 2017; Jones 2020). For example, educational credentials or performance on civic knowledge exams may confer additional votes, or committees of experts might be given special powers to oversee legislation. This is not an altogether surprising suggestion. From the fact that diverse deliberating groups perform well, it does not follow that they would not perform better if some voices were given additional weight. Indeed, many leading approaches to forecasting problems such as weather prediction suggest that the opinions of forecasters should be weighted based on their credentials or past performance before aggregating opinions (Clemen 1989).

The models in this paper suggest a novel challenge to epistocracy. By linking ability to a central form of civic value, namely the ability to vote or influence legislation, epistocracy runs a significant risk of increasing the level of ability emphasis in democratic deliberation. This, as we saw, decreases the expression of epistemic humility, which in turn may hamper group deliberation. As a consequence, even if it would be better *ceteris paribus* to give more weight to better-informed voters, it does not follow that epistocracy is the best system once the full consequences of epistocracy are considered.

In particular, this discussion suggests a methodological lesson: epistemological arguments for democracy and competing political systems must not take group epistemic facts as fixed. Changing from a democratic to an epistocratic system can influence group dynamics such as the level of ability emphasis, and these changes have tangible epistemic impacts. It may be correct, even provable that certain forms of epistocracy outperform equal-weighting schemes in the abstract, but it need not follow that epistocracy is preferable once the effects of increased ability emphasis are considered.

8.3 Extensions

Like all models, the models in this paper highlight certain empirical phenomena over others. It may be productive to extend the models in this paper to incorporate related phenomena. These extensions would increase the robustness of the modeled results and put these results into dialogue with a number of related literatures. I want to conclude by discussing three possible extensions of these models.

First, we might enrich the Discrimination Model to incorporate further effects of discrimination. The Discrimination Model studies how discrimination can impact field composition by influencing investigators' decisions to enter or exit a field. We might also consider effects of discrimination on testimony, such as a tendency for the findings of minoritized investigators to be marginalized or ignored (Fazelpour and Steel 2022; Fricker 2007; Wu 2023). Alternatively, we could consider how discrimination affects experimentation, for example by allowing ability emphasis to reduce the performance of minoritized investigators through increased levels of stereotype threat (Schmader et al. 2008; Spencer et al. 2016). Incorporating these factors would allow us to reassess whether discrimination alone may be enough to counteract the epistemic benefits of ability emphasis.

Second, recent work has suggested that the network structure of inquiring communities can bear significantly on epistemic outcomes (Zollman 2007, 2010). The models in this paper assumed a fully connected network, in which all investigators can communicate with all others. It may be productive to explore the implications of these models within different network structures, for example in communities with sparse testimonial connections (Zollman 2007, 2010) or where investigators are clustered into tightly connected subgroups (Nguyen 2020).

Finally, we might explore a variety of communication types beyond direct testimony. The models in this paper assumed that investigators directly communicate their results to one another. However, other models have considered indirect forms of testimony, such as journal publication (Coates forthcoming; Zollman 2009) or competition for grant funding (Avin 2019; Harnagel 2019). Future work might ask whether the optimal level of ability

emphasis changes in the context of journal publication or grant funding.

References

Antony, Louise. 2012. "Different voices or perfect storm: Why are there so few women in philosophy?" *Journal of Social Philosophy* 43:227–25.

Arnold, Eckhart. 2008. *Explaining altruism: A simulation-based approach and its limits*. Ontos Verlag.

Avin, Shahar. 2019. "Centralized funding and epistemic exploration." *British Journal for the Philosophy of Science* 70:629–56.

Aydinonat, N. Emrah, Reijula, Samuli, and Ylikoski, Petri. 2021. "Argumentative landscapes: The function of models in social epistemology." *Synthese* 199:369–95.

Bala, Venkatesh and Goyal, Sanjeev. 1998. "Learning from neighbors." *Review of Economic Studies* 63:595–621.

Basu, Rima. 2019. "Radical moral encroachment: The moral stakes of racist beliefs." *Philosophical Issues* 29:9–23.

Bian, Lin, Leslie, Sarah-Jane, and Cimpian, Andrei. 2017. "Gender stereotypes about intellectual ability emerge early and influence children's interests." *Science* 355:389–91.

Bowell, Tracy. 2015. "The problem(s) of women in philosophy: Reflections on the practice of feminism in philosophy from contemporary Aotearoa/New Zealand." *Women's Studies Journal* 29:4–21.

Brennan, Jason. 2017. *Against democracy*. Princeton University Press.

Clemen, Robert. 1989. "Combining forecasts: A review and annotated bibliography." *International Journal of Forecasting* 559–83.

Coates, Matthew. forthcoming. "Does it harm science to suppress dissenting evidence?" *Philosophy of Science* forthcoming.

