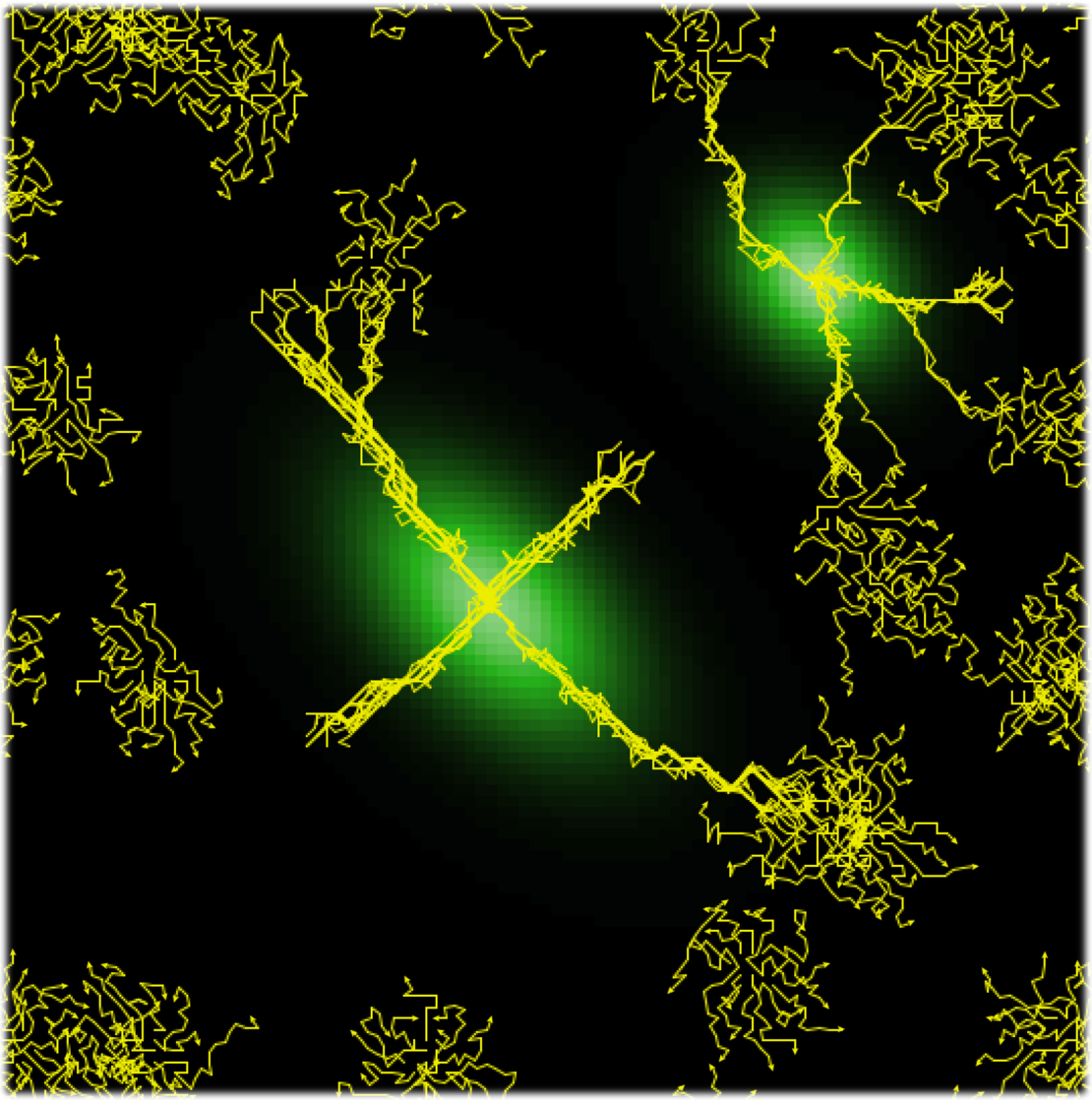
Creativity *versus* Conservatism in Science: A Simulation-Based Investigation



Salvador Gouveia
Brasenose College and Faculty of Philosophy
University of Oxford

A thesis submitted in partial fulfilment
of the requirements for the degree of
BPhil in Philosophy
2020-2022 cohort

Creativity *versus* conservatism in science: a simulation-based investigation[1][2]



[1] To my parents, and to their parents: this is a thesis about diversity, and, together, you could not have been better inspiration.

[2] I thank my supervisor Prof Philipp Koralus for his indispensable advice at all stages of this project. I also thank Dr Nicolas Payette, who provided crucial support on technical issues regarding the programming aspect of this thesis.

Abstract

Scientists and the communities which they integrate may display stronger or weaker tendencies to conform to certain ways of approaching the research of a given scientific topic. Some have claimed that the level of conformity displayed has important epistemic implications. In particular, it has been argued that scientific nonconformity is beneficial for the progress of science, as it is conducive to the adoption of a diversity of scientific approaches, some of which may be truth-conducive and would not be attempted in more conformist settings.

In this thesis, I refine and operationalise an account introduced by Currie(2019b) which frames conformity and nonconformity in science as conservatism and creativity, respectively. In particular, I build an agent-based model from this framework, and I use it to run computer simulations and test whether scientific creativity really promotes scientific progress. Ultimately, by combining the results of these simulations with considerations about the institutions of science, I conclude that in order for a scientific community to reap the epistemic benefits often associated with scientific creativity, its scientists need not (and probably should not) be particularly creative.

Table of Contents

Chapter 1. Introduction *5*

Chapter 2. Conformity in science *11*

2.1. The Research Diversity Problem **11**

2.2. The social epistemology of the Research Diversity Problem **14**

2.3. Creativity and conservatism **16**

2.3.1. Why Currie 16

2.3.2. Nonconformity as creativity; conformity as conservatism 19

2.4. What is ahead **26**

Chapter 3. Modelling creativity and conservatism *28*

3.1. Agent-based models and epistemic landscapes **29**

3.1.1. Agent-based models of science 29

3.1.2. Epistemic landscapes 32

3.2. Currie's framework as an epistemic landscape ABM **36**

3.2.1. The landscape 37

3.2.2. The agents 40

3.3. Creativity in the model **45**

3.3.1. Agents' movement and individual-level creativity 46

3.3.2. Agents' dispersion and community-level creativity 55

3.4. Cognitive diversity in epistemic landscape ABMs **59**

Chapter 4. Simulating creativity *64*

4.1. Does science need creativity? **64**

4.2. Simulations, simulations, simulations **70**

4.2.1. The basic model 72

4.2.2. Experiments and results 76

Chapter 5. Discussion, limitations, and ways forward *98*

5.1. Main lessons **98**

5.2. Towards an optimal arrangement – institutional reality check **102**

5.2.1. Why pockets 103

5.2.2. Why conservatism 104

5.2.3. Anti-conformity arguments revisited 108

5.3. Further reflections **114**

Chapter 6. Conclusion *118*

Bibliography *120*

Chapter 1. Introduction

Robert H. Goddard (1882-1945) is an icon at the National Aeronautics and Space Administration (NASA). NASA's Goddard Space Flight Center as well as its Goddard Institute for Space Studies (Garner 2015) are clear illustrations of Robert Goddard's prominence as a historical figure at NASA. This is no surprise. Goddard was a physicist and engineer whose work in the early decades of the 20th century laid the foundations of modern rocketry, which eventually enabled the success of NASA's *Apollo 11* moon-landing mission in 1969. However, the good light in which Goddard's work is seen today does not reflect the reception it initially got among the academic community and the wider public. Indeed, Goddard was a bit of a solitary, maverick scientist whose theories and inventions did not quite convince the wider academic community. His ideas about the possibility of a rocket flying to the moon (and the associated idea of the possibility of liquid-fuelled rocket propulsion *in vacuum*) were often met with incredulity and ridicule. *The New York Times* even issued an editorial in 1920 mocking Goddard and suggesting that his ideas – in particular, his thesis that propulsion in vacuum was possible – revealed that Goddard lacked 'the knowledge ladled out daily in high schools' (New York Times 1920). This lack of credulity in the engineer's bold theories did not make his life as a scientist easy, and indeed forced Goddard to fund much of his research out of his own pocket. Goddard's ideas were, on the whole, of course, successful. His out-of-the-ordinary theories and inventions were behind much of the later science of rocket propulsion which enabled the success of *Apollo 11*. So compelling was Goddard's work's contribution to the success of this mission that *The*

New York Times redacted its 1920 editorial days after the moon landing, even though Goddard did not live to see it[3].

While Goddard's story is still one of relative success (he is nowadays considered a scientific hero after all), careers such as his are sometimes used to argue that science and its institutions stifle daring innovation and creativity[4], with several explanations in philosophy of science being given to account for this phenomenon (e.g. Stanford 2019). This brings to the fore an interesting problem in the philosophy and social epistemology of science, which I will call the Research Diversity Problem, concerning the question of which ways of approaching the research of a given scientific topic are most effective at discovering the relevant facts about that topic. The Research Diversity Problem has mostly arisen in the philosophy of science literature from the realisation that scientists may adopt different types of research strategy when working on a research topic – they can be more like Goddard, or more like the conformist lab researcher which universities and other research centres are (some claim, excessively) replete with. For now, let me group each of these different strategies into one of two types – respectively, 'nonconformist' or 'conformist'. On the one hand, scientists may adopt the kind of research type which Goddard exemplifies: an audacious way of doing science, which is novel, disruptive, innovative, creative, often adopting unorthodox approaches, new methods and theories which are substantially different from what has been practised before. This type of research strategy tries to achieve scientific

[3] See Clary (2003) for a biography of Robert Goddard focused on his scientific achievements.
[4] For a similar, more recent case see, for example, Haney (2019) for a recount of Jim Allison's career. Allison's ideas about the relationship between immune systems and cancer development, much like Goddard's, were ignored, ridiculed, and considered outlandish throughout much of his career before they earned him the 2018 Nobel Prize in Medicine.

progress via 'science that bucks establishment trends and attempts ambitious projects' (O'Connor 2019: 25) such as the ones achieved by Goddard – this is the nonconformist type. At the other end there is a way of pursuing scientific questions that is cautious, more modest, follows closely from the theories and practices which had been practised before, solving puzzles (to put it in Kuhn's terminology) of established research programmes, taking 'small bites, [moving] along well worn pathways, and [following] establishment rules' (O'Connor 2019: 25) – this is science of the conformist type.

In this essay, I deep dive into the Research Diversity Problem from a social epistemological approach. Specifically, I will build an agent-based model of science and run computer simulations of this model which will allow me to explore the potential epistemic benefits of nonconformity in science. Relatedly, I will combine the results of these simulations with considerations about the institutions of science to suggest what a scientific community's epistemically optimal socio-epistemic arrangement would look like in terms of creativity. To give a preview, I will argue that nonconformity may indeed be beneficial for the progress of science. However, my investigation suggests a more nuanced account of the epistemic benefits of creative, nonconformist science than what is usually presented in the literature. It turns out that nonconformity at the community level – which I operationalise, roughly, as the tendency of a scientific community to adopt a diversity of approaches to a given scientific topic – generally contributes significantly to the progress of science. However, this can be achieved even when most or even all the scientists of that community are individually conformist. In fact, the results suggest that individual nonconformity – roughly, the tendency of a scientist to try significantly different and new approaches from what

had been tried before – may often be detrimental to the progress of the scientific understanding of a research topic. Thus, it is not at all clear that nonconformity in the more traditional, individual sense should be promoted for the benefit of science. In this sense, my thesis will depart from previous literature in its attention to the distinction between individual- and community-level creativity and, based on that, in its claim that the potential epistemic benefits brought about by scientific nonconformity may be most effectively achieved when individual scientists themselves display conformist tendencies.

The thesis is structured as follows. In Chapter 2, I present the general motivation for the investigation of the thesis and how the Research Diversity Problem arises in the social epistemology of science. I also present a framework developed by Currie (2019b) which accounts for scientific conformity and nonconformity as scientific conservatism and scientific creativity, respectively.

This account becomes relevant because, in Chapter 3, I use it as the basis on which I develop my agent-based model. The model-building aspect of this thesis will allow me to appraise Currie's framework and suggest ways in which it can be refined, but also why it is a particularly helpful framework to think about nonconformity and conformity in science. In particular, I highlight how such a framework allows one to explicitly disambiguate between two possible readings of the conformity/nonconformity distinction, one which focuses on the individual level, and one which focuses on the group level. That is, one may think of conformity in science as an individual characteristic – something which describes how a certain individual scientist or

research team behaves –, or as a group-level characteristic, describing a feature of a certain scientific community as a whole. These two notions are not mutually incompatible, but, as I will argue, they may come apart, in the sense that there may be individual-level conformity without group-level conformity, and *vice versa*. Indeed, one relatively neglected aspect of the Research Diversity Problem has precisely been the relationship between individual- and community-level scientific (non)conformity. With this gap in mind, I will develop my model in order to explore how scientific conformity and nonconformity are related to scientific progress. While considerable research has been devoted to studying the degree to which epistemic outcomes are sensitive to variations in *individual*-level scientific conformity (see, e.g., Weisberg and Muldoon 2009; Thoma 2015; Pöyhönen 2017), far less attention has been given to the potential epistemic benefits of nonconformity as conceived at the group level, partly because such a distinction has been, as I have mentioned, obfuscated. My model overcomes such a limitation, and it allows me to consider both notions of nonconformity, their interaction, and their epistemic effects. I end the chapter by highlighting a further advantage of using Currie's framework as the basis of my model, which is that it reveals how the topic of nonconformity in science bears on current debates in philosophy of science regarding diversity and the cognitive division of labour in science.

