DOI: 10.1017/psa.2025.10174
This is a manuscript accepted for publication in *Philosophy of Science*.
This version may be subject to change during the production process.

ARTICLE

Can Scientific Communities Profit from Evaluative Diversity?

Matteo Michelini[1,2*] Javier Osorio[3]

[1]Ruhr University Bochum, [2] Eindhoven University of Technology, [3] Autonomous University of Madrid
*Corresponding author: Matteo Michelini; matteo.michelini@edu.ruhr-uni-bochum.de

Abstract
Current models of scientific inquiry assume scientists to all share the same evaluative standards. However, science is often characterised by multiple ones, that is by *evaluative diversity*. We investigate how scientific success is affected by evaluative diversity through computer-based simulations. Our results show that communities with diverse standards profit immensely from scientists sharing all the approaches they explored, regardless of whether they considered them valuable. Moreover, we find that even a moderate degree of evaluative diversity can, under certain conditions, lead scientists to reach more satisfying results than those they would reach in homogeneous communities.

1. Introduction

Science is far from being an individual enterprise. Scientists spend a large part of their time talking to colleagues, attending conferences, and, more generally, learning from others. In light of this, significant efforts have been made to understand the conditions for successful collective inquiry (Kitcher, 1993).

To this end, philosophers have developed numerous formal models of group inquiry (Šešelja, 2023). With the help of these models, they have generated a range of insightful findings, such as the idea that a restricted flow of information can enhance inquiry (Zollman, 2010), or that diversity in learning strategies tends to benefit science (Pöyhönen, 2017; Weisberg and Muldoon, 2009). Ultimately, these findings have been used to formulate normative recommendations (Wu and O'Connor, 2023; Petrovich and Viola, 2018; Smaldino et al., 2022).

In all of these models, scientific practice is understood as an instance of collective problem-solving, where scientists aim to find the objectively best approach available by relying both on social learning and individual exploration. Consequently, scientists with the same evidence are assumed to agree on the value of each explored approach (Bedessem, 2019; Politi, 2021).

Yet, this framework fails to capture a fundamental aspect of scientific practice. At times, there may be no unique way to assess existing approaches, and scientists may disagree on their epistemic value, even when they engage with the same approaches and possess the same evidence (Kellert et al., 2006). Because no superior set of standards exist, scientists' evaluations may diverge either due to differing epistemic or non-epistemic values (Chang, 2012; Longino, 1987, 2019), or because they pursue different research goals (Parker, 2006; Nickelsen, 2022), even if their values and goals are all scientifically admissible (Ward, 2022, 2021). In short, scientific communities often reflect a range of values and goals (Longino, 1990), which leads scientists to apply diverse evaluative criteria to the same set of methodological frameworks, theories, or models. We refer to scenarios where these approaches are assessed using different but legitimate standards as instances of *evaluative diversity*.[1]

Although present models have neglected it so far, evaluative diversity introduces specific mechanisms that influence collective scientific inquiry. However, a purely conceptual analysis might struggle to determine its impact, for evaluative diversity presents an unusual trade-off of costs and benefits. On the one hand, it is reasonable to expect scientists with different criteria to explore many different approaches (Reijula et al., 2023), which is usually considered a good recipe for success (Zollman, 2007; Wu and O'Connor, 2023). On the other hand, evaluative diversity may lead to a situation in which each scientist develops her own personal approach, which might not be relevant to others. While evaluative diversity reduces the risk of herding (Strevens, 2013), it may also make social learning superfluous, as scientists are likely to not adopt approaches developed by those who aim to meet different evaluative criteria. As a consequence, several questions are still open. Do the normative recommendations and observations that we obtained from existing models also apply to contexts of evaluative diversity? Does evaluative diversity slow down inquiry? Consider a scientist who is researching a particular problem in her field. Would she gain more by collaborating with peers who share similar standards and goals, or by engaging with those whose criteria and objectives differ?

This paper introduces a computational model of scientific problem-solving in contexts of evaluative diversity.[2] We develop a novel extension of the NK model (Kauffman and Levin, 1987; Wu, 2023) where agents explore approaches by themselves and learn about approaches developed by others, while they may assign different scores to the same approach. Accordingly, we consider the success of a community as the average success of its members, where the success of each member is measured based on the coherence of the approach one adopts with respect to one's standards.

Our results suggest that scientists working in a multi-standards community maximise their success through constant communication. Crucially, agents should not only communicate the approach they adopt and consider valuable, but every approach they discover. In doing so, scientists can benefit from instances of *proxy serendipity*, where an approach deemed of little value by an agent proves invaluable to another. We show that even a moderate degree of evaluative diversity allows members of a community—under

[1] When dealing with formal models, philosophers often use the terms 'theory', 'research approach', 'methods', and 'models' interchangeably when discussing the targets of scientific investigation. In fact, most models of scientific inquiry allow for diverse interpretations, as does the present model (section 3). We will refer to these multiple targets as *scientific approaches*, or simply *approaches*, following Wu (2023) and Thoma (2015).

[2] All data, code and supplementary material are available in the following OSF repository: `https://osf.io/6urwj/?view_only=01cd2aa373ae4a77ba41c53ecf1fb72a`.

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

certain conditions—to reach more satisfying results than those reached by scientists who all share the same standards.

This paper is organised as follows. section 2 elaborates on present models of scientific inquiry, and discusses the notion of evaluative diversity. section 3 introduces the baseline model and the extension designed to incorporate evaluative diversity. section 4 details the main results of the extended model. section 5 contextualises our findings within the existing literature. section 5 concludes.

2. Scientific Inquiry with Evaluative Diversity

Formal models of scientific inquiry share the key assumption that scientific communities are governed by a shared fixed set of standards. According to these standards, the available scientific approaches are objectively ranked and the success of a community is determined by the objective quality of the approach its members ultimately adopt (Smaldino et al., 2022; Šešelja, 2023).

This assumption strongly affects the results obtained by such models. To see this, let us consider one of the most well-known findings in the literature, the 'less-is-more' effect, which suggests that agents should only rarely share information with other agents (Zollman, 2007, 2013), even though only when the problem at stake is especially hard (Rosenstock et al., 2017; Frey and Šešelja, 2020). Because evaluative standards are shared, scientists in existing models are usually highly likely to converge on a successful approach if given the same information. As a consequence, when information flows continuously (e.g. in a complete network), scientists are at risk of rapidly converging on approaches that may initially seem promising, but that could be sub-optimal. Instead, when the information flow is sparse, there is a slower but more thorough exploration of every approach, increasing the likelihood of uncovering optimal approaches especially when two approaches are very similar. In short, under the assumption of a fixed shared set of values, a sparse information flow helps scientists achieve a temporary but efficient division of cognitive labour.

While such assumption suits many instances of collective inquiry, it does not cover all of them. Specifically, it overlooks situations where no unique evaluation standard exists, and scientists explore and discuss the same scientific approaches but assess them differently (Politi, 2021; Bedessem, 2019). In such scenarios, scientists may evaluate the same approach with respect to different, but equally legitimate aims and values, which may lead them to disagree about its quality. Yet, this does not prevent them from learning from each other. In short, present models of scientific inquiry fail to explore collective inquiry in the context of *evaluative diversity*.

These contexts are numerous, as science naturally allows a multiplicity of scientifically legitimate values and objectives (Chang, 2012; Longino, 1987, 2019, 2006; Ward, 2022). In climate modelling, for example, scientists have very different aims, e.g. prediction of global average parameters or simulation of regional climate change. As a consequence, they develop a wide range of different types of simulations for the evolution of Earth's climate, while still learning and possibly adopting models developed by others (Parker, 2006; Winsberg, 2012).

Another illustrative example can be found in the history of photosynthesis research, which stretches from the mid-18th century to 1960, and led to an accurate model of plant photosynthesis (Nickelsen, 2021, 2022). Notably, most scholars contributing to the field

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

did not aim to develop a comprehensive model of photosynthesis: they focused on making significant contributions using familiar methods or addressing specific sub-goals relevant to their own research. For instance, chemist Justus von Liebig, who developed a model for photosynthesis in 1843 (Liebig, 1843), was interested in the topic only with the intention of enhancing crop production. In contrast, Adolf Baeyer, who put forward his model of photosynthesis in 1870, aimed to explain the role of formaldehyde (Bayer, 1870). As Nickelsen notes, both Liebig and Baeyer's models had been considered valuable models for photosynthesis until the 1920s while, at the same time, new models were proposed so as to satisfy different aims. To this end, scientists would simply adapt, modify, and integrate (parts of) available models to their own goals. Previous models would provide 'building blocks' for new ones, developed with completely different goals in mind (Nickelsen, 2022).