Dotson, Kristie. 2011. "Concrete flowers: Contemplating the profession of philosophy." *Hypatia* 26:403–9.

—. 2014. "Conceptualizing epistemic oppression." *Social Epistemology* 28:115–38.

Estlund, David. 2007. *Democratic authority: A philosophical framework*. Princeton University Press.

Fazelpour, Sina and Steel, Daniel. 2022. "Diversity, trust and conformity: A simulation study." *Philosophy of Science* 89:209–31.

Frey, Daniel and Seselja, Dunja. 2018. "What is the epistemic function of highly idealized agent-based models of scientific inquiry?" *Philosophy of the Social Sciences* 48:407–33.

Fricker, Miranda. 2007. *Epistemic injustice*. Oxford University Press.

Friedman, Marilyn. 2013. "Women in philosophy: Why should we care?" In Katrina Hutchinson and Fiona Jenkins (eds.), *Women in philosophy: What needs to change?*, 21–38. Oxford University Press.

Gabriel, Nathan and O'Connor, Callin. 2024. "Can confirmation bias improve group learning?" *Philosophy of Science* 91:329–50.

Gelfert, Axel. 2016. *How to do science with models: A philosophical primer*. Springer.

Gendler, Tamar Szabó. 2011. "On the epistemic costs of implicit bias." *Philosophical Studies* 156:33–63.

Grim, Patrick, Singer, Daniel J, Bramson, Aaron, Holman, Bennett, McGeehan, Sean, and Berger, William J. 2019. "Diversity, ability, and expertise in epistemic communities." *Philosophy of Science* 86:98–123.

Haggard, Megan et al. 2018. "Finding middle ground between intellectual arrogance and intellectual servility: Development and assessment of the limitations-owning intellectual humility scale." *Personality and Individual Differences* 124:184–93.

Hannak, Aniko, Joseph, Kenneth, Larremore, Daniel B., and Cimpian, Andrei. 2023. "Field-specific ability beliefs as an explanation for gender differences in academics' career trajectories: Evidence from public profiles on ORCID.org." *Journal of Personality and Social Psychology: Attitudes and Social Cognition* 125:681–98.

Harnagel, Audrey. 2019. "A mid-level approach to modeling scientific communities." *Studies in History and Philosophy of Science Part A* 76:49–59.

Haslanger, Sally. 2008. "Changing the ideology and culture of philosophy: Not by reason (alone)." *Hypatia* 23:210–23.

Hong, Lu and Page, Scott. 2004. "Groups of diverse problem solvers can outperform groups of high-ability problem solvers." *Proceedings of the National Academy of Sciences* 101:16385–9.

Huang, Alice C.W. 2024. "Landscapes and bandits: A unified model of functional and demographic diversity." *Philosophy of Science* 91:579–94.

Hutchinson, Katrina and Jenkins, Fiona (eds.). 2014. *Women in philosophy: What needs to change?* Oxford University Press.

Jenifer, Jalisha B., Jaxon, Jilana, Levine, Susan C., and Cimpian, Andrei. 2023. "'You need to be super smart to do well in math!' Young children's field-specific ability beliefs." *Developmental Science* 27:e13429.

Jones, Garrett. 2020. *10% less democracy: Why you should trust elites a little more and the masses a little less*. Stanford University Press.

Kitcher, Philip. 1990. "The division of cognitive labor." *Journal of Philosophy* 87:5–22.

Landemore, Hélène. 2012. *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton University Press.

Leary, Mark et al. 2017. "Cognitive and interpersonal features of intellectual humility." *Personality and Social Psychology Bulletin* 43:793–813.

Leslie, Sarah-Jane, Cimpian, Andrei, Meyer, Meredith, and Freeland, Edward. 2015. "Expectations of brilliance underlie gender distributions across academic disciplines." *Science* 347:262–5.

Leuschner, Anna. 2019. "Why so low? On indirect effects of gender bias in philosophy." *Metaphilosophy* 50:231–49.

List, Christian and Goodin, Robert E. 2001. "Epistemic democracy: Generalizing the Condorcet jury theorem." *Journal of Political Philosophy* 9:277–306.

Lux, Thomas and Zwinkels, Remco C.J. 2018. "Empirical validation of agent-based models." In Cars Hommes and Blake LeBaron (eds.), *Handbook of computational economics*, volume 4, 437–88. Elsevier.

Mayo-Wilson, Conor and Zollman, Kevin. 2021. "The computational philosophy: Simulation as a core philosophical method." *Synthese* 199:3647–73.

Moulton, Janice. 1983. "A paradigm of philosophy: The adversary method." In Sandra Harding and Merrill B. Hintikka (eds.), *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology, and philosophy of science*, 149–64. Kluwer Academic Publishers.