In Chapter 4, after considering the details of some main arguments in favour of scientific nonconformity and how scientific institutions stifle creativity, I run a series of experiments which allow me to explore the link between nonconformity and scientific progress.

In Chapter 5, alongside with some methodological reflections, I highlight the implications of these results in conjunction with considerations about the institutions and incentive-structure of science. I draw the main conclusions and suggest that an optimal socio-epistemic arrangement approximates what I call a 'pockets of conservatism' model of a scientific community – i.e. one in which scientists behave conservatively, but which is such that community-level creativity still obtains by having different subcommunities of scientists, each of which exploring a fairly distinct set of approaches from those explored by the other subcommunities.

Chapter 6 concludes the thesis.

Chapter 2. Conformity in science

2.1. The Research Diversity Problem

As mentioned in the Introduction, social scientists and philosophers of science have explored the idea that strategies to study scientific questions may be distinguished based on the extent to which they are instances of more conformist, or, by contrast, more nonconformist science[5]. I should note that I have been deliberately vague in defining what I mean by conformist and nonconformist science. That is because there is actually no consensus on this matter, and it is one of my aims in this essay to put forward a conception which makes a more precise and systematic sense of this distinction. For now, though, it is sufficient to appeal to the intuitive idea that the kind of research which individual scientists as well as wider scientific communities adopt in order to study a given research topic may vary. In particular, it may vary in terms of the extent to which it converges to a given set of relatively well-explored practices, background theories, methodologies. In other words, research may be more or less conformist. This gives rise to what I call the Research Diversity Problem. To put it concisely, the Research Diversity Problem concerns which research arrangements (more or less conformist) are most likely to promote truth-finding in science, as well as which (if any) institutional setups in science are compatible with and conducive to those research arrangements. It concerns diversity of research in a variety of senses, which will be specified later. Intuitively, though, the idea is that there is a diversity of ways in which a scientific topic may be studied, and the challenge is to understand

[5] For allusions to this distinction (some more explicit than others), see, e.g., Hull (1988: Ch.12); Brezis (2007); Weisberg and Muldoon (2009); Luukkonen (2012); Muldoon (2013); Alexander et al. (2015); Thoma (2015); Kummerfeld and Zollman (2016); Avin (2019); Bedessem (2019); Currie (2019a); Currie (2019b); Harnagel (2019); Heesen (2019); O'Connor (2019); Stanford (2019); Avin (2020); Bedessem (2021).

which of these ways are preferable, as well as to understand how these preferable ways may be achieved. The most obvious example of this is that scientists may be more or less aligned with the scientific *status quo*, so to speak. That is, they may be more or less conformist in their individual strategies to study a certain research topic; they may be more or less like Goddard. Scientific communities, too, can be considered more or less conformist, as they may be more or less inclined to converge on a narrow set of approaches to study a given topic.

However, there is a prior question raised by the Research Diversity Problem, which concerns exactly how to cash out the distinction between conformity and nonconformity in science. Indeed, there is some variability in how this distinction has been explicitly drawn in the philosophy of science. For example, some authors draw a distinction between 'high risk, high reward/impact' science on the one hand, and low risk, low reward/impact science (Hull 1988; Heesen 2019; O'Connor 2019). Some align the distinction with a trade-off between 'exploration' and 'exploitation' of theories, methods, models, etc. (Polanyi 2000; Alexander 2013; De Langhe 2014a; De Langhe 2014b; Kummerfeld and Zollman 2016; Kaaronen 2018; Avin 2020). Some oppose 'ground-breaking and potentially controversial' research to 'uncontroversial' research (Luukkonen 2012). Some also put it in terms of followers versus mavericks (Weisberg and Muldoon 2009), or, similarly, as extractors versus explorers (Thoma 2015). Some cash it out in terms of a distinction between 'innovation' and 'conformity' in science (Brezis 2007; Luukkonen 2012), and others also find parallels between the conformist-nonconformist distinction and the Kuhnian (Kuhn 1962) distinction between

revolutionary and normal science (Luukkonen 2012; Heesen 2019; O'Connor 2019; Bedessem 2021).

These ways of cashing out the distinction are not necessarily mutually exclusive, and my goal is not to proclaim one specific formulation as the most appropriate way to frame the Research Diversity Problem. Instead, in order to make progress beyond the definitional issue, I will develop a particular framework initially presented by Currie (2019b) to address this question. This will allow me to explore the Research Diversity Problem and especially put to test arguments which one may call *anti-conformity arguments*. Such arguments claim that nonconformity in science (or at least some degree of nonconformity in science) is epistemically beneficial. The idea here is that the study of a given research topic benefits, epistemically speaking, from the exploration of a diversity of approaches, rather than the investigation of only a narrow set of similar ways of investigating that topic[6]. One straightforward interpretation of this is that some degree of nonconformity in science allows for the discovery of relevant facts and phenomena which would otherwise be left uncovered, or which (all else equal) would require much more effort and resources in order to be discovered. Conformity (or excessive conformity), the argument goes, is therefore epistemically inefficient.

If one such anti-conformity argument is combined with the claim that the scientific *status quo* does not produce nonconformist science (or does not produce conformist science to high enough a degree), then one may conclude that science is *too*

[6] As I have mentioned, a more precise account of what nonconformity amounts to will be given later.

conformist relative to the ideal that is implied by the *anti-conformity arguments*. From this conclusion, one may further derive the normative upshot that the institutions of science should (if possible) be adapted in order to accommodate and promote more nonconformist research. Later on, I will provide examples of such arguments. For now, I want to make clearer why this is an interesting topic from a *social* epistemological point of view.

2.2. The social epistemology of the Research Diversity Problem

The case for studying the Research Diversity Problem from a social epistemological perspective is hopefully relatively straightforward. Still, it is worth being clear about the ways in which this social epistemological approach takes shape.

Firstly, this is partly an investigation into scientists' strategies to approach the study of a given topic and the *epistemic* consequences of varying such strategies. This highlights the epistemological nature of my investigation. But it is also relevant that scientists operate within a *social* structure. Scientific knowledge is a 'system property' (Pöyhönen 2017: 4521), which results not only from individual scientific work, but also from community-level conditions, such as how the scientific community is organised (Longino 2002). As such, it is crucial to acknowledge that scientists are social learners (Weisberg and Muldoon 2009), they learn not only from their own experiments and investigations, but also from those of other scientists who they communicate with. An investigation into the epistemic consequences of adopting different research types

should thus be attentive to the social nature of scientific research[7]. For example, the degree to which conformist scientists are aware of nonconformist scientists' results (and *vice versa*) may affect the ideal distribution of research types across a scientific community (Thoma 2015). With this in mind, conformity is intuitively seen as a social phenomenon. When scientists conform to a set of practices and methods, for example, they may do it because other scientists adopt precisely such practices.

I also mentioned considerations regarding the institutions of science, and how different institutional arrangements may be more or less conducive to conformist behaviour in science. This is another way in which my investigation fits squarely within a social epistemological approach. Philosophers and sociologists of science have long recognised that the institutions of science and the incentives to which scientists are subject at least partially affect the questions they deem worthy of being pursued, the methods they adopt, etc. The institutional setup thus plays an important role in determining epistemic outcomes in science[8]. As such, an investigation into how more

[7] For work on the epistemic consequences of the social and communicative structure of science see, for example, Lazer and Friedman (2007); Zollman (2007); Zollman (2010); Grim et al. (2013); Mayo-Wilson et al. (2013); Mayo-Wilson (2014); Rosenstock et al. (2017); Baumgaertner 2018; De Langhe (2018); O'Connor and Weatherall (2018); Santana (2021); Weatherall and O'Connor (2021); Fazelpour and Steel (2022). Many of these works are based on a model introduced by Bala and Goyal (1998), and imported to philosophy by Zollman (2007), which frames communities as networks of communicating individuals. See Zollman (2013) for a (slightly outdated) review.

[8] There is a vast literature on the role that incentives and institutions (and, more broadly, instrumental rationality) play in determining epistemic outcomes in science. See Merton (1957) for a pioneering sociological approach. Some prominent topics explored from this perspective have been the division of cognitive labour in science (e.g. Kitcher 1990; Kitcher 1993; Strevens 2003; Strevens 2013); peer review and its effects on the production of scientific knowledge (Heesen and Bright 2020; Santana 2022); the role of credit incentives and the desire for credit for one's scientific work (Zollman 2018); the institutional incentives to commit scientific fraud (Bruner 2013; Bright 2017; Zollman 2019; Heesen 2021); the effects of funding arrangements on evidence-based public policy outcomes (Holman and Bruner 2017; Holman and Elliot 2018; Weatherall et al. 2020); the replication crisis (Romero 2017; Heesen 2018; Romero 2020); and, of course, conformity in science (Currie 2019b; Harnagel 2019; Heesen 2019; Stanford 2019; Avin 2019; Avin 2020).

and less conformist ways of doing research differ in terms of their epistemic consequences should be sensitive to such institutional factors.

In this thesis, then, I will attend to the social aspects of science and place them at the fore of my investigation.

2.3. Creativity and conservatism

2.3.1. Why Currie

I will now make the discussion about the contrast between conformist and nonconformist science more precise. I will work on a framework introduced by Currie (2019b), which I call the 'nonconformity as creativity' account, as it frames nonconformity as scientific creativity and conformity as scientific conservatism. As it will become clear throughout the thesis, Currie's framework is particularly useful for my purposes for several reasons.

Firstly, Currie's framework lends itself to being productively studied via what Currie calls the 'economic approach' to social epistemology of science. This 'economic approach' is a broad term used to refer to formal ways of studying the behaviour of scientific individuals and communities, by taking into account the factors mentioned in the previous section, such as the set of incentives to which scientists are subject, or the structure of the community which these scientists populate. Such an approach is advantageous in allowing one to model scientific communities in ways that enable one to isolate particular features of certain communities (such as scientific ones) and

investigate their epistemic consequences in a precise and systematic manner[9]. In this case, such models will be useful in studying clearly and precisely the epistemic aspects surrounding conformity and nonconformity in science. To be more specific still, I will show that Currie's framework is naturally translatable into an agent-based model, and that this model offers a particularly promising way of exploring the intuitions about the Research Diversity Problem and the associated notions of conformity and nonconformity in science, and of testing specific claims regarding the benefits or harms of scientific nonconformity. More generally, the fact that Currie's framework is easily adaptable into such a model allows me to explore the normative question of how a scientific community *should* approach the study of a certain scientific topic.

A second main reason for the appeal of Currie's account is that it provides conceptual tools which allow one to think of nonconformity at the individual and group levels as different, potentially independent phenomena. Previous accounts of nonconformity in science have focused almost exclusively on the individual level. Currie's framework, on the other hand, will allow me to make clear why it makes sense to conceive of community-level creativity and individual-level creativity as sometimes potentially coming apart. This feature is especially pertinent if one's focus is a social epistemological one. At the same time, the framework allows me to highlight the features that these phenomena at different levels have in common, which justify

[9] I will say more about the methodological advantages of employing formal models to studying philosophical questions in the next chapter, in which I will use Currie's relatively informal framework to build a formal model of a scientific community.

subsuming both phenomena under the same conceptual distinction of scientific creativity versus scientific conservatism.

A third advantageous aspect derives from the first two advantages. In short, it is that Currie's framework, once translated into a formal model, allows one to draw a natural link between the creativity/conservatism contrast and two types of research diversity commonly treated in the social epistemology of science – these two types are, in Thoma's (2015: 456) terms, 'diversity of research strategy' and 'diversity of type of research'. Previous treatments of this issue have either been ambiguous about which kind of diversity they referred to, or, while acknowledging that they are distinct phenomena, they have failed to recognise what links both types of diversity. Currie's framework, in allowing one to distinguish between creativity at the individual and community levels, brings conceptual clarity in allowing one to disambiguate between both notions of diversity, while also allowing one to highlight what there is in common between them. This will become clear in Chapter 3, after I have introduced the model I will build on the basis of Currie's framework.

Despite these advantages, it should be clear that I will not take Currie's framework for granted. On the contrary, the model which I will build from my appreciation of Currie's framework will also allow me to appraise it and identify its strengths and limitations more precisely. As such, while my goal is admittedly not to elect the best account of scientific conformity and nonconformity, I will not take Currie's framework uncritically either. Rather, I will use it to advance in the exploration of the Research Diversity Problem, while reflecting on the potential limitations of the framework itself.