Additionally, evaluative diversity may come in degrees (Ludwig and Ruphy, 2021). At one end of the spectrum, in contexts with high diversity, a scientific community may evaluate the same objects in radically different ways, driven by strongly divergent values and goals. At the other end, there may be scientific communities that share an almost unified set of values and goals, whose members hold highly similar evaluative criteria. For example, two climate scientists may evaluate models differently if they aim for different levels of precision and tractability. Instead, climate scientists who aim to model the climate of structurally similar regions with the same levels of precision are likely to have rather similar but not identical evaluative criteria (Parker, 2006).

As mentioned, while scientists may have varying values and goals, this does not prevent them from discussing and adopting each other's approaches. In fact, scientists can learn as much from each other in *heterogeneous communities*, where diverse standards coexist, as in *homogeneous communities*, which rely on a unified set of values and aims. Hence, it is natural to wonder what are the conditions that grant a successful inquiry for heterogeneous communities, and how do they fare with respect to homogeneous ones. To elicit such conditions, we turn to agent-based modelling, in particular to the NK framework, a type of epistemic landscape.

3. The Model

In formal philosophy of science, epistemic landscape models have been used to investigate the division of cognitive labour, the communication structure of scientific inquiry, and both cognitive and social diversity in scientific communities (Weisberg and Muldoon, 2009; Grim et al., 2018; Alexander et al., 2015; Pöyhönen, 2017; Huang, 2024). These models study epistemic communities that explore a set of locations, where each location represents a research approach. Accordingly, each location has a specific height, representing the objective fruitfulness of each approach.[3] Here, we use an NK framework, a type of landscape that represents each approach as the combination of different components, and its quality as an aggregate of the quality of each component. The NK framework was first proposed in biology to study gene evolution (Kauffman and Levin, 1987), and has lately been used by philosophers (Wu, 2023; Reijula et al., 2023).

[3]While the locations of the landscape have been consistently called 'approaches', their heights have received multiple different names in view of slightly different interpretations. While Pöyhönen (2017) and Weisberg and Muldoon (2009) talk about epistemic significance of each of them, Wu (2023) talks of scores.

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

As our aim is to explore the impact of evaluative diversity, this model incorporates novel features not present in the standard NK framework. Traditional NK models assume an evaluation that univocally determines the value of each approach. In contrast, while we also model agents as navigating a shared space of approaches either through social learning or individual exploration, each agent evaluates this space through their own standards. Hence, different agents may assign different values to the same approach. In this sense, agents may be interpreted as navigating different landscapes, even though the set of available approaches remains identical across the community. We refer to 'different landscapes' in later sections as an intuitive shorthand. Accordingly, the degree of satisfaction of a scientist corresponds to the score she assigns to the approach she adopts, and the success of a community is measured as the average degree of satisfaction of its members.

Each *research approach* represents a potential way for an agent to tackle the problem at hand and may be understood differently depending on the context. First, a research approach can be taken broadly as a way of proceeding in a field, i.e. a research programme or methodology (Weisberg and Muldoon, 2009). Scientists may disagree about the quality of a research programme because they may disagree about the relevance of the results that can be obtained with it (Longino, 2019). Second, an approach can represent a full-fledged theory. Scientists often value the same theory differently, e.g. because they value its epistemic virtues differently (Chang, 2012; Schindler, 2022). Finally, an approach can be understood as a possible model for a target phenomenon on which scientists may have contrasting opinions, e.g. a climate evolution model.

We discuss the basic structure of our NK model in section 3.1. We talk about the different learning strategies in section 3.2, and finally about how to evaluate success in section 3.3.

3.1. Approaches and Scores

In an NK landscape, each approach results from N design choices, where each choice selects one component from two available options. Formally, an approach is uniquely defined by a binary string $A = (a_1, a_2, \ldots, a_N)$, where each position $i \in [1, N]$ corresponds to a design choice and a_i represents the component selected for that choice, 0 or 1. A landscape, therefore, consists of all possible binary strings of length N, with each string (i.e. each approach) representing a specific location.

An agent evaluates an approach based on its components, assigning a *score* to each component depending on how well it aligns with her evaluative criteria. The overall score assigned to an approach is the average score of all its components. In turn, the score an agent assigns to a component is influenced not only by the component itself but also by the components chosen for K other design choices (when $K = 0$, the score depends solely on the component). An *atomic evaluative unit* for a component a_i is a string that specifies all the components necessary to determine the score of a_i. Hence, an approach uniquely determines an evaluative unit for every component.[4]

Parameter K determines the degree of complexity of the problems agents face (Wu, 2023; Reijula et al., 2023). A high K represents highly complex challenges, where the

[4] A more technical version of how scores are assigned can be found in an online-only Appendix, available in the OSF repository.

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

quality of a component depends on many other components, and the overall scores of two approaches may be highly different even if they differ only in one component. A low K represents a low complexity challenge, as the scores associated with the components are (almost) independent.

The *personal evaluation* of an agent is a function that assigns to each atomic evaluative unit of each component a specific score, and univocally determines the overall score the agent assigns to any possible approach. A personal evaluation assigns to each evaluative unit a real number between 0 and 1. Although we do not explicitly model agents' evaluative criteria in the model, an agent's personal evaluation is the product of such evaluative criteria: the extent to which the agents' personal evaluations diverge reflects how much their evaluative criteria do so.

Agents are assigned a personal evaluation as follows. For each evaluative unit of each component, we toss a biased coin with a probability of landing head equal to $d \in [0, 1]$, which we call the *diversity coefficient*. If the coin lands head, agents evaluate the unit differently because its value is controversial (e.g. due to unshared background assumptions). If it lands tail, agents evaluate the unit in the same way and assign the same value (e.g. as with an experimental design yielding high statistical power).

A low degree of diversity (d) characterises contexts in which scientists subscribe to (almost) the exact same standards. A high degree of diversity represents communities where scientists disagree over the value of almost every evaluative unit (be it a method, theory, or research programme). While we consider a full range of values for the diversity coefficient, real scientific communities are unlikely to exhibit a very high degree of diversity (e.g. $d > 0.7$), for scientists usually subscribe to a minimum core of shared standards (Kitcher, 2013).

When all scientists agree on the score of a unit (coin lands tail), this score is randomly drawn from a uniform distribution over $[0, 1]$. For each unit on which they disagree (coin lands head), we create a set of real numbers, themselves randomly drawn from a uniform distribution over $[0, 1]$. Each real number in the set represents the outcome of assessing the unit from a specific distinct perspective. Each scientist is then randomly assigned a score from this set, ensuring that each value is equally likely to be assigned. Therefore, for each controversial unit, different groups emerge where members agree internally but disagree with other groups on the score to assign. In fact, in real scientific contexts, whenever the evaluation of a specific component is controversial, scientists usually cluster around a limited set of perspectives (Chang, 2017, 2012; Longino, 2019). The size of this set is determined independently for each evaluative unit and is randomly chosen from the natural numbers between 2 and a maximum value r, a model parameter. This parameter. which we denote as *richness of available perspectives*, controls the maximum number of distinct evaluations per unit. A low value for r (e.g. 3) represents a community where only a few different evaluative perspectives are available, and scientists usually split into very few large groups when evaluating controversial components, while a high value for r (e.g. 20) allows for a greater variety of perspectives.[5]

[5] Although the quantitative results presented depend on this specific set-up, our main findings hold for other different procedures. As a test, we performed simulations under the assumption that agents would draw their score from a normal distribution, and obtained strikingly similar results. See the Supplementary Analysis for details.

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

Consider how this framework accounts for cases of evaluative diversity. If each approach is understood as a research programme, then different components may represent different methods, different research questions, or background assumptions (Wu, 2023). Consequently, scientists may subscribe to different standards when it comes to evaluating a specific method. The same applies to cases where each approach is understood as a scientific theory or a specific model. Each component may then represent a specific modelling choice, or scientific tenet, and different positions may be held with respect to their quality. Diversity in personal evaluations reflects the plurality of values and goals. Specifically, the diversity coefficient (d) determines the probability of disagreeing on the value of a specific unit, while the richness (r) controls the sheer number of perspectives available to evaluate each unit.

3.2. Agents

There are two main behavioural procedures governing agents' actions in the model: a local search algorithm and a social learning protocol. In each step of our simulation, agents first engage in local search, and then in social learning.

The local search rule is an internal optimisation mechanism by which the agents attempt to modify their specific approach, seeking a new approach with a higher score. First, an agent modifies a single component of her approach and lands on a new approach. The agent then compares the score of the new approach with the score of the original one and decides how to proceed. If the score for the new approach is higher, she adopts that approach; if not, she reverts the changed component and returns to the initial approach. Local search, in this context, aims to represent the typical explorations that scientists conduct on their own: scientists change their approaches one component at a time to explore new possible ones (Reijula et al., 2023).