Muradoglu, Melis, Arnold, Sophie H., Leslie, Sarah-Jane, and Cimpian, Andrei. 2023. "'What does it take to succeed here?': The belief that success requires brilliance is an obstacle to diversity." *Current Directions in Psychological Science* 32:379–86.

Muradoglu, Melis, Horne, Zachary, Hammond, Matthew D., Leslie, Sarah-Jane, and Cimpian, Andrei. 2022. "Women – particularly underrepresented minority women –

and early-career academics feel like impostors in fields that value brilliance." *Journal of Educational Psychology* 114:1086–1100.

Nguyen, C. Thi. 2020. "Echo chambers and epistemic bubbles." *Episteme* 17:141–61.

O'Connor, Callin and Weatherall, James. 2017. "Scientific polarization." *European Journal for Philosophy of Science* 8:855–75.

—. 2019. *The misinformation age: How false beliefs spread*. Yale University Press.

Porter, Tenelle and Cimpian, Andrei. 2023. "A context's emphasis on intellectual ability discourages the expression of intellectual humility." *Motivation Science* 9:120–30.

Porter, Tenelle, Leary, Mark, and Cimpian, Andrei. forthcoming. "Teachers' intellectual humility benefits adolescents' interest and learning." *Developmental Psychology* forthcoming.

Porter, Tenelle et al. 2022a. "Clarifying the content of intellectual humility: A systematic review and integrative framework." *Journal of Personality Assessment* 104:573–85.

—. 2022b. "Predictors and consequences of intellectual humility." *Nature Reviews Psychology* 1:524–36.

Reijula, Samuli and Kuorikosi, Jakko. 2022. "The diversity-ability trade-off in scientific problem solving." *Philosophy of Science* 88:894–905.

Rosenstock, Sarita, Bruner, Justin, and O'Connor, Callin. 2017. "In epistemic networks, is less really more?" *Philosophy of Science* 84:234–52.

Schmader, Toni, Johns, Michael, and Forbes, Chad. 2008. "An integrated process model of stereotype threat effects on performance." *Psychological Review* 115:336–56.

Seselja, Dunja. 2022. "Agent-based models of scientific interaction." *Philosophy Compass* 17:e12855.

Skyrms, Brian. 2010. *Signals: Evolution, learning and information*. Oxford University Press.

Spencer, Steven J., Logel, Christine, and Davies, Paul G. 2016. "Stereotype threat." *Annual Review of Psychology* 67:415–37.

Storage, Daniel, Charlesworth, Tessa E.S., Banaji, Mahzarin R., and Cimpian, Andrei. 2020. "Adults and children implicitly associate brilliance with men more than women." *Journal of Experimental Social Psychology* 90:104020.

Storage, Daniel, Horne, Zachary, Cimpian, Andrei, and Leslie, Sarah-Jane. 2016. "The frequency of 'brilliant' and 'genius' in teaching evaluations predicts the representation of women and African Americans across fields." *PLoS One* 11:e0150194.

Surowlecki, James. 2004. *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Doubleday.

Thompson, Abigail. 2014. "Does diversity trump ability? An example of the misuse of mathematics in the social sciences." *Notices of the AMS* 61:1024–30.

Vial, Andrea, Muradoglu, Melis, Newman, George E., and Cimpian, Andrei. 2022. "An emphasis on brilliance fosters masculinity-contest cultures." *Psychological Science* 33:595–612.

Weisberg, Michael and Muldoon, Ryan. 2009. "Epistemic landscapes and the division of cognitive labor." *Philosophy of Science* 76:225–52.

Whitcomb, Dennis, Battaly, Heather, Baehr, Jason, and Howard-Snyder, Daniel. 2017. "Intellectual humility: Owning our limitations." *Philosophy and Phenomenological Research* 44:509–39.

Wu, Jingyi. 2023. "Epistemic advantage on the margin: A network standpoint epistemology." *Philosophy and Phenomenological Research* 106:755–77.

Zhao, Siqi, Setoh, Peipei, Storage, Daniel, and Cimpian, Andrei. 2022. "The acquisition of the gender-brilliance stereotype: Age trajectory, relation to parents' stereotypes, and intersections with race/ethnicity." *Child Development* 93:e581–97.

Zollman, Kevin. 2007. "The communication structure of epistemic communities." *Philosophy of Science* 74:574–87.

—. 2009. "Optimal publishing strategies." *Episteme* 6:185–99.

—. 2010. "The epistemic benefit of transient diversity." *Erkenntnis* 72:17–35.

—. 2018. "The credit economy and the economic rationality of science." *Journal of Philosophy* 115:5–33.

—. forthcoming. *Network epistemology: How our social connections shape knowledge*. Oxford University Press.