2.3.2. Nonconformity as creativity; conformity as conservatism

Let me, then, introduce Currie's (2019b) framework to understand conformity in science and its epistemic implications. Currie frames the discussion about different types of research in science in terms of scientific creativity and scientific conservatism. Conformity is the expression of scientific conservatism; nonconformity is the expression of creativity. In what sense are creativity and conservatism at the root of nonconformity and conformity, respectively? Take the case of an individual scientist (say, for familiarity, Robert Goddard). Creativity is expressed in the extent to which that scientist is willing to try out different methods, conceive of alternative theories (Stanford 2019), and be relatively unconstrained by the theories and methods which one has applied and studied before. Conservatism, on the other hand, translates into caution, into an unwillingness to expand too much the boundaries of what one conceives of as promising science, into a commitment to a relatively stable set of theories and methods as the ones that should be tried, based on what one already knows.

Currie (2019b) aims to capture this idea via a metaphor of what he calls 'hot' and 'cold' searches within of a given solution space. He says:

> 'These terms are inspired from thermodynamics, and refer, metaphorically, to the kinetic energy of a molecule (a scientist) or a collection of molecules (a research community). A hot molecule, with plenty of stored energy, will move

through a space in erratic bounds. A cold molecule will move more slowly. How the metaphor plays out should become obvious' (Currie 2019b: 40).

Lest the metaphor be not so obvious after all, it is worth saying a bit more about this. We can think of the solution space for a given research topic as the set of potential solutions or approaches to a given problem of scientific relevance. Some solutions will work, others will fail. Not all scientists were successful in advancing the field of genetics, for example. Or, to give a more fine-grained example, not all attempts to understand the structure of the DNA molecule were successful. Some solutions will indeed work very well, as they will constitute great advances in the understanding of a certain scientific field of research. Others will be totally fruitless.

The talk of a solution *space* also suggests some spatial distribution of the solutions along a certain number of dimensions. As such, a certain solution can be more or less distant from another, different solution. This reflects how similar those attempted solutions are to each other in terms of the dimensions that are being considered. Currie (2019b), however, is not very clear on what these dimensions should represent. That is a gap which I will fill in the next chapter, in which I will build a specific model from Currie's framework.

In order to get a more concrete idea of what this framework captures about scientific research, consider Goddard's example once more. Using the cold and hot searches metaphor, one possible account of his maverick approach is precisely that his research strategy is rather creative in the sense just specified. He 'jumped' across the solution

space, trying out different theories and hypotheses, experimenting with different methods and materials, looking for new ways of learning about his topic of research.

In any case, what I have described so far is an account of *individual* creativity and conservatism. In studying how predisposed to try out new and different theories, methods, etc., a given scientists is, one is studying creativity and conservatism as individual characteristics. This is certainly an important aspect of the Research Diversity Problem, in the sense that it identifies individual creativity as an axis of diversity, but it would be rather limiting to study creativity only on that level. Ultimately, and especially in the context of a social epistemological investigation, one would want to understand community-level behaviour and its epistemic outcomes. Of course, such outcomes may arise from individual characteristics (such as each individual's scientific creativity). In fact, perhaps the reason why there has been an overwhelming focus on scientific creativity and nonconformity at the individual level is precisely that it has been thought to be a major, if not the sole, source of creativity at the global level. But it is not at all obvious that that is the case. The socially relevant notion of creativity might well arise independently of individual creativity. In order to see this, I need to be precise about what I mean by a 'socially relevant' notion of creativity. Indeed, Currie's (2019b) framework provides the conceptual tools to adopt an approach which is also sensitive to group-level creativity.

Currie (2019b: 40) characterises creativity at the *community* (or population or group) level in the following way:

> ‘A population of agents' creativity is proportional to the likelihood that those agents will explore a wide solution-space’.

This definition reveals that conservatism at the individual level does not necessarily result in conservatism at the community level. Nor is a community overwhelmingly made up of individually creative scientists necessarily creative.

In order to see this, note that a scientific community made up of conservative agents may end up being creative, for while each individual scientist explores a rather narrow sub-space of the solution-space, the aggregate result may be one in which the whole of the solution space is widely explored – scientists’ attempted solutions to a given research topic need only be widely dispersed across the solution space. Conversely, a scientific community made up of individually creative agents may end up not being creative, if, for example, agents do not register, share, or remember the results of previous investigations (Currie 2019b: 41). That is because, if previously attempted solutions are not registered, then cumulative learning is impeded and the chances of exploring a solution-space for a given problem widely are naturally limited.

This disambiguation is rather helpful. Whereas previous treatments of the Research Diversity Problem have explored the idea that individually creative scientists may bring about beneficial epistemic outcomes, they have typically ignored that the benefits arising from the creative exploration of a given scientific research topic by a whole community may be achieved even when no *individual* scientist behaves in a particularly creative fashion. In fact, all things considered, it may well be that the

optimal socio-epistemic arrangement is one in which individual scientists behave conservatively – that is, they employ the kind of 'cold' research strategies described above – but in which scientists explore solutions that are widely dispersed across the solution-space. In other words, the optimal socio-epistemic arrangement may be one which is conservative at the individual level but still creative at the community-level[10]. My model will allow me to explore precisely whether this is a plausible hypothesis, and that is one reason why the translation of this framework into a formal model will be valuable.

To be clear, the conservative-creative distinction is not really a binary in Currie's framework. An individual agent may explore the solutions to a given problem within a really narrow space, trying out only small variations of previously attempted solutions and working on a rather stable methodological and theoretical framework; or she may be really erratic, coming up with unprecedented approaches which have very little in common with past attempts to work on a given scientific question; or she might be somewhere in the middle. The same applies to community-level conservatism and creativity: a scientific population may be highly concentrated around a narrow set of similar approaches to a scientific problem, or its scientists may be highly dispersed across the space; or they might be dispersed to some moderate extent. While it is useful to keep this in mind, I should disclose that I will work with a rather sharp

[10] Indeed, as Godfrey-Smith (2008: 142) puts it, '[w]e have become used to the idea that a community or population can embody epistemic properties that no individual has' – see, for example, Bishop (2005); Mayo-Wilson et al. (2011); Levy and Alfano (2020).

distinction between the two types of research at the individual level, as a workable idealisation of the actual scientific state of affairs.

One might worry about whether creativity and conservatism really capture nonconformity and conformity, respectively. In fact, Currie's framework is not unprecedented. Creativity and conservatism are often opposed concepts in a variety of disciplines, from information technology to psychology (Dollinger 2007). And, indeed, it is not rare to encounter such an association between creativity and nonconformity on the one hand and conservatism and conformity on the other in the sociology and social epistemology of science literature (see, e.g., Loye 2007; Popova et al. 2017; Avin 2019; Schneider 2021). I hope to have conveyed the intuitive appeal of such an association. Conservatism captures a tendency to 'take small bites', tweaking one's methods slowly and gradually, relying on what is already known as guidance for one's own research, rather than adventuring oneself in the unknown. This captures a tendency to *conform* to specific methods, theories, approaches, paradigms. Creativity, on the other hand, translates into the ability to imagine alternative ways in which the world could be, to devise innovative ways to explore the world, to break away from established norms and traditional practice. In this sense, creativity is a measure of nonconformity.

Admittedly, this way of pumping the intuition in favour of the adequateness of Currie's framework for capturing the conformity-nonconformity distinction relies heavily on the individual level. That is, in order to convince you that nonconformity and conformity are well captured, respectively, by creativity and conservatism, I have relied on the

individual-level a lot. One question that comes up, then, is why one should care about group-level creativity when thinking about the issue of conformity in science. Perhaps these are very different phenomena, and to subsume both under the common label of 'creativity' is a mere terminological trick to bring together two notions that have very little in common indeed. In sum, one might wonder what group-level creativity actually has to do with scientific conformity.

But, as I have mentioned, one of Currie's framework's appealing features is precisely that it allows me to treat individual- and group-level creativity separately. So it is not particularly problematic that two different phenomena have been identified. The challenge here is to show that these two phenomena have enough in common that they can indeed be considered instances of the same general phenomenon of nonconformity in science. The way in which these two notions are closely related becomes clearer when ones thinks about this framework in the context of the formal model which I will present in the next chapter. But, merely relying on the intuitions presented in this chapter, notice that both individual- and community-level creativity capture the tendency to explore a 'wide solution-space'. At the individual-level, that means jumping across the solution-space, attempting different solutions; at the community-level, that means having a community of scientists spread across a sufficiently diverse set of solutions within the solution-space.

The criterion for creativity – a wide exploration of the solution-space – in turn captures relevant aspects of nonconformity. We have seen how that is so in the individual case. In the case of community-level, creativity also seems to be a natural candidate for a

measure of nonconformity. To say that a community of scientists is nonconformist is to say that it does not converge on a narrow set of attempted solutions to a given problem. Creativity in the sense proposed by Currie (2019b: 40) captures this well. Similarly, a conservative community will tend to focus on a relatively narrow and uniform subset of the solution-space, and this too captures the dynamic of conformity in science.

While some elements of conformity and nonconformity may be left out of this framework, it seems to me that it captures crucial features of the distinction which any plausible account of scientific conformity should attend to. In the next chapter, I will suggest a way in which the notions of scientific creativity and conservatism can be operationalised in an even more complete and intuitive way. However, on approximation I would claim that Currie's framework does rather well. Hence, while it is not this thesis' goal to elect the best way to account for scientific conformity, Currie's framework's simplicity combined with the fact that it captures central elements of the distinction rather intuitively renders it a promising basis on which to further investigate the Research Diversity Problem.

2.4. What is ahead

I am now well positioned to advance my investigation. As I have pointed out, the realisation that scientists and communities of scientists may be more or less conformist (or, to use the terms of the framework which I have adopted, more or less conservative) raises the question of what the epistemically optimal way for a scientific community to study a particular research topic is, and which institutional

arrangements are compatible with and conducive to it. I have referred to this as the Research Diversity Problem, because there is a variety of research arrangements (more creative; less creative) which individual scientists as well as communities may employ in order to study a given scientific topic, but not all such arrangements may be equally productive in their goal of uncovering significant facts about such topic. Perhaps an optimal socio-epistemic arrangement is, as some suggest, one where the kind of work which Goddard presumably practised – i.e., unorthodox, risky, daring, etc. (in other words, nonconformist) – is promoted to a great extent; or maybe, actually, survivorship bias is at play here and Goddard was just lucky that his individual creativity led him to successful approaches – certainly his discoveries were important, but, on the whole, it may be argued, science is better off when more resources are put into a more cautious, conservative way of conducting research, where individual scientists do not waste resources attempting outlandish methods that only very rarely may be successful. On top of these questions, there is another factor which further complicates things, which is the fact that the potential benefits of individual creativity may be more efficiently achieved via other routes. In particular, I have noted that group-level creativity may still be achieved without individual creativity. This interaction between individual- and community-level creativity in determining epistemic outcomes will be a central theme of my investigation into the Research Diversity Problem. What socio-epistemic arrangements (with respect to both individual- as well as group-level creativity) most efficiently promote scientific discovery? In the next chapter I will present a model which will allow me to explore this question.

Chapter 3. Modelling creativity and conservatism

As I have mentioned, one advantage of working on the basis of Currie's account of scientific nonconformity and conformity to think about the Research Diversity Problem is that it enables one to see how individual- and community-level scientific nonconformity may be independent of each other. But, as I have flagged in the previous chapter, another major advantage of working on the basis of Currie's proposed framework is that it translates naturally into a formal model – in particular, an agent-based model. There are a number of potential methodological advantages that come from the use of models in philosophy. For this essay, one good reason to use models is that social systems such as scientific communities are highly complex and the individual interactions between members of each community intractable (Martini and Pinto 2017). Furthermore, model-building allows one to make precise and explicit the assumptions that one makes about the target-system, and isolate these from potentially confounding factors, and thus find what conclusions follow from such assumptions, and, importantly, what conclusions do *not* follow (Williamson 2017). The use of models also helps raise awareness of and work out the different ways in which one might conceive of the problem at hand, and in that way try and figure out what the most promising and adequate ways of thinking about the problem are.

In this chapter, then, I will try and realise the potential benefits that come with building a model of a scientific community to formally explore the Research Diversity Problem. In order to do that, I will first introduce kind of model that I will build – an epistemic landscape agent-based model. I will then review some of the relevant literature that

has used such models to investigate the Research Diversity Problem, so that I am able to then set up my own model and highlight what it will add to the discussion.

3.1. Agent-based models and epistemic landscapes

3.1.1. Agent-based models of science

Agent-based models (ABMs) are a method of growing prominence in social epistemology, especially as applied to the epistemology of scientific communities[11]. Generally, ABMs model systems of communities made up of individual decision-makers (the agents), which follow certain decision rules that may differ from those of other individuals. ABMs also allow one to explore the interactions between the different agents, as well as interactions between agents and the environment which they populate. As Dilaver et al. (2018: 1134) put it, ABMs allow researchers to model 'heterogeneous agents that can respond to internal and local as well as aggregate variables' and thus 'simulate adaptive behaviour, inter-dependent decision making, spatial patterns and social networks in a broad range of contexts'.