Social learning represents the collaborative aspect of scientific research, where scientists can draw on the work of others. In this procedure, agents observe and potentially adopt the approaches of other agents they are connected with. First, each agent selects the approach(es) she shares. We consider two possible procedures: *partial sharing*, which has been the standard procedure in previous models (Wu, 2023), and *total sharing*.

- **Partial Sharing.** An agent shares the best approach she has found so far. This always corresponds to the approach she employs at the moment of social learning.
- **Total Sharing.** An agent shares up to two approaches: the best one she has found so far (i.e. the one she employs at the moment) and the approach she explored during the last iteration of local search. If these two approaches correspond, the agent shares only one.

Agents share their approach(es) with whoever they are connected with. Then, an agent learns about all the approaches shared by the agents she is connected with, and also learns immediately their corresponding scores based on her own personal evaluation. Consequently, she compares the most valuable among them (with respect to her personal

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

evaluation) to the one she presently adopts. If the new one is more valuable than the present, she moves to the new approach.[6]

Finally, we explore two possible networks: a complete network, which represents systematic communication among all members, and a cycle network, which represents an instance of limited and sporadic communication.

3.3. Model Overview and Success

To determine how evaluative diversity affects scientific inquiry, we model the inquiry process with a computational simulation. First, we assign values to the relevant parameters (Table 1), then initialise the community, and finally let the agents explore their environment while monitoring the progress of the investigation.

During initialisation, each agent is assigned a personal evaluation and a randomly selected starting approach. In line with our focus, we interpret all agents' evaluations as adhering to legitimate scientific standards, i.e. as the result of scientifically admissible aims and values.[7] Once initialised, agents begin their inquiries, performing a local search at each step, followed by social learning. As a result, they explore new approaches and move around in the landscape.

We take the score s_i an agent i assigns to their current approach as an indicator of how successful their inquiry has been thus far, i.e. as an indicator of the agent's *satisfaction*. Consequently, we define the success S of a community as the average value of the satisfaction of individuals, i.e.

$$S = \frac{\sum_i s_i}{n},$$

where n is the number of agents. We choose this criterion—which we name *multiple satisfaction measure*—inspired by Chang (2012), who defines a successful inquiry as one in which a scientist adopts a theory, model, or approach that meets their standards. Thus, community success is determined by the satisfaction of each member with their preferred approach, measured by the score each agent assigns to it. Because all scientists' standards are equally legitimate, their progress should be evaluated based on their own standards.

This measure of success departs substantially from those traditionally used in models of scientific inquiry. These models assume the presence of a unique evaluation criterion (section 2) and measure the success of each scientist based on it.[8] However, as we focus on cases where this assumption does not hold, we resort to the multiple satisfaction

[6] In real scientific communities, the total sharing procedure is likely to require scientists more effort than the partial sharing procedure.

Similarly, learning about new approaches from people with different evaluative standards is also likely to require more effort than learning from people with similar standards. In fact, scientists with different standards are likely to struggle to communicate about the value of their approaches. However, in the present work we leave effort and communication difficulties out of the picture.

[7] Although we are aware that distinguishing a legitimate scientific research programme from a non legitimate one may be rather complicated in real science (Chang, 2017), we leave this matter aside and simply assume that a filter has been applied already. Combining scientifically admissible and inadmissible standards (e.g. due to financial incentives) represents a stimulating future direction of our research.

[8] To highlight the implications of this difference, we later compare our results with those derived from a more conventional measure based on a unique epistemic standard (section 4.5). We are grateful to an anonymous reviewer for suggesting the inclusion of this comparative analysis.

measure. As a consequence, our notion of success is somewhat relative, insofar as the progress of a community is evaluated with respect to the standards that exist within it, rather than against an external, 'objective' one. Yet, we contend that this is appropriate for our scenario, as scientific pluralists have extensively argued (Longino, 1990, 2002; Solomon, 2007; Chang, 2012).

Thus, the ideal outcome of scientific inquiry may involve scientists settling on different approaches. In homogeneous communities, the ideal outcome is one in which each scientist selects the objectively best approach, the success being evaluated according to a shared standard (Wu, 2023; Smaldino et al., 2022). However, when agents' evaluative standards diverge, they may adopt different approaches to maximise their individual scores.

Parameter	**Description**	**Values**
n	Number of agents	5, 10, 20, 30, 40
N	Number of components per approach	10, 15
K	Number of interdependencies	3, 5, 7, 10
d	Diversity coefficient	0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0
Sharing protocol	Approach(es) shared	total, partial
Network	Communication Structure	complete, cycle
r	Richness of available perspectives	2, 5, 7, 10

Table 1. Parameters description and value range explored

For each combination of parameters we run 500 simulations, and for each of these simulations we measure the community success at any given step. Then, we compute the average value for success within a specific set of parameters. In particular, each simulation stops after reaching 150 steps, at which point the community typically reaches a stable state, as agents do not change their approaches any more. We conduct simulations across a range of parameter values, as shown in Table 1. For the baseline scenario, we set $n = 30$, $N = 10$, $K = 5$, $r = 5$. Unless specified otherwise, this configuration generates the presented results.

Thus, the average scientific success of a community serves as a reliable estimate of its expected success under a given set of parameters. Since the multiple satisfaction measure defines community success as the average success of individuals, this value also provides a reliable estimate of an individual's expected success. Under the assumption that each agent is equally likely to start from any approach and hold any possible personal evaluation, the expected value of the multiple satisfaction measure aligns with the expected satisfaction of a single agent. Therefore, comparing the success of different communities under specific parameters offers insight into collective as well as into individual success within each community.

4. Results

In line with previous models, homogeneous communities display well-known dynamics, such as rapidly converging on suboptimal solutions when communication is dense (Lazer and Friedman, 2007; Wu, 2023). However, the situation changes markedly when heterogeneous communities are introduced. These communities perform better when

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

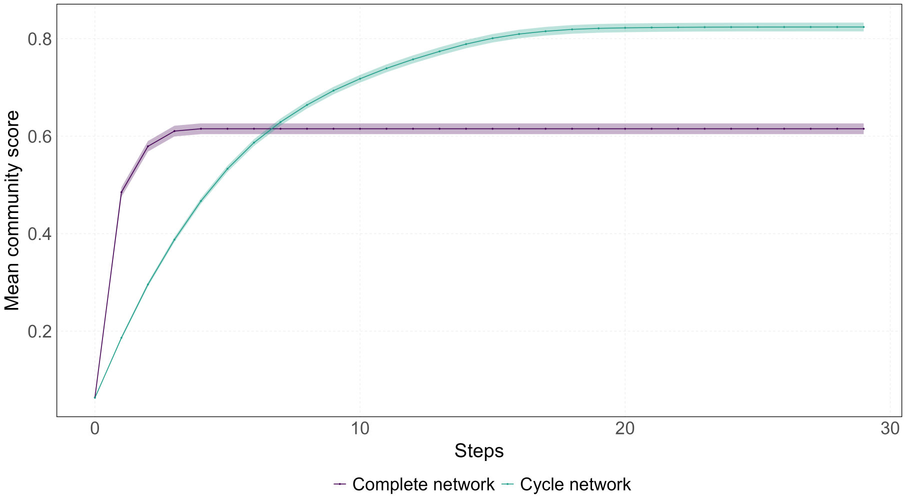

Figure 1. The success for homogeneous communities under different network combinations ($d = 0$, *sharing protocol* = total). Shaded areas represent the standard error of the mean.

groups are highly connected and under conditions of total sharing, as members can benefit from the diverse exploration paths taken by others. Furthermore, we find that communities with a moderate level of diversity in their evaluative criteria outperform homogeneous ones, as long as both are placed in conditions that are optimal for their performance.

To explain how these dynamics play out, we first discuss the behaviour of homogeneous communities and then move to heterogeneous ones.

4.1. Homogeneous Communities

The process of local search can be visualised by imagining a person climbing a landscape of peaks and valleys, where each location corresponds to an approach, and the location's height to the score the agent assigns to it. Accordingly, a peak corresponds to an approach whose score is higher than the score of all the neighbouring approaches, i.e. of all the approaches that can be obtained only by changing one component of the starting approach. In the landscape, an agent can only move upward, meaning she can switch to another approach only if it has a higher score than her current one. Once an agent reaches a peak in her landscape, she cannot move any further using local search alone.

In a homogeneous community, agents assign the same scores to each approach regardless of their groups. Hence, if two agents start from the same location, they are likely to visit the same locations and reach the same peak. Because of this, social learning involves a trade-off. On the one hand, when an agent shares her current adopted approach, it allows others to reach a location that they couldn't have reached by themselves, and consequently increase their success. On the other hand, social learning hinders the exploration potential of a homogeneous community, i.e. the number of approaches a community can still explore, because it leads agents to converge on the same locations.

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

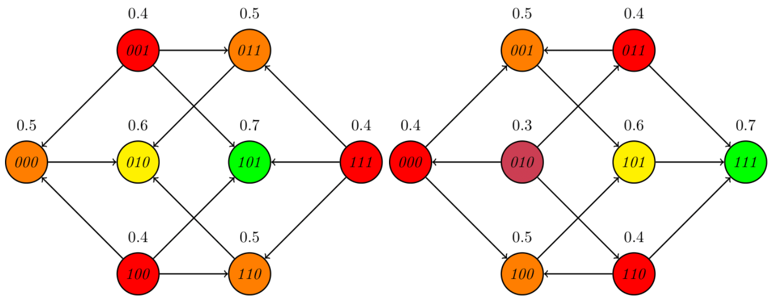

Figure 2. The same set of eight approaches with two different set of scores, obtained through two different personal evaluations. Arrows show available exploration paths.

The effect of the communication structure reflects this trade-off. In the early stages of the simulations, a highly connected community rapidly converges on the highest-scoring approach among those available, often getting stuck in sub-optimal locations. In contrast, communities organised in a cycle network have greater exploration potential, but slower growth. Due to more sporadic communication, they do not immediately exploit each other's findings, leading to extensive exploration of their respective starting areas. This makes the community more likely to converge to higher peaks in the long run (figure 1). This result replicates the 'less-is-more' effect 2.

Total and partial sharing protocols show negligible differences. In fact, regardless of the protocol, agents in homogeneous communities adopt only the approaches that other agents also consider the most valuable.

4.2. Heterogeneous Communities

The first crucial difference between a community with diverse criteria and one with homogeneous ones lies in the way agents explore approaches through local search. While agents with the same evaluative criteria 'see' the same valleys and peaks, agents with diverse criteria may 'see' different peaks, as the score assigned to an approach is determined by their evaluative criteria. Agents within heterogeneous communities are thus likely to follow different exploration paths even if starting from the same approaches.

Example 1. *Consider a set of eight approaches (figure 2), and two agents, i and j, who both start at approach* 000*. If i and j share the same evaluative criteria, with scores as shown in the left graph of figure 2, they converge to the approach* 010 *and stop there. Instead, if agent i assigns scores as indicated in the right graph, while agent j retains the scores from the left one, while j follows the same path, arriving at* 010*, agent i may explore* 001 *or* 100*, then move to* 101*, and finally reach* 111*. The two agents end up on different approaches.*

Different evaluations produce different search heuristics: even if agents follow the same local search rule, this leads them to different positions. As Hong and Page (2004) put it, different preferences may produce different associations between approaches and may lead agents on different exploration paths.

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

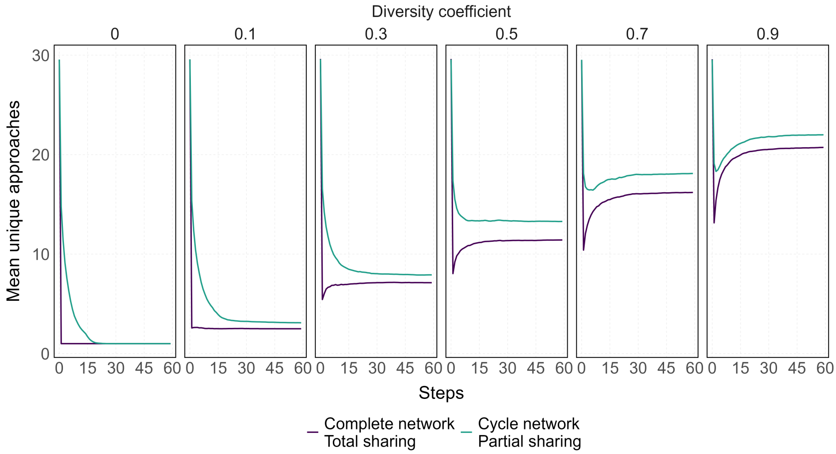

Figure 3. Number of unique approaches adopted at each step by communities with different degrees of diversity.

Hence, scientists with different evaluative criteria rarely converge. To see this, consider figure 3, which shows the average number of unique approaches that are being used in a community.[9] When evaluative criteria are shared, this number decreases rapidly, quickly converging to one or two. Instead, when agents have different evaluative criteria, they settle on different approaches: an increase in evaluative diversity (d) or in the richness of perspectives (r) usually results in an increase in the average number of unique approaches adopted by a community.[10] As a consequence, more diverse communities explore on average a larger number of approaches, as shown in figure 4.

Yet, at the same time, diverse criteria are likely to lead agents to ignore the approaches chosen by their neighbours. As agents try to maximise different criteria, they settle on approaches that are only valuable to themselves. Consider again Example 1. If agent i and agent j have different evaluative criteria, they do not benefit from learning about the approaches chosen by each other: agent i has no interest in moving to approach 010, and agent j has no interest in moving to approach 111. How can then diverse communities benefit from their own exploration potential?

Scientists with different evaluative criteria can profit from each other's explorations as long as they can learn about every approach that other agents explore. The degree to which a heterogeneous community profits from its exploration potential depends

[9] As indicated in the legends of the figures, we usually compare a community organised on a cycle network with partial sharing to a community organised on a complete network with total sharing, as they capture two extremes in terms of information flow: the former involves minimal information exchange, while the latter corresponds to extensive information exchange. Figure 5 provides an overview in terms of community success of all possible combinations. An overview of the behaviour of communities for any combination of network and sharing protocol can be found in the Supplementary Analysis.

[10] Lazer and Friedman observe that in fully connected networks, agents initially converge on few approaches but then go on exploring new ones, producing a 'bouncing' effect—an initial decline in the number of unique approaches followed by a temporary increase—before ultimately settling on a single approach (Lazer and Friedman, 2007, 679). In our model, this effect is absent in homogeneous and minimally diverse communities ($d < 0.3$), but becomes more pronounced as diversity increases. Greater evaluative diversity sustains communal exploration, making the bouncing effect increasingly visible.

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

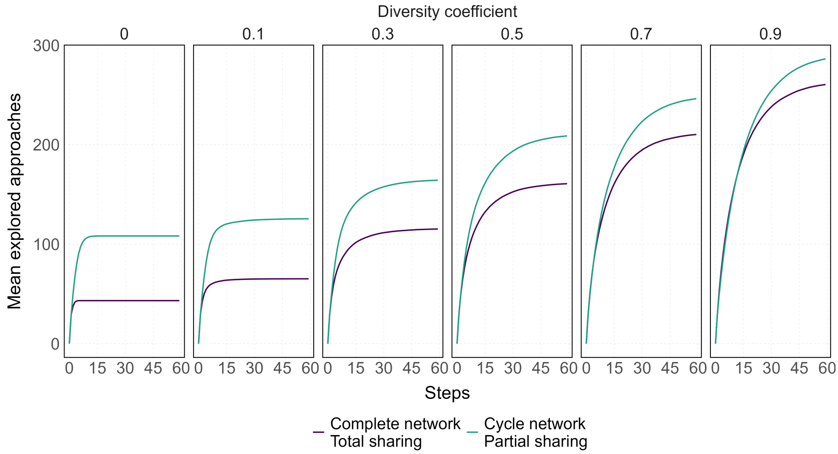

Figure 4. Number of total approaches explored by communities with different degrees of diversity.

on the type of sharing protocol and the communication structure. Whenever scientists share only the approaches they value the most, heterogeneous communities cannot fully exploit their exploration potential because agents are likely to deprive someone else of valuable indications.

Example 2. *Consider the set of approaches in Example 1, such that agent i's personal evaluation is visualised by the left graph of figure 2, and agent j's personal evaluation by the right one. Agent i starts from position* 011 *and explores approach* 111 *before adopting approach* 010*. If she does not share with j details about approach* 111*, she will deprive j of the possibility of learning about a very valuable approach.*

Communication networks have a similar effect. In a homogeneous community, regardless of whether agents are organised on a cycle or a complete network, whenever someone finds a highly valuable approach, sooner or later everybody will converge on it. Valuable information travels either fast or slowly, but it always reaches everyone. Instead, if the community is heterogeneous and agents are not completely all interconnected, much valuable information is at risk of getting lost.

Take again Example 2, and suppose that this time agents share all the approaches they explore but are organised on a cycle. Suppose, for instance, that agent k is connected with both j and i, while i is not connected to j. If agent k does not have interest in approach 111, she may receive information about it from i, but will not adopt it. As a consequence, in the next round of social learning she won't share the details of approach 111 with agent j. Even if the community uses total sharing, the lack of a direct connection between i and j prevents agent j from learning about the approach 111 discovered by i, from which j would have greatly benefited. Instead, if agent i, j, k had the same personal evaluation, the lack of a direct connection between j and i would not harm j. Whenever j would benefit from learning about 111, then agent k would also benefit from adopting it, which would result in k moving on approach 111 and sharing information about it with agent j.