Agent-based models are predominantly built as computational models (see Weisberg 2012: Ch. 2). This allows harnessing the computational power of computers to capture dynamics that would be intractable if one only used pure analytical mathematical modelling (Axelrod 1997). This, in turn, permits inferring the macro-level outcomes that emerge from such interactions rather than follow directly from specific assumptions in the model (De Marchi and Page 2014). Still, what is distinctive about

[11] For some reviews of the literature employing ABMs in philosophy and social epistemology of science, see, for example, Payette (2012); Martini and Pinto (2017); Klein et al. (2018); Pinto and Pinto (2018); or Šešelja (2021).

ABMs is *not* that they are computational models – that need not be the case (Bonbeau 2002: 7210). Rather, what is special about these models is that they rely on modelling at the micro-level (what Bonbeau 2002: 7210 calls 'microscopic modelling') to make inferences about the macro-level phenomena (Kuhlmann 2021), the latter usually being the population behavioural patterns that emerge from the interactions at the individual (agent-based) level. The fact that ABMs allow for such heterogeneous microscopic modelling often allows these models to better capture the complexity of their target systems (Gilbert 2007). Not only do ABMs potentially offer the possibility of more realistic modelling, but they also allow, via computer simulations and the statistical or visual processing of data arising from such simulations, to deal with information which, as mentioned, would be intractable from an analytical approach (Axelrod 1997). Partially for that reason, the macro-level results that emerge from ABMs are often puzzling if one focuses only on individual behaviour and incentives. A famous example is Schelling's (1971) finding that strong levels of residential segregation can emerge in populations whose members have only a 'mild' preference towards living near other members of their own group.

Scientific research – being 'a form of coordinated social behavior arising from the cognitive mechanisms and interpersonal processes involved' (Reijula and Kuorikoski 2019: 244) – is thus a good candidate for being fruitfully modelled via ABMs. Moreover, the fact that I am partially concerned with how the (macro) phenomenon of community-level creativity relates to individual-level creativity renders the ability of ABMs to capture emergent phenomena especially interesting. Furthermore, the fact that agents (typically conceived of as individual scientists or individual teams of scientists) are the unit of modelling significance in ABMs allows me to easily model

heterogeneous populations of scientists and study the epistemic dynamics that arise from their interaction from computer simulations. This is a step away from more traditional approaches to formal social epistemology of science (of which Kitcher 1990 or Strevens 2003 are paradigmatic examples), which rely on analytical solutions to mathematical problems, often assuming an ideally rational representative scientist. Not only do ABMs allow departing from such assumptions, but they also enable focusing more on the epistemic dynamics generated by the interaction between different agents and how that can form new incentives in response to which scientists may adapt their behaviour (Kuhlmann 2021).

For example, such ABMs of science have been used to study the dynamics of agreement and disagreement and opinion diversity in science (e.g. Zollman 2010; Santana 2021), of scientific journal publication strategies (e.g., Zollman 2009), of scientific collaboration (e.g. Zamzami and Schiffauerova 2017) or of methodological spread across scientific disciplines (e.g., Smaldino and O'Connor 2020). Another prominent topic in the social epistemology of science which has been studied via ABMs is much more closely related to my own project in this thesis. These models investigate what is commonly referred to as the division of cognitive labour (e.g. Weisberg and Muldoon 2009). Such models explore the aggregate epistemic effects of different individual research strategies and of different combinations of such strategies across a scientific community. This explains the use of the term 'division of cognitive labour' – different types of research strategies may be divided across the members of a scientific community differently[12]. Such models, because they are so intimately related

[12] Although, as I will note, such a term has been ambiguously used in the literature.

with my own purposes in this essay, will be highly instructive as guidance and term of comparison for my own investigation. As such, it is now pertinent to introduce the general framework of such ABMs in greater detail.

3.1.2. Epistemic landscapes

Epistemic landscape ABMs were adapted from evolutionary biology[13] and originally introduced in social epistemology of science by Weisberg and Muldoon (2009)[14]. In this subsection, I will focus mainly on the structural aspects of epistemic landscape ABMs, leaving the interpretation of the model to the next section.

As the name suggests, epistemic landscape ABMs rely on the notion of a 'landscape'. The landscape consists in a square grid made up of a certain (whole) number of equally sized patches. Thus, if, for example, the side of the grid is 10 patches long, the grid will be made up of 100 homogenously distributed patches (i.e. a 10 by 10 grid). Each patch is therefore identifiable by a pair of coordinates, (x, y), which represent, respectively, the position of the patch along the x-axis and the y-axis of the grid. Moreover, each patch is associated with a level of elevation (which can be zero or a positive number), which determines the position of each patch along a z-axis. We thus end up with a three-dimensional landscape, as shown below.

[13] See Wright (1932) for the original introduction of these models in evolutionary biology, and Smith et al. (2002) for a more contemporary example of such evolutionary fitness landscape models.
[14] Following Weisberg and Muldoon's (2009) pioneering model, epistemic landscape models have been used to study social epistemology of science by Weisberg (2013); Alexander et al. (2015); Thoma (2015); Pöyhönen (2017); Currie and Avin (2019); Harnagel 2019; and Avin (2020). See Grim et al. (2013) for a closely related model.

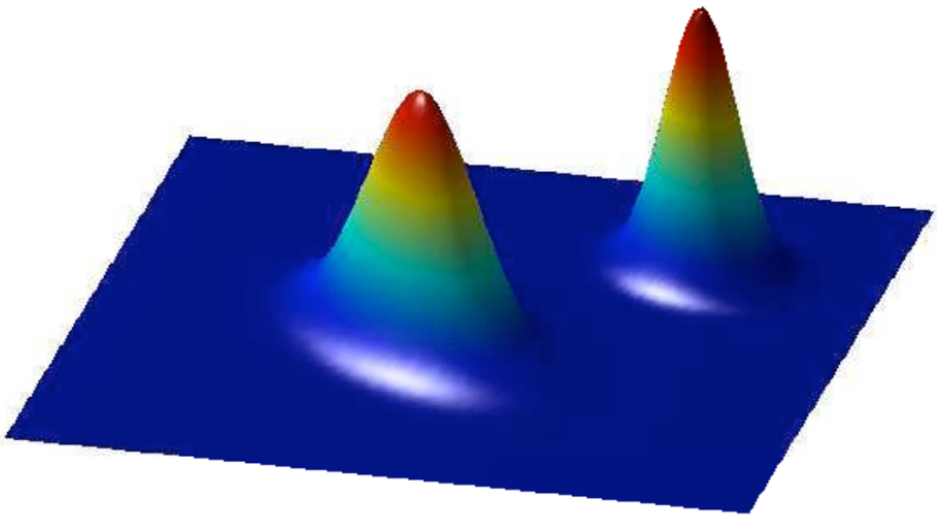


Figure 1 (from Weisberg and Muldoon 2009). A 3D representation of an epistemic landscape.

The three-dimensional representation can be turned into a two-dimensional one, like in a topographic map, as in figure 2 below. Coloured patches represent approaches with a positive level of elevation, and the lighter the patch, the higher the elevation of that patch.

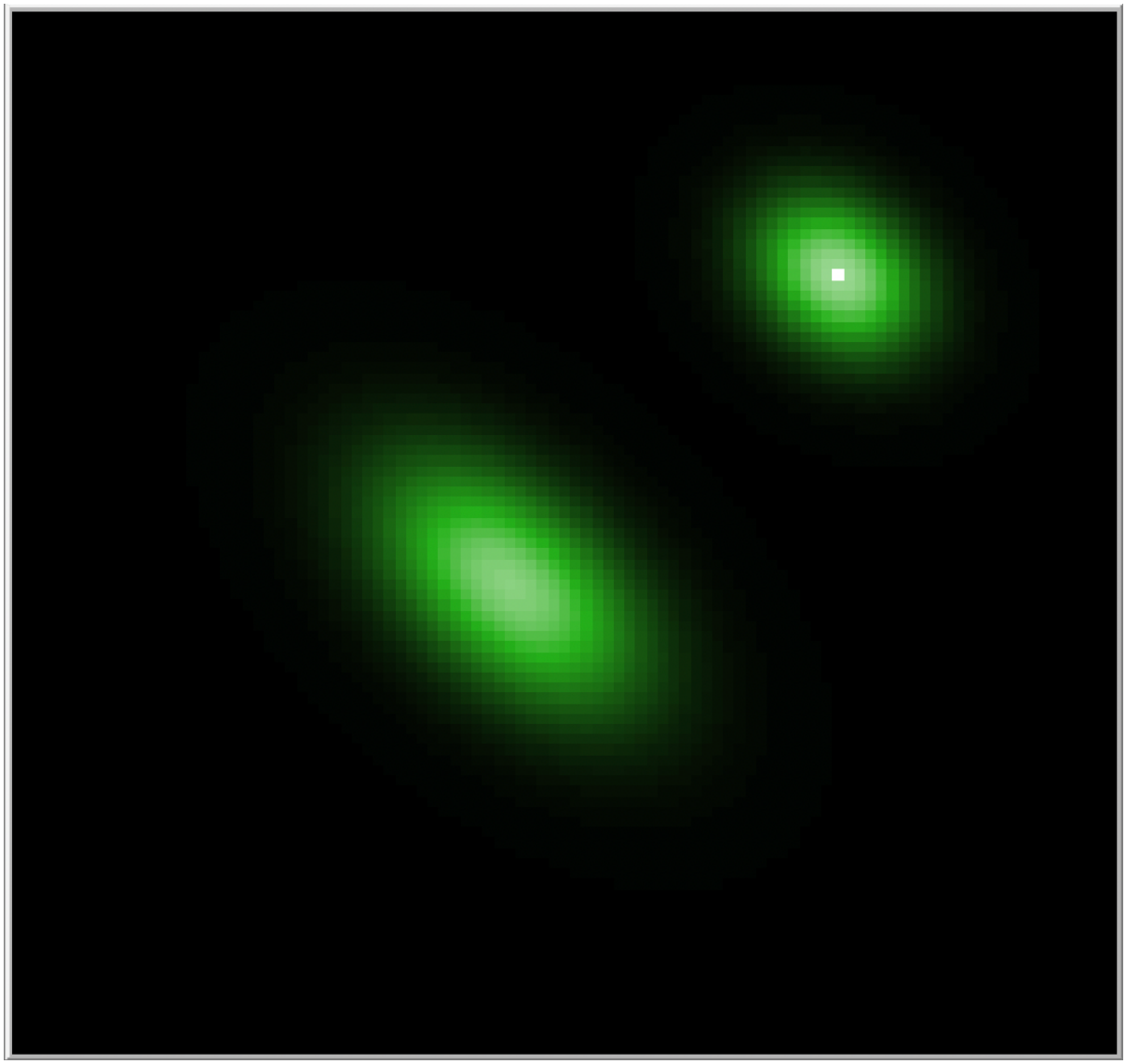


Figure 2. A two-dimensional representation of an epistemic landscape where epistemic significance is bi-modally distributed.

Having focused on the *landscape* aspect of epistemic landscape agent-based models, let me now say a little more on the agent component. Agents populate and interact with the landscape in certain ways. In epistemic landscape ABMs, agents typically move around the landscape, inspecting the elevation of certain patches, and the elevation of the patch of the landscape which each agent occupies at each point influences their movements in the landscape, depending on their objectives and their movement heuristics. Before they start moving, however, agents need to be

distributed across the landscape. For example, an initial random distribution of agents across low-significance patches of the landscape looks as below.