In short, although the great exploration potential of heterogeneous communities (figure 4), their performance is suboptimal whenever agents share only the adopted approaches and are not completely interconnected: a limited flow of information prevents the community from benefiting from exploration.

Combining total sharing with a highly connected network completely changes the picture. As long as every agent shares each new approach she explores, most agents are always informed about the explorations of everyone else. Consequently, agents can benefit from approaches discovered by others, even if the discoverer neither explicitly sought out nor valued those approaches. We refer to these cases as instances of *proxy serendipity*, as they resemble serendipitous discoveries—discoveries that occur to a scientist that was looking for something else (Copeland, 2017, 2023))—carried out by a 'proxy' agent. Consider Example 2: while agent j could benefit greatly from learning about approach 111 from agent i, agent i does not value it at all.

The possibility of proxy serendipity makes social learning an added value for heterogeneous communities. Social learning enables agents to capitalise on others' explorations and even increases the exploration potential of a diverse community. In fact, an agent who moves on a new approach is likely to explore its neighbourhood through new paths. While a dense information flow in moderately heterogeneous and homogeneous communities generates some degree of convergence, it opens up new paths for exploration in radically heterogeneous ones.

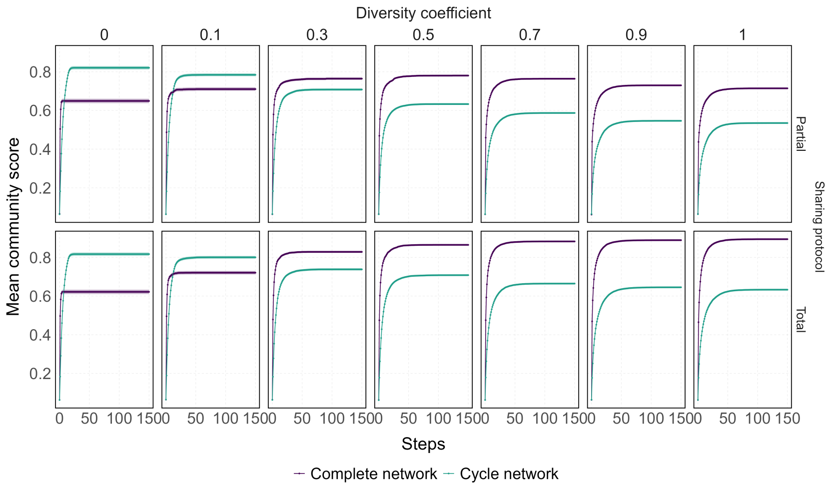

Figure 5. The average success of communities with different degrees of diversity is plotted based on different values for diversity, networks and sharing protocols.

This suggests a crucial difference between homogeneous communities and diverse ones. In homogeneous communities, limiting the information flow does not prevent the community from capitalising on exploration, it simply slows down its ability to do so (figure 1). Important information always finds a way to reach everybody, even if indirectly or after some time. In diverse communities, limiting information can and most of the time does impede exploitation (figure 5). Because information is not equally valuable to everyone, if an approach is not immediately communicated to the person who values it, the information is lost. This suggests the importance of circulating unsuccessful explorations. We return to this point in the Discussion.

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

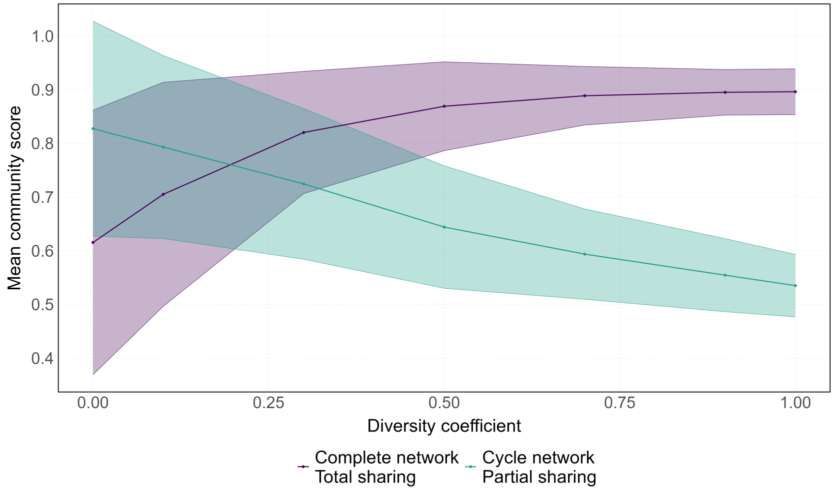

Figure 6. The average success of a community at 150 steps. Shaded areas represent the standard deviation.

4.3. Diverse Communities Beat Homogeneous Ones

How do then homogeneous communities fare with respect to heterogeneous ones? The answer depends on a number of parameters. Nonetheless, figure 5 already hints at the main result of our work, that is that diverse communities may systematically outperform homogeneous ones. The combination of a complete network with total sharing benefits diverse communities to the point that they perform much better than any homogeneous community. This implies that an agent's expected success can be higher in a completely connected, diverse community using total sharing than in a community where every member shares her criteria.

Figure 6 allows us to determine the degree of diversity required for heterogeneous communities to outperform homogeneous ones. For the parameter combination considered in the figure, this threshold lies around 0.3, indicating that even a moderate level of diversity can be sufficient for a community to surpass a homogeneous counterpart organised on a cycle network. However, this value also depends on the complexity of the problem (K)—which we discuss in the next section—as well as the maximum number of unique evaluations available for each unit on which the community disagrees (r).

The effect of r on the success of a community is illustrated in figure 7. An increase in the richness of perspectives corresponds to greater success for heterogeneous communities with total sharing and a complete network, and a lower one for heterogeneous communities with partial sharing and a cycle network. In fact, a greater richness leads to a wider divergence in evaluations, promoting a more thorough exploration of different approaches. This fosters a greater epistemic success whenever agents are all informed of other agents' explorations, or a less successful inquiry, when communication is scarce. The richness of perspectives functions as an amplifier of the effect of evaluative diversity. Consider again figure 7, and compare the performance of a homogeneous community with partial sharing and cycle network (green line in the first box)

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

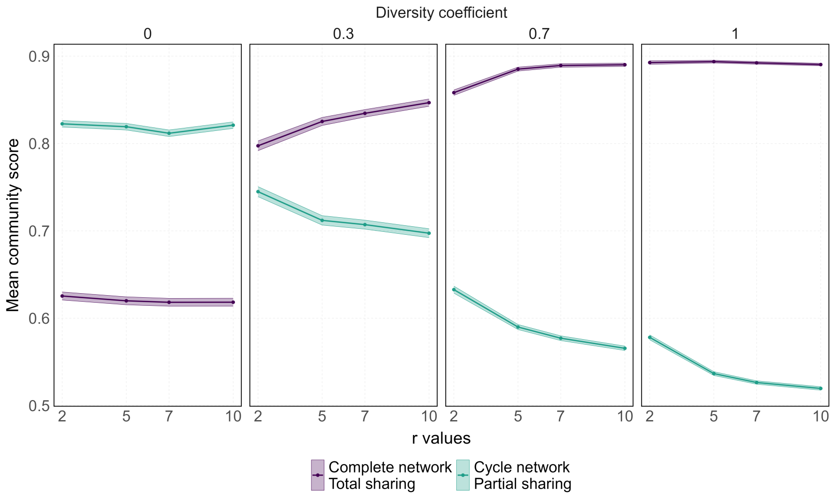

Figure 7. Impact of the maximum number of unique evaluations (r) on the success of communities. Shaded areas represent the standard error of the mean.

with that of a diverse community with $d = 0.3$, complete network, and total sharing (purple line in the second box). The homogeneous community outperforms the diverse one if $r = 2$, while the opposite is true if $r = 10$. Whenever agents can capitalise on diversity, a wide variety of perspectives amplifies its beneficial impact.

figure 6 also provides us with information about the variability of our results. It shows that while the success of a homogeneous community can be highly variable (large standard deviation), increasing the evaluative diversity results in a decrease in variability (increasingly small standard deviation). In fact, homogeneous communities' performance heavily depends on their starting conditions. If agents begin in a region filled with high-quality approaches, they achieve great success; if they start in a less favourable area, the entire community gets stuck with low-success approaches. On the other hand, the diversity in search paths within heterogeneous communities results in much more exploration, preventing the starting conditions to affect strongly the results. The greater variability in homogeneous communities is primarily due to their sensitivity to initial conditions.