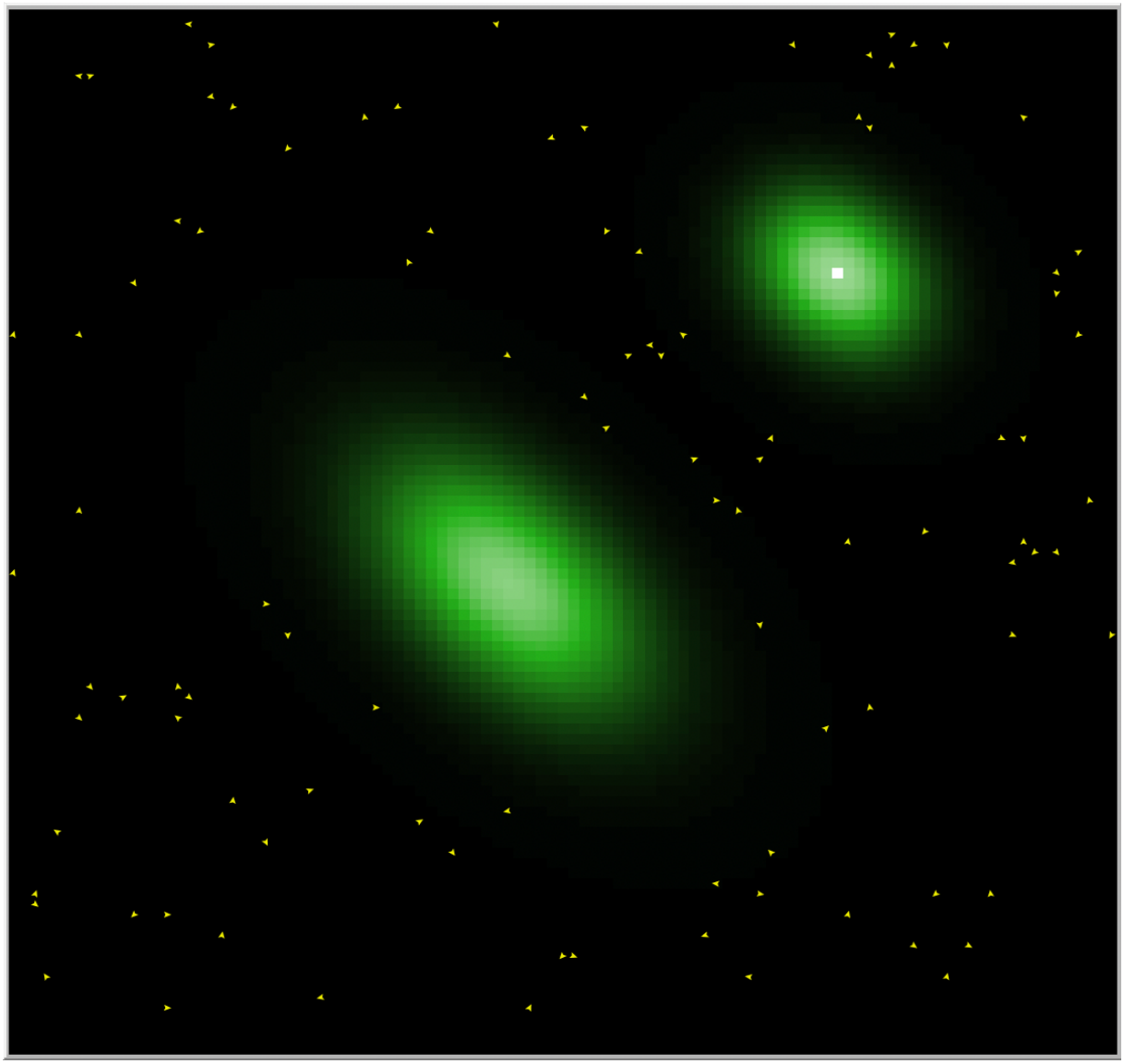


Figure 3. The same epistemic landscape as in figure 2, with 100 individual agents (each agent represented by a yellow tick) added to low-significance patches of the landscape.

In this setting of the model, there are no edges to the landscape, so that the scientists are never 'trapped' against an edge. This means that the square landscape above is a representation of a torus wrapped around the edges.

Even though agents move across the landscape, they need not all move by following the same heuristics. Indeed, as remarked, one potential advantage of ABMs is that they allow one to introduce some complexity at the level of individual decision making. This feature of epistemic landscape ABMs is in turn crucial to exploring the Research Diversity Problem, as it allows me to model individually creative and individually conservative research behaviours. Since this is indeed a crucial element, I will say more about it in the next section.

At this point, the link between Currie's framework and Epistemic Landscape ABMs hopefully comes naturally. In the next section, then, I will specify how exactly I conceive of the translation of Currie's framework into an epistemic landscape agent-based model, and how that will allow me to investigate the Research Diversity Problem.

3.2. Currie's framework as an epistemic landscape ABM

I have mentioned that some advantages come with using models to study certain phenomena. In particular, I argued that the social epistemology of science is a field which might benefit especially from the use of models due to the complexity of the phenomena it studies, and I argued that ABMs might be especially suitable for the task. I then further specified a type of ABM – epistemic landscape ABMs –, and I am now going to show why they are particularly useful for studying the Research Diversity Problem, especially if one has Currie's as the basis of one's model. So let me set up

the epistemic landscape ABM with reference to Currie's framework which I introduced in the previous chapter[15].

3.2.1. The landscape

I will start with the landscape. The landscape is intended to represent a given topic of scientific research broadly construed – for example, molecular biology, or behavioural welfare economics, or perhaps gastrointestinal oncology. One question that may be raised is how broadly one should conceive of a given research topic as represented by a landscape (contrast, for example, Weisberg and Muldoon's (2009) with Thoma's (2015) answer to this question). For example, one may conceive of a landscape as representing a topic as broad as virology, or as narrow as the structural biology of the SARS-CoV-2 virus. In the next chapter, I will suggest a way in which to deal with this challenge. For now, let us grant that research topics are broad enough that one particular approach in the model corresponds to more than one particular paper or narrow research question that each scientist might extract from it. In this sense, the epistemic landscape I have in mind is *not* such that, once an approach is tried, it is no longer worthy of being adopted again. Rather, scientists will still derive relevant knowledge about the topic from productive approaches even if they have been previously visited (I will indicate shortly what I mean by 'productive').

Each patch, in turn, represents a different research approach to the research topic in question. This captures naturally Currie's 'solution space'. Each approach offers a

[15] I will note when a certain feature of my model follows directly from a previous instance of an epistemic landscape ABM.

possible 'solution' to the topic in question. To be clear, and especially given that I am interpreting each research topic relatively broadly, 'solutions' are not to be interpreted as all-or-nothing answers to given questions so much as ways in which one can extract interesting and important, but usually non-exhaustive, facts about a given topic. The elevation of each patch, on the other hand, represents how the extent to which, if adopted, the approach represented by the patch would contribute to the progress of science regarding the research topic under investigation. In particular, following Weisberg and Muldoon (2009), one may think of each approach as having associated to it a level of 'epistemic significance'. The more epistemically significant (or, to use the term used in the previous paragraph, productive) an approach is, the higher the elevation of the patch which represents it. The epistemic significance of an approach, in turn, is to be understood as a measure of how significant the discoveries associated with that approach are to the understanding of the topic under research[16].

It is pertinent to say more about the patches and their representation of 'approaches' or 'solutions' to a given topic of scientific research. I have noted that every patch is identifiable by a unique ordered pair (x, y), referring to the x- and y-coordinates of that patch, respectively. We can then conceive of these two axes as representing two dimensions along which approaches are distinct from each other. In fact, this is a simplification of the fact that approaches may vary along a number of dimensions. For example, approaches may be distinguished in terms of what background theories they

[16] As Weisberg and Muldoon (2009: 6) note, working with a single landscape common to all scientists requires the assumption that scientists make approximately the same judgements as to the significance of the facts uncovered by a given approach. In real life, this might not always exactly hold. Scientists may disagree on the extent to which a scientific contribution is significant to the progress of science. However, for the sake of simplicity, this is an idealisation worth making.

assume, what kind of data is gathered, how that data is analysed, what questions within the topic are investigated (Weisberg and Muldoon 2009), etc. For modelling purposes, though, all these dimensions are collapsed into two dimensions so that the totality of possible approaches can be modelled as a two-dimensional grid of patches.

This way of mapping possible approaches onto a landscape of patches resonates a lot with Currie's conception of a 'solution space' being explored by a community of scientists. Moreover, this interpretation allows one to overcome the difficulty that I flagged in the previous chapter concerning Currie's framework. As I mentioned, Currie is unclear about what makes certain solutions closer to a given reference point on the solution space. The interpretation suggested makes sense of this metaphor. The closer two patches are to each other, the more similar (in terms of the dimensions under consideration) are the scientific approaches they represent.

Another point of natural affinity between Currie's framework and epistemic landscape ABMs is in the fact that Currie suggests that some solutions work better than others at solving a scientific problem. Similarly, in epistemic landscape ABMs some approaches will be highly apt for studying a given scientific area of research – in which case they yield high epistemic significance–, while others won't. This is captured by the elevation of the patch on the landscape that stands for that approach. Thus one may think of Currie's solution-space as an epistemic landscape, where the third dimension (the height) represents how enlightening the solutions are to the general problem at hand.

Regarding the shape of the landscape, I also draw heavily from Weisberg and Muldoon's model, as I conceive of the landscape as a two-peaked landscape (as in figure 1), where the significance of non-zero significance patches progresses smoothly according to the two-dimensional Gaussian function (see Weisberg and Muldoon 2009: 11, footnote 9). This serves the point of simplicity, but it also has some theoretical motivations. Firstly, if an approach yields significant discoveries, then it is plausible that similar approaches (i.e. close-by patches) will also yield significant discoveries. Epistemically significant patches are thus clustered, not randomly dispersed. Moreover, it is also plausible that there is not just one single cluster of approaches yielding significant truths. If one thinks of the landscapes as representing relatively broad research topics, there might be substantially different approaches discovering important facts about the topic. For example, behavioural approaches to welfare economics are significantly different from approaches assuming ideally rational economic agents. Nevertheless, both approaches have yielded significant results and contributed to the progress of the discipline. The fact that the landscape has more than one peak is supposed to capture this. Of course, the fact that there are *exactly* two peaks might not be particularly meaningful, but this is an idealisation which promotes simplicity in a first approach to exploring Currie's conception of creativity[17].

3.2.2. The agents

Having spent some time interpreting the landscape, I will now briefly say something about the agents. Intuitively, the agents in epistemic landscape ABMs represent

[17] For papers exploring the consequences of different landscape shapes, see, for example, Grim et al. (2013); Alexander et al. (2015); or Pöyhönen (2017).

individual scientists or lab teams working on the topic represented by the landscape which they populate. The totality of the agents makes up a scientific community. As I have noted, agents move around the landscape and they learn about the elevation of certain patches. This information, in turn, affects their future movement decisions. But I have also noted that the exact way in which this happens may vary across the same population of agents, depending on what 'type' of scientist each agent is. And, indeed, it will become clear in the next section that this variability is a way to study the dynamics of creativity. On a general level, though, the fact that agents move around the landscape inspecting the elevation of patches and adapting their movement is intended to represent their individual exploration of the topic in question.

Before diving into the different types of movement strategies, I should first say more about which patches scientists can learn about at each time. In order to motivate my modelling decisions concerning this aspect, I highlight a few intuitive factors. Firstly, regarding the epistemic significance of approaches, if scientists can know anything, it is the significance of the discoveries enabled by the approach that they have just adopted. It also seems plausible to assume that a characteristic aspect of scientific research involves trial and error (see e.g. Popper 1963: Ch.1) and the uncertainty of the epistemic benefits brought about by conducting certain experiments and engaging in certain kinds of research. In other words, scientists are not able to determine *a priori* which approaches are going to lead to the most significant discoveries. If they were, scientific research would be much less uncertain and more efficient than it is.

The way I model this (following previous, similar models) is by assuming that agents are able to assess the elevation of the patches which they are occupying at a given time but are not able to know the exact elevation of unvisited patches in the landscape. What of previously visited patches?

In order to address that question, I should restate that scientific communities are also social structures – scientists learn not only from their own research, but also from the research of others when that research is registered and communicated. As such, in general, the knowledge a scientist has about how promising a given approach is for studying the research topic in question is not limited to the approaches that that scientist has adopted herself. Rather, individual scientists or lab teams may also be able to appreciate the epistemic significance of the discoveries achieved by other scientists employing different approaches. However, there are limits to what scientists can know about the approaches that they may adopt to study a given topic of research. In particular, scientists are not necessarily equipped to know the epistemic significance of *all* the approaches that have been tried within the research topic in question, especially if one conceives of approaches relatively broadly. After all, scientists do not have unlimited information or cognitive abilities (Payette 2012). As such, agents with a particular methodological and disciplinary focus will plausibly be able to assess the epistemic significance of only so many patches, and it is natural that their ability to assess this significance correlates with how close these other approaches are to their own. In this way, the fact that agents rely on 'local information' can be seen as capturing a certain degree of bounded rationality of the agents (see Epstein 2006: Ch. 1). A scientist working in the general field of robotics is not

necessarily aware of all tried approaches to the topic, and she is likely unable to gauge the importance of certain discoveries, especially if they follow from approaches which are quite different from one's own. Someone working in *bio-inspired robotics*, for example, may not be especially good at keeping up with the latest developments in *quantum robotics*, and therefore they may not be particularly good at identifying significant discoveries and the approaches that led to them.

Therefore, it makes sense to model scientists as being able to gauge the epistemic significance of a previously tried approach, provided they are sufficiently familiar with it. The way I translate these considerations into the model is the following. Scientists adopt the approaches represented by the patches which they occupy. In so doing, they discover some things about the research topic in question, and they are able to determine the elevation of that patch, as I have mentioned. Moreover, agents can inspect the elevation of those previously visited patches in the landscape which represent approaches sufficiently similar to the one being adopted by that agent. In the model, I take this group of sufficiently similar approaches to be represented by the neighbourhood of one's patch. To summarise, when a given approach is beyond one's realm of expertise (which may approximately be represented by one's neighbourhood), one becomes less able to assess the merits and demerits of such an approach, and that is why an agent's knowledge of the landscape is limited to its (previously visited) neighbourhood. A visual illustration of the model is represented in figure 4 below.