Finally, it should be noted that while heterogeneous communities outperform homogeneous ones under specific conditions, they only do so after a certain number of steps. As illustrated in figure 5, in the very short run (e.g. in the first 30-40 steps—although the precise number depends on the parameter combination), homogeneous communities fare much better than diverse ones. This is due to different search patterns: homogeneous communities converge rather easily on superior peaks, while diverse communities may continue exploring for a much longer time (figure 4).

4.4. Community Size and Complexity

Evaluative diversity also affects the way other factors influence the community. First, while homogeneous communities perform rather worse when the complexity of the problem increases (section 4.1), heterogeneous communities with a dense information

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

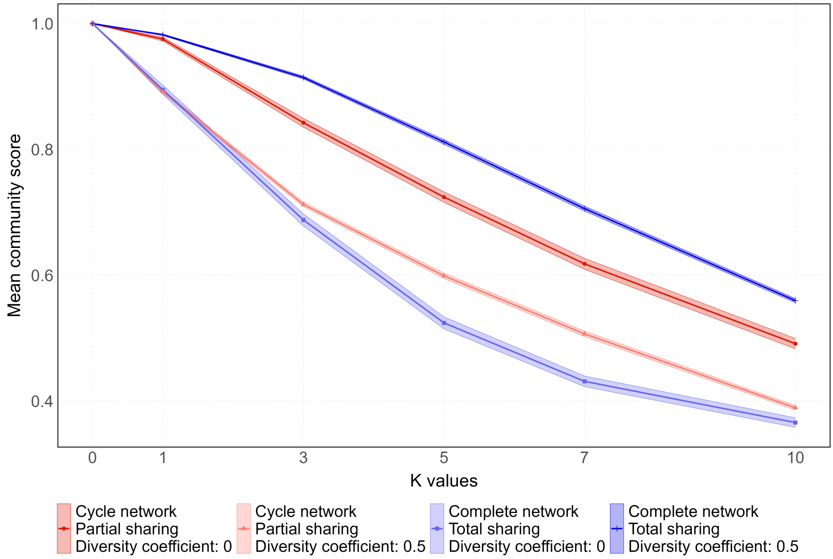

Figure 8. Impact of K on the success of communities ($N = 15$). Shaded areas represent the standard error of the mean.

flow are less affected by it. Figure 8 shows this pattern. Homogeneous communities perform consistently worse because complexity makes individual agents more likely to land on suboptimal local peaks. An increase in K corresponds to a decrease in the potential of local search. Since agents see the same landscape, they are likely to get stuck on the same suboptimal peaks. This effect is less pronounced in the case of heterogeneous communities with a dense information flow because agents follow different paths.

An increase in the size of a community impacts the results in a similar way. As shown in figure 9, a larger number of agents tend to benefit the most homogeneous communities, while heterogeneous communities only gain efficiently from this increase when the flow of information is sufficiently dense. This is the case because adding more agents typically leads to a greater number of locations being explored. This expansion in explored areas is particularly significant in homogeneous communities, where the overall exploration potential is usually lower, allowing them to benefit from the increased size. In contrast, heterogeneous communities only profit from a larger number of explored approaches when every agent has access to that information. Since this is not the case with limited communication, increasing the number of agents has little or no effect.

4.5. Reassessing Community Performance

As discussed in section 3.3, our results are based on a measure of individual success determined by each agent's own evaluation, reflecting the idea that scientists can adhere to different standards, yet epistemically admissible. This approach differs significantly from how NK frameworks have traditionally been used and raises a natural question: how would our results change if epistemic success were assessed using a single, universally correct standard? To address this, we introduce the single-evaluation measure.

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

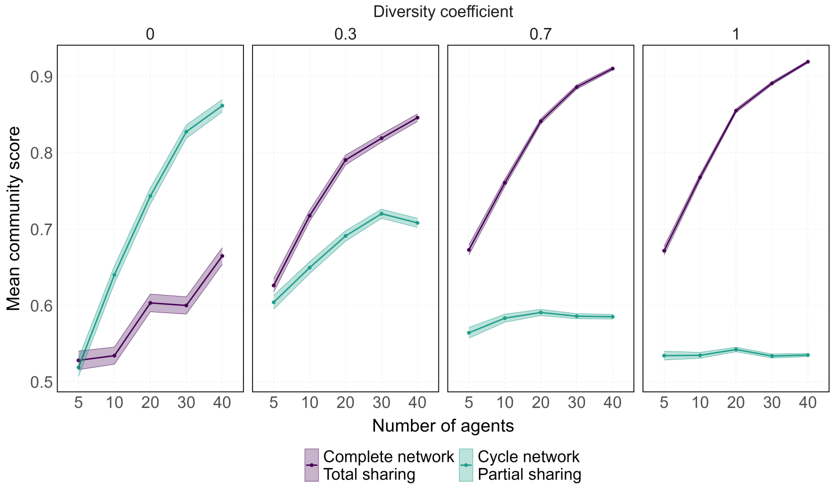

Figure 9. Impact of the number of agents on the success of communities. Shaded areas represent the standard error of the mean.

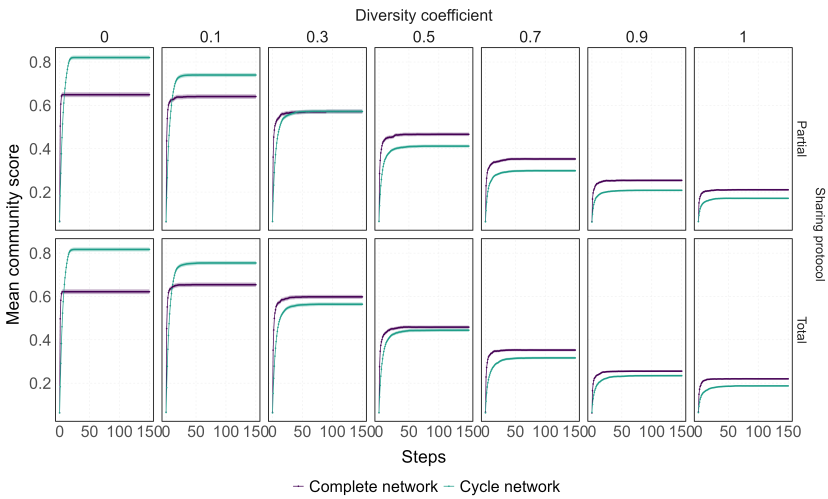

Figure 10. Community performance under the single-evaluation measure. The average success of communities with different degrees of diversity is plotted based on different values for diversity, networks and sharing protocols.

Under this measure, diverse communities never outperform homogeneous ones. While the single-evaluation measure does not reflect the kinds of inquiry our model is designed to capture (section 2), the discussion that follows highlights the importance of the multiple satisfaction measure and serves as an initial robustness check of our findings (Frey and Šešelja, 2020).

Here, we retain the behavioural rules of our model (agents still act according to their own evaluative criteria), but the epistemic success of each agent is now assessed according to only one evaluation, which we shall understand as the only *epistemically correct*

https://doi.org/10.1017/psa.2025.10174 Published online by Cambridge University Press

evaluation. For every community, we randomly select one of the agents and take her personal evaluation to be the correct one. All other evaluations are to be considered epistemically inaccurate. Then, we evaluate every agent's individual success based on the correct evaluation. For homogeneous communities nothing effectively changes from what we had before—as all agents share the same evaluation. Conversely, agents in heterogeneous communities may now be evaluated based on an evaluation they themselves do not use when selecting which approach to adopt.

Under this modified setup, heterogeneous communities perform worse on average as diversity increases (figure 10). This result reflects an increasing misalignment between the standards agents use to guide their search and the one used to measure their epistemic success. In other words, agents in diverse communities may adopt approaches that best meet their own criteria, but these approaches may score poorly under the epistemically correct standard. Figure 10 also illustrates that heterogeneous communities still profit more from a complete network and total sharing than from a cycle network and partial sharing.

This analysis confirms that *if* a single epistemic standard is assumed to be correct, evaluative diversity offers no epistemic benefit. Thus, whether diversity and dense communication structures are epistemically beneficial or not depends on the underlying notion of epistemic success. This suggests that while our results might provide insights concerning the contexts we discussed above (section 2), one should be very careful in over-generalising them to different contexts of inquiry.

5. Discussion

This model explores collective problem-solving in a context of scientific evaluative diversity.[11] Our results can be summarised as follows. First, scientists with diverse standards can effectively collaborate to achieve satisfying outcomes, particularly when they are highly connected and willing to share intermediate results. The key mechanism driving this success, which we call proxy serendipity, allows scientists to benefit from approaches they might not have pursued independently but that ultimately prove valuable to them. Under these conditions, diverse communities may demonstrate greater success in addressing complex problems and consistently outperform homogeneous communities.

These results come with certain limitations. First, it is important to acknowledge the highly idealised nature of our model (Frey and Šešelja, 2020; Martini and Fernández Pinto, 2017). As such, it is designed to explore logical possibilities rather than to provide direct explanations for real-world phenomena. Specifically, our model operates under the idealised assumptions that communication among scientists is always successful and that all the adopted evaluative standards are equally legitimate (section 4.5). Yet, effective communication can be challenging, especially when scientists adhere to different standards, potentially leading to misinterpretations. Additionally, even in diverse scientific communities, not all evaluative standards may be scientifically admissible—some may be influenced by financial incentives rather than epistemic

[11] We argue that our communities engage in *collective* problem-solving despite having different evaluative criteria, as their search processes remain interdependent. Since agents learn from one another's discoveries, problem-solving retains a fundamentally collective dimension.

considerations. We recognise these limitations and plan to relax both assumptions in future work.

Nonetheless, a number of observations can be made in view of our findings. In what follows, we first discuss evaluative diversity in relation to the broader literature on diversity. Next, we explore the potential implications of our findings for general philosophy of science.

5.1. Evaluative Diversity as a New Type of Diversity

To our knowledge, our model is the first to formally examine the impact of a plurality of values and aims on scientific practice. As such, it contributes to the growing body of formal literature on the role of diversity in science by adding a missing piece to it.

Our results suggest that the impact of diversity in personal evaluations is related to, but not reducible to, the effects of other kinds of diversity (Steel et al., 2018), e.g. demographic diversity (Huang, 2024; Steel et al., 2021) or cognitive diversity (Pöyhönen, 2017). Diversity in evaluations naturally leads scientists to adopt different approaches, and consequently results in a thorough exploration of various options. Furthermore, scientists with different evaluations explore the available approaches as if they follow different search heuristics (section 4.2).This supports Hong and Page's (2004) hypothesis that different preferences may lead to different search strategies and complements the formal literature on learning heuristics (Reijula et al., 2023; Weisberg and Muldoon, 2009; Reijula and Kuorikoski, 2022).

Because evaluative diversity naturally fosters the exploration of various alternative approaches, looking at contexts characterised by it requires a shift in the modeller's mindset. In homogeneous communities, one is usually looking for mechanisms that could prevent scientists from herding (Wu and O'Connor, 2023)). In contexts of evaluative diversity, herding is no longer a problem. When studying in diverse communities, it is necessary to explore what mechanisms allow them to benefit from the exploration potential they naturally have (section 4.2).

5.2. Scientific Communities Can Profit from Evaluative Diversity

Our results also suggest a number of observations concerning the social epistemology of scientific inquiry. First, they further qualify our general understanding of different communication protocols. They indicate that the optimal communication structure depends not only on the complexity of the problem at hand (Rosenstock et al., 2017), the stubbornness of scientists (Frey and Šešelja, 2020), the way they interpret evidence (Michelini et al., 2023), and their research heuristics (Pöyhönen, 2017), but also on the diversity of epistemic standards. In fact, previous work has shown that communities where agents share the same standards may perform better with limited information flow (Wu, 2023), while we find that communities characterised by a plurality of goals and values may be better off communicating as much as possible.

According to our results, heterogeneous communities may greatly profit from agents sharing every approach they explore, regardless of whether they value those approaches (section 4.2). Under our formal assumptions, agents with diverse standards are better off sharing not only their substantive findings but also those they might not personally consider substantive. Any finding could still prove valuable to others, leading to instances of proxy-serendipitous discoveries. Proxy serendipity, where an agent adopts an approach discovered by someone else searching for something different, may be crucial to the success of heterogeneous communities.

These results are consistent with historical research on scientific progress, which shows that scientists can derive significant benefits from approaches initially developed for different purposes (Copeland, 2023; Nickelsen, 2022). Chang describes such cases as instances of 'co-optation', citing, for example, Lavoisier and colleagues, who used phlogistonist experimental results to advance new theories (Chang, 2012). Our model advances this literature by providing a formal framework that can capture this phenomenon and identify the conditions under which it becomes critical for successful scientific inquiry.

Nonetheless, in practice, scientists may not have an incentive to communicate the result of all their explorations (Strevens, 2017). Communicating the details of an approach is costly, and scientists may do so if only they believe this could grant them some rewards. Indeed, scientists seldom share the results of their unsuccessful explorations, as they expect little to no recognition for them, a problem which is usually known as the 'file-drawer' problem (Rosenthal, 1979). In other words, heterogeneous communities may face a collective prisoner's dilemma. While everyone benefits from learning about others' (unsuccessful) explorations, there may be little to no personal benefit in sharing them. This dilemma resembles the one discussed by Boyer (2014), Heesen (2017) and Strevens (2017), who wonder what could motivate scientists to share intermediate results even if this exposes them to the risk of being scooped. Among the solutions, they suggest that information withholding may be prevented through some sort of Hobbesian contract between scientists (Strevens, 2017), or by adequately rewarding sharing practices (Heesen, 2017; Boyer, 2014). In a similar vein, Copeland (2017) argues that institutions should, upon a new discovery, reward all scientists who contributed to it, even if only by inspiring the researchers who directly worked on the project. Our results suggest that this proposal may be promising: if a scientist is rewarded based on the extent to which she contributes to others' successes, she may be more willing to consistently share her approaches in the hope that someone else finds them valuable.

Finally, our results show that, under specific conditions, heterogeneous communities outperform homogeneous ones. This observation contributes to the ongoing debate on the effectiveness of scientific pluralism (Chang, 2017), particularly with respect to how much we can and should support scientific groups that pursue divergent standards (Kourany, 2010). Chang (2012, 268) challenges the idea that scientific resources should be concentrated on a single line of inquiry adhering to a specific set of standards. Instead, he argues that society should support multiple lines of inquiry, as scientists with different standards can benefit from each other. Our results may be taken to formally refine this argument. Specifically, they indicate that a scientist following a particular set of standards may be more likely to achieve greater success—that is, to discover a more valuable approach based on her standards—in a heterogeneous community rather than a homogeneous one, especially if the problem at stake is complex. However, this advantage holds only if the heterogeneous community is fully connected, with agents actively sharing any approach they explore. In short, funding research with diverse goals can be an optimal strategy, but only when accompanied by a system of incentives that encourage scientists to share all the approaches they explore with as many colleagues as possible.

6. Conclusion

In this paper, we explored the effect of a plurality of values and goals on scientific problem-solving. We developed a formal model to simulate scientific inquiry in contexts where agents have different evaluative standards and, consequently, evaluate the same approaches differently. Our results suggest that a moderate degree of evaluative diversity, combined with a dense communication structure and willingness to share intermediate approaches, benefits the entire community.

This work represents an initial step toward adapting formal models of collective inquiry to the diverse nature of scientific practice. A possible direction for future research is to integrate in the present model the possibility for scientists' standards to change. This would allow to study collective inquiry as a process in which scientific standards and scientific exploration influence each other continuously.

Acknowledgments. We would like to thank Dunja Šešelja, Christian Straßer, Wybo Houkes, Sanaa Jukola, Klee Shöppl, Leon Assaad, Rafael Fuchs, Jonas Stein, Felix Kopecky, Cristoph Merdes, and three anonymous referees for valuable feedback on earlier versions of this article.

Declarations of competing interests. None to declare.

Funding information. This research was supported by the Ministerio de Economía y Competitividad (grant number FFI2017-87395-P) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 426833574.

References

Alexander, J. M., Himmelreich, J. and Thompson, C. (2015) Epistemic landscapes, optimal search, and the division of cognitive labor. *Philosophy of Science*. *82*(3), 424–453. https://doi.org/10.1086/681766.

Bayer, v. A. (1870) Ünteration of water and its significance for plant life and fermentation. *Ber. d. deutsch. chem. Ges*. *3*(63), 1870.

Bedessem, B. (2019) The division of cognitive labor: Two missing dimensions of the debate. *European Journal for Philosophy of Science*. *9*, 1–16. https://doi.org/10.1007/s13194-018-0230-8.

Boyer, T. (2014) Is a bird in the hand worth two in the bush? or, whether scientists should publish intermediate results. *Synthese*. *191*(1), 17–35. https://doi.org/10.1007/s11229-012-0242-4.

Chang, H. (2012) *Is Water H2O? Evidence, Pluralism and Realism*. Springer.

Chang, H. (2017) Is pluralism compatible with scientific realism? . In *The Routledge handbook of scientific realism*. Routledge. pp. 176–186.

Copeland, S. (2023) Serendipity and the history of the philosophy of science . In *Serendipity Science: An Emerging Field and its Methods*. Springer. pp. 101–123. https://doi.org/10.1007/978-3-031-33529-7-6.

Copeland, S. M. (2017) On serendipity in science: Discovery at the intersection of chance and wisdom. *Synthese*. (6), 1–22. https://doi.org/10.1007/s11229-017-1544-3.