Neighbour1 100	Neighbour4 ?	Neighbour6 ?
Neighbour2 300	My patch 250	Neighbour7 ?
Neighbour3 ?	Neighbour5 ?	Neighbour8 ?

Figure 4. An agent (blue triangle) occupies a certain patch. Of the eight neighbours (what is called the Moore neighbourhood), only Neighbour1 and Neighbour2 have been visited previously (either by the agent represented or by another agent of the population). The scientist represented can thus know the epistemic significance of those two neighbouring patches (100 and 300 respectively), but she does not know the significance of any of the remaining neighbours, as they have not been visited previously by any member of the scientific community.

One final point about agents' movement in general is that agents are not stationary, and they typically search for higher and higher patches. This represents scientists' interest in making more and more significant discoveries on the topic they are researching. There are a few possible roots of that interest. Scientists may be intrinsically motivated to know more and more significant things about the topic in question; or perhaps they are interested in making a big difference to the field and being acknowledged as having done such; or perhaps they want to maximise the likelihood of getting funding for future projects and a way to do that is to have as good a research track record as possible. Thus there are several reasons which could motivate agents' movement, in particular while they have not yet reached a peak of the landscape. As such, even though the epistemic significance of a patch is not assumed to be completely depleted once that patch is visited, agents are still plausibly modelled as having an interest in keeping moving through the landscape.

How exactly they move in response to what they learn about the landscape, though, is an important aspect of the model. And, importantly, different agents may adopt different strategies to respond to the information they gather about the shape of the landscape they navigate. This, in turn, will play an important role in my approach to modelling creativity and diversity. This is the task I turn to now.

3.3. Creativity in the model

As I have hinted, it is in modelling the agents' movements and their distribution across the landscape that one can start thinking about creativity and conservatism (as conceived by Currie 2019b) more explicitly in the model. This, in turn, will make clear how my proposed model naturally captures the remaining elements of Currie's framework. So let me approach individual- and group-level creativity at a time and how I conceive of these in the model. I start precisely by introducing how the model can be set up in such a way that is sensitive both to individual- as well as group-level creativity. I will then show how each of these two notions of creativity, when evaluated through the perspective of the epistemic landscape ABM, latch onto two important and distinct notions of research diversity. This will allow me to highlight at the end of this chapter yet another advantage of thinking about the problem of conformity in science through my proposed model – that it captures two notions of research diversity that have so far been either considered independently of, or confused with, each other.

3.3.1. Agents' movement and individual-level creativity

Epistemic landscape ABMs allow one to model different individual behaviours and explore interactions between them. Weisberg and Muldoon (2009), for example, employ their pioneering model precisely to study how different kinds of research strategies result in different levels of epistemic success. In particular, they explore what the epistemic consequences are of having more 'maverick' or, on the other hand, more 'follower' individual scientists, two notions which are very closely linked with individual-level creativity and conservatism, respectively. Indeed, epistemic landscape ABMs have been employed mainly to study the macro-level effects of such variations in individual research behaviour. Currie's cold and hot searches can thus be captured by defining different decision heuristics that individual scientists may adopt.

In chapter 2, I characterised individual-level creativity and conservatism, and I highlighted that, even though creativity is a matter of degree, for the purpose of my model I will simplify that feature into a binary. Thus the conservative-creativity distinction at the individual level is captured in my model by the different research strategy types a scientist can adopt. I consider three types of scientists, which differ solely in terms of the research strategy they employ, i.e. the way in which scientists navigate the epistemic landscape. The three types are:

- Controls
- Conservative scientists
- Creative scientists

Let me, then, consider each type of scientist at a time.

I borrow the control rule from Weisberg and Muldoon's (2009) and subsequent models. Controls are scientists who do not take into account any kind of social information. In particular, it does not matter to their decision of which patch to visit next whether their neighbouring patches have been visited before. They are 'controls' in the sense that they control for the effect of using social information to navigate the landscape. It is useful to have the pseudocode of the control scientists (see Alexander et al. 2015: 11). In each cycle of the simulation, they follow the rule below[18].

Control Rule

```
Control_scientist checks [significance of current patch]
        a. If [significance of current patch] = [significance of
           previous patch]
              i. With 98% chance, don't move
             ii. With 2% chance, move one patch in random
                 direction
        b. If [significance of current patch] > [significance of
           previous patch]
              i. Continue in the same direction for one patch
        c. If [significance of current patch] < [significance of
           previous patch]
              i. Go back to previous patch
```

[18] Cycles in the model are interpreted as a unit of time, where each agent in the model follows its respective moving heuristic once. In this model, one research cycle corresponds to the time taken to explore one particular approach, as in every new cycle agents attempt to move.

The pseudocode makes clear that all that matters for control scientists' movement is how the significance of the patches which they currently occupy compares to the significance of the previous patch. They do not consider any information regarding whether patches have been visited before and what the elevation of previously visited neighbouring patches is. We can then compare this behaviour with the more socially informed behaviour which real-life scientists are more likely to exhibit. Still, I would like to contrast a more (individually) conservative behaviour with a more (individually) creative one. Thus I model two other types of scientists. Those who navigate the landscape in a conservative way, taking small steps, searching for approaches in a 'cold' fashion, following the steps of previous successful scientists around their area of the landscape; and those who search the landscape creatively, jumping across the landscape. While the pseudocodes for both the creative and the conservative types are too convoluted to be productively transcribed here, it is still useful to note their strategies on a less formal level.

Conservative Rule

For each cycle of the simulation, conservatives check the epistemic significance of the patch they are currently occupying.

1. If the epistemic significance of the current patch is greater than the epistemic significance of the patch they had (immediately) previously occupied, their strategy to move is the following.

a. Each conservative agent checks which of its (eight) neighbouring patches have been previously visited (either by them or by any other scientist).
    1) If no neighbour has been visited, they go ahead[19].
    2) If some neighbouring patches have been previously visited, then conservatives inspect whether the significance of at least one such visited patch is higher than the significance of the patch they currently occupy.
        a) If so, then conservatives will go to the previously visited neighbouring patch with the highest significance. (The idea behind this is that conservatives like what has been proven to work well; if, on top of that, it works better than the approach they are currently employing, then they are happy to occupy that patch rather than exploring an unvisited patch and risk landing on a less significant patch than they are currently occupying.)
        b) If, on the other hand, of all visited neighbouring patches, there is none whose epistemic significance is higher than that of the patch which they are occupying, then they choose an unvisited path at random. If there are no unvisited neighbouring patches, then conservatives will have to do with a visited neighbouring path of equal significance to the currently occupied patch. If all neighbouring patches are,

[19] Given that conservatives move between adjacent patches, the only time a conservative scientist is exclusively surrounded by previously unvisited neighbours is in the first cycle of each simulation.

however, of a lower significance, then conservatives stay where they are, as they have reached a local peak.

2. If, on the other hand, the epistemic significance of the patch they are currently occupying is not higher than the epistemic significance of the previously occupied patch,
    a. then they incur what I call a 'moving likelihood cost', which, like with the controls, is of 98%. That is, with 98% likelihood, they do not move in the next cycle. With 2% likelihood, they follow the moving strategy described under (1.a.) above.

Finally, the creative research strategy.

Creative Rule

For each cycle of the simulation, creatives check the epistemic significance of the patch they are currently occupying.

1. If the epistemic significance of the current patch is greater than the epistemic significance of the patch they had (immediately) previously occupied,
    a. then they jump the equivalent of 10 patches across the landscape in a random direction. They might land on a previously visited patch, in which case they go back the patch from which they jumped. Otherwise, they stay on the patch on which they landed.

2. If, on the other hand, the epistemic significance of the current patch is not greater than the epistemic significance of the patch they had (immediately) previously occupied,
    a. then with 99% likelihood, they do not move in that round. With 1% likelihood, they follow the moving strategy described in (1.a.).

Having provided the movement rules for all three research types, I will now discuss some important aspects of this element of the model. Firstly, a note on the moving likelihood cost. Weisberg and Muldoon (2009) only apply (what I call) a moving likelihood cost to controls[20]. The original purpose of this cost is to introduce a degree of exploration to controls' research strategy, but which is low enough that it would not lead them to abandon local peaks (Weisberg and Muldoon 2009: 9). The problem, as Alexander et al. (2015: 11) note, is that if a comparable moving likelihood cost is not applied to the other research strategies, then the different types of scientists are operating at fundamentally different time-scales. As such, in my model, both conservative and creative scientists (and not merely control scientists) face a moving likelihood cost after they move to a patch of equal or lower significance to the one of the patch they had previously occupied. This equalises time-scales across researcher types.

This moving likelihood cost need not be seen as an *ad hoc* solution to the problem raised by Alexander et al. (2015). Indeed, it can be given an interpretation that takes

[20] In fact, what I call a moving likelihood cost Weisberg and Muldoon (2009:9) call an 'experimentation rate', since control scientists move in a *random* (exploratory) direction with 2% probability.

into account the institutional and cognitive factors that are relevant in social epistemology of science. Under this light, the idea of a moving likelihood cost is that scientists adopting approaches with increasing levels of epistemic significance enjoy a level of epistemic and institutional success which, in turn, allows them to progress more quickly. Increasingly successful scientists can attract funding and talent for new projects more easily, which also enables better learning of the topic they are exploring. If, on the other hand, they do not reveal such development and fail to uncover increasingly significant facts about the research topic, they face greater difficulties to engage in new projects and exploring their interests further. It is in this scenario that they incur a moving likelihood cost.

This intuition also helps explain why in my model creatives incur a moving likelihood cost of 99%, while conservatives incur a cost of 98%. The motivation for this increase is that being creative is particularly costly[21]. If one is a scientist who 'jumps across approaches', one requires a lot of resources, both cognitive as well as financial and reputational, for example in learning and developing new background theories, becoming familiar with different methods, learning new programming languages, acquiring new experimental equipment, etc. If that new approach ends up not being fruitful – or, at least, not more fruitful than one's previous approach –, then the scientist will face difficulties doing another 'jump'. It may be difficult finding capable scientists willing to join the team, it may be difficult to get funding, and, of course, it

[21] See Poyhönen (2017: 4527) for an alternative way of introducing an 'exploration cost' in an epistemic landscape ABM.

will require effort to learn a whole new, substantially different set of theories and methods for the next approach.

Another point to note concerns the main differences between the conservative and the creative scientist's moving strategies more generally. Firstly, the conservative moves 'coldly', 'taking small bites', exploring the solution space patch by patch, tweaking her approaches slowly and gradually, while the creative moves ten steps at a time, each jump leading her to significantly different approaches than she had occupied previously. This nine-patch difference is intended to capture the difference between 'cold' and 'hot' searches which Currie highlights.

While this is a relevant difference between the two types, I introduce a feature into the movement strategies which captures another important difference between conservatives and creatives. This is a feature which Currie's framework fails to capture (at least explicitly) and which relates to the social nature of science. In this sense, this is another way in which my model can be seen as a refinement of Currie's framework. The added feature is that conservative and creative scientists (unlike controls) do care about what has been discovered previously, either by them or by others; however, they use this information differently. Conservative scientists, on the one hand, are strongly guided by what has been previously done and proven to work in the field they are exploring, even if it means only slightly improving their approaches each time. In this way they follow the 'well worn paths' that are highly likely to guarantee a satisfactory epistemic return. Thus I model conservative agents as highly sensitive to their surroundings, preferring moving to previously visited patches whenever they

guarantee that they will move to a higher patch than they were previously occupying. Only when that is not possible do conservatives venture into the unknown.

Creatives, on the other hand, are more interested in unleashing new and promising, potentially ground-breaking, ways of studying a given topic. In this sense, creatives are much less reliant on previously tried approaches (either by themselves or by other scientists), even if that risks wasting the 'well worn paths' that would safely provide them with results that would be considered interesting by the community. As such, I model creative scientists as actively avoiding previously visited patches. Therefore, like conservatives, creatives use information regarding whether other patches have been visited by the other community. Unlike conservatives, though, creatives use this information in order to know, not what patches to visit, but what patches *not* to visit.

At this stage I have presented the main elements of individual-level creativity. As I have noted, creativity and conservatism at this level are captured in the model by what type of scientist one is. Scientist types are, in turn, differentiated in terms of the rules that determine their movements. Conservative scientists move in short steps and tend to follow previously visited paths if they seem promising. Creative scientists, in turn, jump across the landscape and prefer the unexplored.

While individual-level creativity is an important aspect of Currie's framework, it is not the only one, and I will now present the second notion of creativity of prominence in Currie's framework.

3.3.2. Agents' dispersion and community-level creativity

The model which I have presented so far displays features which have close parallels in extant epistemic landscape ABMs. These models explore the dynamics of nonconformity in science by adopting an individual-level conception of creativity (albeit somewhat different from the one I have presented here). Thus, one important notion that has been relatively neglected in the study of the Research Diversity Problem and the issue of conformity in science has precisely been the idea that communities as a whole may display different degrees of nonconformity and that this might be independent of individual scientists' nonconformity. As I have highlighted in this this thesis, one of the reasons why Currie's framework is so appealing to work with is that it allows one to attend more explicitly to this notion.