Frey, D. and Šešelja, D. (2020) Robustness and idealizations in agent-based models of scientific interaction. *The British Journal for the Philosophy of Science*. *71*(4), 1411–1437. https://doi.org/10.1093/bjps/axy039.

Grim, P., Singer, D., Bramson, A., Holman, B., McGeehan, S. and Berger, W. (2018) Diversity, ability, and expertise in epistemic communities. *Philosophy of Science*. *86*. https://doi.org/10.1086/701070.

Heesen, R. (2017) Communism and the incentive to share in science. *Philosophy of Science*. *84*(4), 698–716. https://doi.org/10.1086/693875.

Hong, L. and Page, S. E. (2004) Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*. *101*(46), 16385–16389. https://doi.org/10.1073/pnas.0403723101.

Huang, A. C. W. (2024) Landscapes and bandits: A unified model of functional and demographic diversity. *Philosophy of Science*. *91*(3), 579–594. https://doi.org/10.1017/psa.2023.169.

Kauffman, S. and Levin, S. (1987) Towards a general theory of adaptive walks on rugged landscapes. *Journal of theoretical Biology*. *128*(1). https://doi.org/10.1016/S0022-5193(87)80029-2.

Kellert, S. H., Longino, H. E. and Waters, C. K., editors (2006) *Scientific pluralism*. University of Minnesota Press. Minneapolis.

Kitcher, P. (1993) *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press. New York.

Kitcher, P. (2013) Toward a pragmatist philosophy of science. *Theoria*. *77*, 185–231.

Kourany, J. A. (2010) *Philosophy of Science after Feminism*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732623.001.0001.

Lazer, D. and Friedman, A. (2007) The network structure of exploration and exploitation. *Administrative science quarterly*. *52*(4), 667–694. https://doi.org/10.2189/asqu.52.4.667.

Liebig, J. (1843) Die wechselwirthschaft. *Annalen der Chemie und Pharmazie*. *46*, 58–97.

Longino, H. (2002) *The Fate of Knowledge*. Princeton University Press. Princeton.

Longino, H. (2006) Theoretical pluralism and the scientific study of behavior In *Scientific Pluralism*, Kellert, S., Longino, H., and Waters, C. K. (eds). University of Minnesota Press. pp. 102–31.

Longino, H. E. (1987) Can there be a feminist science? *Hypatia*. *2*(3), 51–64. https://doi.org/10.1111/j.1527-2001.1987.tb01341.x.

Longino, H. E. (1990) *Science as Social Knowledge*. Princeton University Press. Princeton.

Longino, H. E. (2019) *Studying human behavior: How scientists investigate aggression and sexuality*. University of Chicago Press.

Ludwig, D. and Ruphy, S. (2021) Scientific Pluralism In *The Stanford Encyclopedia of Philosophy*, Zalta, E. N. (eds). Metaphysics Research Lab, Stanford University. Winter 2021 edition.

Martini, C. and Fernández Pinto, M. (2017) Modeling the social organization of science: Chasing complexity through simulations. *European Journal for Philosophy of Science*. *7*, 221–238. https://doi.org/10.1007/s13194-016-0153-1.

Michelini, M., Osorio, J., Houkes, W., Šešelja, D. and Straßer, C. (2023) Scientific disagreements and the diagnosticity of evidence: How too much data may lead to polarization. *Journal of Artificial Societies and Social Simulation*. *26*(4), 5. 10.18564/jasss.5113.

Nickelsen, K. (2021) Cooperative division of cognitive labour: The social epistemology of photosynthesis research. *Journal for General Philosophy of Science*. *53*(40). https://doi.org/10.1007/s10838-020-09543-1.

Nickelsen, K. (2022) *Explaining Photosynthesis. Models of biochemical mechanisms, 1840-1960*. Springer.

Parker, W. (2006) Understanding pluralism in climate modeling. *Foundations of Science*. *11*(4), 349–368. https://doi.org/10.1007/s10699-005-3196-x.

Petrovich, E. and Viola, M. (2018) Social epistemology at work: from philosophical theory to policy advice. *RT. A Journal on Research Policy and Evaluation*. *6*(1). https://doi.org/10.13130/2282-5398/9828.

Politi, V. (2021) Formal models of the scientific community and the value-ladenness of science. *European Journal for Philosophy of Science*. *11*(4), 97. https://doi.org/10.1007/s13194-021-00418-w.

Pöyhönen, S. (2017) Value of cognitive diversity in science. *Synthese*. *194*, 4519–4540. https://doi.org/10.1007/s11229-016-1147-4.

Reijula, S. and Kuorikoski, J. (2022) The diversity-ability trade-off in scientific problem solving. *Philosophy of Science (Supplement)*. https://doi.org/10.1086/714938.

Reijula, S., Kuorikoski, J. and MacLeod, M. (2023) The division of cognitive labor and the structure of interdisciplinary problems. *Synthese*. *201*(6), 214. https://doi.org/10.1007/s11229-023-04193-4.

Rosenstock, S., Bruner, J. and O'Connor, C. (2017) In epistemic networks, is less really more? *Philosophy of Science*. *84*(2), 234–252. https://doi.org/10.1086/690717.

Rosenthal, R. (1979) The file drawer problem and tolerance for null results. *Psychological bulletin*. *86*(3), 638. https://doi.org/10.1037/0033-2909.86.3.638.

Schindler, S. (2022) Theoretical virtues: Do scientists think what philosophers think they ought to think? *Philosophy of Science*. *89*(3), 542–564. 10.1017/psa.2021.40.

Smaldino, P. E., Moser, C., Pérez Velilla, A. and Werling, M. (2022) Maintaining transient diversity is a general principle for improving collective problem solving. *Perspectives on Psychological Science*. p. 17456916231180100. https://doi.org/10.1177/17456916231180100.

Solomon, M. (2007) *Social empiricism*. MIT press.

Steel, D., Fazelpour, S., Crewe, B. and Gillette, K. (2021) Information elaboration and epistemic effects of diversity. *Synthese*. *198*, 1287–1307. https://doi.org/10.1007/s11229-019-02108-w.

Steel, D., Fazelpour, S., Gillette, K., Crewe, B. and Burgess, M. (2018) Multiple diversity concepts and their ethical-epistemic implications. *European Journal for Philosophy of Science*. *8*(3), 761–780. https://doi.org/10.1007/s13194-018-0209-5.

Strevens, M. (2013) Herding and the quest for credit. *Journal of Economic Methodology*. *20*(1), 19–34. https://doi.org/10.1080/1350178x.2013.774849.

Strevens, M. (2017) Scientific sharing, communism, and the social contract In *Scientific Collaboration and Collective Knowledge*, Boyer-Kassem, T., Mayo-Wilson, C., and Weisberg, M. (eds). Oxford University Press. pp. 3–33.

Thoma, J. (2015) The epistemic division of labor revisited. *Philosophy of Science*. *82*(3), 454–472. https://doi.org/10.1086/681768.

Ward, Z. B. (2021) On value-laden science. *Studies in History and Philosophy of Science Part A*. *85*, 54–62. https://doi.org/10.1016/j.shpsa.2020.09.006.

Ward, Z. B. (2022) Disagreement and values in science In *The Routledge Handbook of Philosophy of Disagreement*, Baghramian, M., Carter, J. A., and Cosker-Rowland, R. (eds). Routledge New York.

Weisberg, M. and Muldoon, R. (2009) Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*. *76*(2), 225–252. https://doi.org/10.1086/644786.

Winsberg, E. (2012) Values and uncertainties in the predictions of global climate models. *Kennedy Institute of Ethics Journal*. *22*(2), 111–137. https://doi.org/10.1353/ken.2012.0008.

Wu, J. (2023) Better than best: Epistemic landscapes and diversity of practice in science. *Philosophy of Science*. https://doi.org/10.1017/psa.2023.129.

Wu, J. and O'Connor, C. (2023) How should we promote transient diversity in science? *Synthese*. *201*(2), 1–24. https://doi.org/10.1007/s11229-023-04037-1.

Zollman, K. J. (2007) The communication structure of epistemic communities. *Philosophy of science*. *74*(5), 574–587. https://doi.org/10.1086/525605.

Zollman, K. J. (2010) The epistemic benefit of transient diversity. *Erkenntnis*. *72*(1), 17–35. https://doi.org/10.1007/s10670-009-9194-6.

Zollman, K. J. S. (2013) Network epistemology: Communication in epistemic communities. *Philosophy Compass*. *8*(1), 15–27. https://doi.org/10.1111/j.1747-9991.2012.00534.x.

Šešelja, D. (2023) Agent-Based Modeling in the Philosophy of Science In *The Stanford Encyclopedia of Philosophy*, Zalta, E. N. and Nodelman, U. (eds). Metaphysics Research Lab, Stanford University. Winter 2023 edition.