Based on Currie (2019b), I have noted that there may be individual-level creativity without community-level creativity. For example, if scientists do not register and share their findings, information is lost as time goes by and so, in the long term, the 'wide exploration of the solution space' does not obtain. However, for the purposes of this thesis, I will not explore this possibility. I will assume that scientists do register and share their results and that, as such, if a significant number of individuals are creative, then, given the nature of their erratic research trajectory, that eventually leads to a wide exploration of the solution space at the community level. Instead, I will build my model such that it is sensitive to the other possibility, i.e. that of having community-level creativity without necessarily having individual-level creativity.

Before implementing this feature of the model, it is pertinent to motivate it. What does a scientific community that is creative at the group-level but not at the individual-level look like? The answer depends in part on how broadly one conceives of a research topic, but, again, consider something relatively broad such as particle physics, labour economics, or perhaps potamology (the study of rivers). We want to conceive of possible scientific communities which explore various and different approaches to that study. But this does not require that individual scientists always try significantly different approaches from the ones they had tried before. Instead, there might be 'pockets of conservatism' across the solution space. By this I mean that scientific communities may be such that they do, collectively, attempt different solutions, but they do it while having individual scientists conservatively explore their areas of expertise. Thus some labour economists may study the dynamics of unemployment via agent-based models based on a search theoretical framework (e.g. Kant et al. 2020), while others may approach it via micro-theoretic models of monopsony (e.g. Manning 2013), and others may explore the sociological roots and implications of the phenomenon (Boland and Griffin 2015), etc. The scientists learn primarily about the developments in their sub-area of research, and proceed to change their approaches gradually, taking insights from academics who have done similar research. In short, scientists act conservatively. However, that does not mean that the community as a whole behaves conservatively. If there is enough dispersion across possible approaches to studying a topic, then the topic in question is studied creatively. This is precisely what is suggested by Currie (2019b), but which has been largely neglected in the literature concerning scientific creativity, especially the one employing ABMs to study it.

The way I conceive of community-level creativity in the model, then, is in terms of the dispersion of agents across the epistemic landscape. Essentially, scientists do not need to be creative themselves in order for such a community-level creativity to be achieved. With a dispersed distribution of scientists, all scientists may behave individually conservatively by exploring the patches that are immediate neighbours of one's patch at a certain time. Provided scientists are sufficiently widely distributed and in enough number, community-level creativity in Currie's sense is still achieved.

As such, even if one assumes that individual-level creativity does tend to lead to community-level creativity, as it leads to scientists spreading widely across the epistemic landscape, the more important point is that this is not the only way of achieving such wide dispersion. Indeed, I suspect that it is because of this intuitive link between individual-level and community-level creativity that community-level creativity or similar notions have been largely ignored as relevant concepts in the literature regarding ABMs of conformity in science – insofar as it is an important notion, it has been assumed to be somehow reducible to individual-level creativity. This perhaps explains that previous epistemic landscape ABMs have failed to capture this relevant nuance related to the Research Diversity Problem. Thus, while extant models have explored to quite some detail the dynamics of individual-level creativity in science (see Weisberg and Muldoon 2009; Grim et al. 2013; Thoma 2015; Pöyhönen 2017; Avin 2020), none has focused on the driver of community-level creativity which achieves it via a wide dispersion of scientists. Indeed, Weisberg and Muldoon (2009) (and the mentioned subsequent variations of their epistemic landscape ABM), set up

their model by randomly distributing scientists across the landscape (in low-significance patches) as in figure 3.

Yet, given what I have said in this section, it is not at all obvious that this is a harmless way of implementing the model. In fact, by Currie's definition, the initially dispersed distribution of scientists featured in extant Epistemic Landscape ABMs means that current models consider highly creative communities by default. Instead, one should allow for the possibility that, in at least some research topics, scientists start out rather more concentrated around a set of similar approaches. I thus introduce a feature to the model, the cluster variable, which allows one to decide how 'clustered' a certain scientific community is before scientists start moving and exploring the landscape, as represented in figure 5 below[22].

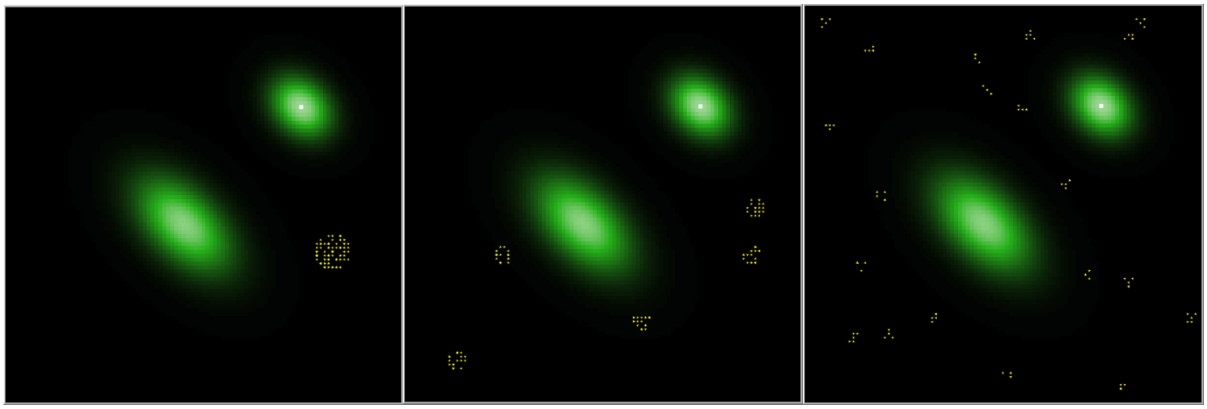


Figure 5. The same epistemic landscape as in figures 2 and 3. Similarly to figure 3, there are 100 individual agents added to the landscape, but, in this case, they are distributed across one, five, and 20 clusters, respectively.

Of course, this feature of the model is not a perfect indicator of (initial) community-level creativity. Suppose for example, that I model the landscape to be populated by

[22] The model is designed to minimise the discrepancy in the number of scientists per cluster. For example, if a given landscape is modelled so as to be populated by 100 scientists across three clusters, they are randomly distributed across the clusters so that two of those clusters have 33 scientists each, and the third one has 34 scientists. Scientists are distributed randomly, regardless of their type (control, conservative, creative).

two clusters of scientists. Due to the random nature of the assignment of agents to the patches, these two clusters may be further from or closer to each other in one landscape than in a second landscape with exactly the same number of scientists distributed across two clusters. Nevertheless, on approximation, this is a simple and accurate way of capturing the extent to which scientific communities are creative.

This completes the presentation of the model and the main features of science it is supposed to represent. Before putting the model to practice, I turn briefly to the topic of cognitive diversity, which is often discussed in the context of the Research Diversity Problem, and I will show how my model helps understand different notions of research diversity more clearly and systematically.

3.4. Cognitive diversity in epistemic landscape ABMs

I began this essay by presenting what I called the Research Diversity Problem. I initially formulated such a problem in intuitive terms: scientists can adopt different strategies to study a given research question. They may be more or less conformist. They may be more or less inclined to challenge the theoretical, methodological, experimental *status quo* (if there is one). Scientific communities as a whole, too, can have different coordination arrangements to explore a given scientific topic. Based on the framework developed by Currie (2019b), I then showed that one may account for these different ways in which scientific questions may be approached in terms of creativity and conformity. Furthermore, I further developed these notions of creativity and conservatism and translated them into an epistemic landscape agent-based model of a scientific community approaching a given research topic.

The model I developed also sheds light on the kinds of research diversity that are entailed by the Research Diversity Problem. The Research Diversity Problem is so called because it deals with the diversity of arrangements and strategies – at both the individual and the group levels – which may be employed for the study of a given research topic by a scientific community. The problem, then, lies in finding out which such arrangements are epistemically optimal and institutionally feasible. What I want to highlight at this point is that the model I presented in the previous section actually conveys two important but distinct notions of research diversity. I will show how each notion of diversity follows from each of the two conceptions of creativity (one at the individual, the other at the group level).

A main focus of the existing literature in social epistemology and philosophy of science is precisely on the epistemic benefits of diversity broadly construed[23]. In particular, as Alexander et al. (2015) and Heesen (2019) highlight, there are two distinct types of diversity that are conveyed in epistemic landscape ABMs of the kind that I have presented. Unfortunately, there has not been a uniform way in which these two notions have been referred to in the literature. Let me review each notion of diversity at a time.

[23] For a sample of such literature see, for example, Kitcher (1990); De Langhe and Greiff (2010); Zollman (2010); Pöyhönen (2017); Borg et al. (2019); Devezer et al. (2019); Grim et al. (2019); O'Connor and Bruener (2019); Reijula and Kuorikoski (2021); Fazelpour et al. (2022). See Muldoon (2013) or Reijula and Kuorikoski (2019) for reviews on the literature on diversity in science.

The first one is the diversity arising from the different ways in which *each* scientist may choose to move across the landscape. Some scientists may be more 'adventurous' than others, and the way I have captured this (based on Currie's framework) was by modelling different scientist types. Diversity of research at this level has been called 'cognitive diversity' (Alexander et al. 2015: 16), 'diversity of research strategy' (Thoma 2015: 456) or of 'researcher types' (Heesen 2019: 10).

The second notion of diversity refers to how different the projects and approaches adopted by each scientist are from those adopted by other scientists in the community at a given time. In epistemic landscape ABMs, this is captured by how scientists are distributed across the landscape at a given point in time (see figure 5 above). This type of diversity has been identified with notions such as 'epistemic coordination' (Alexander et al. 2015: 16), 'diversity of type of research' (Thoma 2015: 456), or 'division of cognitive labour' (Heesen 2019: 10).

The term 'division of cognitive labour' – used by Heesen to refer to this second kind of diversity – however, is the term used by Wesiberg and Muldoon (2009) to refer to the first kind of diversity I have identified – i.e. what Heesen (2019) identifies as 'diversity of researcher types'. On the other hand, the same term (division of cognitive labour) is borrowed from the work in social epistemology of science pioneered by Kitcher (1990), which is closer to addressing the second kind of diversity I highlighted, in the sense that it is concerned with finding the collectively optimal distribution of scientists across different research projects. This illustrates the terminological ambiguity and conceptual confusion involved in the Research Diversity Problem.

Fortunately, the model I introduced helps subsume these notions of diversity under a unified conception of creativity. Intuitively, each of these two types of diversity follows directly from variations in each of the two conceptions of creativity represented in the model. The first one is a result of the fact that different scientists may display different levels of *individual* creativity, while the second type of diversity is a result of the extent to which a given scientific community is (*qua* community) creative.

This helps bring some clarity to the question regarding research diversity and the division of cognitive labour. The first kind of diversity has implications for the division of cognitive labour in the sense that it concerns the question of how creative each individual's research strategy should be. Here cognitive labour is differentiated in terms of how individually creative it is. The second kind of diversity concerns rather the different projects in which the scientists of a given community may be involved. In this case, cognitive labour is differentiated in terms of the different approaches (patches) one may employ to study a given research question. Working on the basis of the framework and model which I am proposing in this essay is therefore helpful because it helps disambiguate between two notions of diversity and the division of cognitive labour in the social epistemology of science literature. Granted, such disambiguation has been advocated before by Alexander et al. (2015), Thoma (2015) and Heesen (2019). The main contribution of my model regarding this matter, though, is in making it evident that these two types of diversity, while being distinct, can be understood with reference to one common concept – that of creativity in science.

This is a valuable feature, as it strikes a balance between two appealing positions held in the literature. One is that, by calling their analysis of what combination of individual research strategies is epistemically optimal a study in the 'division of cognitive labour', Weisberg and Muldoon (2009) trade on an ambiguity between the two kinds of diversity highlighted (Alexander et al. 2015; Heesen 2019). Thus there is a benefit to keeping these two notions apart, and seeing both notions as related to different levels of creativity makes this evident. On the other hand, it would be a mistake to treat both notions completely independently from one another. Both notions refer to a kind of diversity of research, after all. Both notions are related by the fact that they depend on considerations about how different a set of approaches are from each other. To that extent, it makes sense to identify both as deriving from the one phenomenon of creativity. My account thus overcomes the conceptual and terminological confusions that have affected the Research Diversity Problem so far.

Chapter 4. Simulating creativity

Having described all the main features of the model I have built, it is now appropriate to put it to use. However, the simulations that I am now in a place to run will be more illuminating if I first consider existing arguments concerning the need for creativity in science.

4.1. Does science need creativity?

One of the main reasons why the topic of creativity (as nonconformity) in science has attracted attention in the philosophy and social epistemology of science is that many authors consider that the scientific *status quo* is not creative enough. The argument is, roughly, that creativity is an epistemically beneficial trait in science, and that there are factors in science that prevent creativity from obtaining, resulting in a sub-optimal epistemic context.

While it would be impossible to consider all arguments for creativity in science, in this essay I have already mentioned some such arguments. Here, I will mainly focus on the arguments put by Stanford (2019) and Currie (2019b) regarding the benefits of creativity on the one hand, and the way in which current institutions and norms stifle that creativity on the other. The choice to focus on Currie (2019b) is fairly evident – his arguments are based on a framework which I have adopted and explored in this essay. Stanford's (2019) arguments, too, are based on a framework which maps easily onto the model which I have developed, as I will make clear in this section. As such, one of the reasons for focusing on these two authors is that their points are fairly naturally tested by the model I have developed.

Let me start by considering why one might think that creativity is an appealing epistemic property. Throughout this essay I have, in fact, raised some major points on why this may be the case, but this gives me the opportunity to present them in a summarised and systematic way, and it will allow me to introduce some new points. Stanford (2019) is concerned with creativity in science from the perspective of someone studying the problem of 'unconceived alternatives' (see Stanford 2001; 2006). He argues that throughout the history of science, scientists have failed to conceive of distinct alternatives to the successful theories of their time, even though such alternatives were well-supported by evidence. He thus suggests that the most serious challenge faced by scientific realism is that 'there are probably such unconceived alternatives to even the most successful scientific theories of our own day' (Stanford 2019: 3916). Creativity is pointed as a way to mitigate this problem. In leading to theoretical innovation, scientists' creativity allows them to '[challenge] existing theoretical orthodoxy' (Stanford 2019: 3917), thus diminishing the likelihood of unconceived alternatives remaining unconceived. For Stanford, then, creativity contributes to exhausting the 'space of well-confirmed theoretical possibilities' (Stanford 2019: 3916) of a given research topic.

Stanford's *space* of theoretical possibilities connects well with Currie's solution-space, which, in turn, can be interpreted as a landscape of an ABM, as I have pointed out. Stanford focuses primarily on the theoretical component of scientific research, but that is mainly because he addresses the issue of scientific creativity from the perspective of the debate about scientific realism, which is concerned to a great deal with the metaphysical status of *theoretical* entities (see, e.g., Psillos 1999). Nevertheless, if

such considerations about scientific theories are combined with considerations regarding other dimensions of scientific research (experimental, methodological, etc.), Stanford's space approximates the multi-dimensional landscape I conceive of in my model. And Stanford's argument can thus be translated as claiming, roughly, that creativity promotes the discovery of the patches of significance at least as high as those that have been already visited, if not higher.

Currie (2019b), on the other hand, offers a less general argument for the epistemic benefits of creativity. He does not claim that creativity is generally epistemically beneficial; rather, he contends that some areas of scientific research are such that they require creativity in order to be appropriately studied. So while in some research topics scientific conservatism may be the best way to go epistemically speaking, in some others conservatism will likely be unsuccessful in uncovering the relevant facts about those topics. Yet how is one supposed to know which areas are more conservative-prone and which are more likely to benefit from some degree of creativity? In order to answer this question, Currie (2019b) uses the science of existential risk as a paradigmatic example of an area of science which requires creativity in order to be adequately studied, but, for my purposes, this is not crucial. What is important is to understand what makes scientific topics such as the science of existential risk areas in which creativity is likely to be epistemically beneficial. In Currie's terms, it is important to understand which 'epistemic situations' render creativity an asset in scientific research[24].

[24] Currie defines the epistemic situation of a given scientific enterprise as consisting of '(1) the challenges facing knowledge generation and (2) the resources available in generating knowledge' (Currie 2019b: 44).

Currie highlights a series of potential challenges that the study of a certain discipline or area of research may face. When (at least some of) these challenges hold, then creativity is called for. I will not go through all of them in detail. Instead, I would note that many of these challenges have to do with the weakness of the evidence available about a certain system under study. For example, Currie points out severe uncertainty as one such factor. If there is severe uncertainty involved in the study of a given system (meaning that one cannot assign precise probabilities to the outcomes under study, but only ranges of probabilities), the study of that system benefits from a rather exploratory research type, as a way to 'mitigate a lack of evidence available to construct theories and models of the relevant phenomena' (Currie 2019b: 44). Largely based on his own previous research (e.g. Currie 2015; 2018), he contends that areas of research facing this and similar epistemic challenges should be 'multi-disciplinary, pluralistic, and opportunistic' and that this characterisation 'meets the criteria for creativity' (Currie 2019b: 46). The argument is, therefore, that in the face of weak evidence about a certain area of research, a research strategy which explores the solution-space rather widely is better equipped to find and generate better evidence, to the extent that the uncertainty about the system in question is substantially reduced. Moreover, Currie (2019b) claims, in such systems where evidence is so weak, one might not even be aware of all the major potential factors determining the outcome of interest. This makes the needs for an exploratory, creative study of the research topic even greater, according to him.

Currie's and Stanford's arguments for the benefits of creativity may not coincide on all points, as I have highlighted. However, they do agree on a crucial point – namely, the idea that scientific creativity is (at least sometimes) conducive to the discovery of significant truths about the topic in question. Their agreement, however, is not only on this point, but also on the claim that the current incentives and broader institutional setup in science promotes conservatism to too large an extent, given the beneficial role that creativity has to play. Indeed, on this point many have been the suggested institutional causes for an overly conservative science. For example, peer review-led funding allocation, it has been suggested, increases conformity in inducing the expectations in scientists that their research grant applications will not be successful unless they conform to a narrow set of methodological, theoretical and experimental practices dominant in their field of expertise (see, e.g., Haufe 2013; Gillies 2014; Avin 2019)[25]. Research norms such as the preregistration of studies and the free publication of the data used in one's studies are also thought to affect scientific creativity, with some arguing that these norms stifle creativity (e.g. Goldin-Meadow 2016), (even though some others (e.g. Frankenhuis and Nettle 2018) argue that they actually promote creativity).

The advent of 'Big Science' (see Galison and Hevly 1992) is another institutional factor indicated by Stanford (2019) as partially responsible for an excessive amount of conformity in science. The term 'Big Science' is intended to describe the way in which contemporary science tends to be organised in ever more collaborative projects, the outcome of which makes up a very substantial portion of the total data that is

[25] Although see Heesen and Bright (2020: 652) for a word of caution on this point.

generated in science. Stanford (2019) argues that in order to deal with the sheer volume and diversity of data and information generated, large, centralised, hierarchical structures were developed. Such centralisation and hierarchy, he continues, induce conformist tendencies in scientists, as the considerations about which projects are worthy of being tried ultimately depend on a rather small set of people and institutions.

Finally, Currie (2019b: 43) highlights that strict disciplinary boundaries, allied to the fact that the most prestigious journals of each scientific discipline usually focus on the core of the discipline, also foster conservatism. As scientists are incentivised to adopt monodisciplinary approaches, the scientific community will restrict itself to relatively narrow areas of the research space and, as a result, potentially epistemically significant multidisciplinary approaches of the epistemic landscape will remain uncovered (to put it in epistemic landscape ABM vocabulary).

I should re-emphasise that with this review I do not claim to have exhausted the factors that are often suggested as being the causes of an alleged overly conservative science. However, these aspects constitute a fairly comprehensive overview of the institutional factors that are often indicated[26]. I have thus shown that according to these arguments, on the one hand scientific creativity is epistemically beneficial – it

[26] On top of this, some other mechanisms less focused on specific institutions of contemporary science have been suggested as being behind conservatism in science. O'Connor (2019), for example, offers an account of the evolutionary disadvantage of 'maverick-y' science; and Kummerfeld and Zollman (2016) argue that even in what they call the 'scientific state of nature' – a scenario in which scientists are left to their own devices, with no added institutional pressures to pursue one type of research rather than another – scientists would still conduct research that is less creative than collectively desirable.

leads to the discovery of significant facts about the research topic in question which would otherwise be unlikely to be uncovered. On the other hand, actual scientific institutions are organised in such a way that often discourages creativity. The bottom line: more creativity is needed.

What I suggest, though, is that the awareness that a scientific community may achieve group-level creativity in different ways raises the question of whether all such forms of achieving group-level creativity are equally desirable. I doubt this is the case. In particular, an arrangement which is able to combine the benefits of conservative research with the benefits of creative research intuitively has an edge over other solutions. Such an arrangement is, presumably, organised along the lines described in the previous chapter – i.e. an arrangement whereby individual scientists behave mostly conservatively, so that they are able to follow best research practices and gradually improve their approaches, but in which, as a whole, they are not overly concentrated around a small set of approaches. The simulations which I will present in the next section will help me test this suggestion in a precise way.

4.2. Simulations, simulations, simulations

As I have indicated, the arguments above prescribe a certain level of scientific creativity, without which, it is suggested, scientific communities are epistemically worse off. Part of my goal is to test this initial claim. Is creativity really epistemically advantageous, in the sense that it enables communities to uncover significant truths that would otherwise likely be left uncovered? However, and as I have also previously hinted, this question hides the nuance – not explored in epistemic landscape ABMs so

far – that there are different ways for a scientific community to be creative, and they need not rely on individual scientists being creative. This raises the further question of what ways of achieving creativity at the group-level are most effective at realising the epistemic benefits which creativity purportedly facilitates.

In sum, I am ultimately concerned with the extent to which scientific creativity (conceived in both individually and on a group-level) is conducive to scientific progress, and, relatedly, with what the creativity profile of a scientific community should be if the goal of that community is to uncover epistemically significant facts about a given research topic. These I take to be the questions that sum up the Research Diversity Problem, and I have presented a framework to think clearly about them. Moreover, I have introduced an epistemic landscape ABM which will allow me to explore these questions more systematically. In what follows, I will present a series of computer simulation experiments based on the model described in the previous chapter[27]. I will use these simulations to assess and compare the epistemic performance of scientific communities with different creativity profiles (both at the individual and at the community levels). Importantly, this exercise will allow me to test the plausibility of my hunch that an arrangement which combines both individual-level creativity with community-level creativity is the most efficient for discovering the relevant facts about a given research topic.

[27] The model is implemented and simulations are run using version 6.2.2 of the programme NetLogo (Wilensky 1999). See also Wilensky and Rand (2015).

4.2.1. The basic model

Epistemic landscape ABMs such as those presented by Weisberg and Muldoon (2009) or Thoma (2015) are used to support the idea that at least moderate levels of creativity are indeed conducive to the discovery of epistemically significant truths. In this sense, they offer support to claims such as those made by Currie (2019b) or Stanford (2019) that creativity is epistemically beneficial. However, as I have hinted before, such models conceive of creativity as exclusively rooted in individual behaviour. Moreover, they are not 'neutral' on the initial level of group-level creativity that is assumed in the set-up of their models. On the contrary, they assume approximately maximally creative scientific communities, since they disperse individual across the epistemic landscape.

My model, on the other hand, by introducing the cluster variable, allows me to consider different avenues for achieving group-level creativity. In this way, this model allows me to explore the interaction between group-level and individual-level creativity. With this in mind, I will suggest that individual creativity is not as beneficial as is suggested by previous authors. Rather, the extent to which a certain community is initially dispersed across the epistemic landscape plays a major role in determining epistemic outcomes, and relative to that the benefits of individual creativity are significantly smaller.

In order to support this hypothesis, let me present the basic simulation framework which I will use. I consider a community of scientists exploring the epistemic landscape, and I check the epistemic consequences of different creativity

arrangements (both at the individual level and at the community level). The community I consider is made up of a fixed number of 400 scientists[28], and I consider two variations in the model. Firstly, I vary the composition of that population of scientists in terms of the number of conservative and creative scientists in it. This dimension along which the parameters of the model are changed concerns individual-level creativity. Since the population is of a fixed number, the more conservative scientists there are, the fewer creative scientists there are, and *vice versa*. Given computational and time limitations, I do not consider all possible ratios of [number of conservative scientists]:[number of creative scientists]. Instead, I start with a population exclusively made up of conservative scientists, and I increase the number of creative scientists (and therefore decrease the number of conservative scientists) 25 at a time. I thus consider 17 different ratios of [number of conservative scientists]:[number of creative scientists].

The second dimension of variation concerns the initial degree of group-level creativity of the scientific community. For each of the 17 variations of the population composition (i.e. variations of individual-level creativity), I vary the number of clusters across which the population of scientists is initially distributed. I increase the number of clusters by a factor of 3 each time, starting with only one initial cluster, up to 243 clusters, and I consider one final scenario of maximal dispersion, which is equivalent to having 400 initial clusters (where each cluster is made up of only one scientist). The higher the number of initial clusters, the more creative the community is initially, as it is more widely dispersed across the landscape from the start. In terms of how initially

[28] This is in line with previous simulations, such as Weisberg and Muldoon's (2009) and Thoma's (2015).

dispersed the scientific community is across the landscape, I thus consider seven different scenarios (namely, an initial distribution of scientists across 1, 3, 9, 27, 81, 243, or 400 clusters).

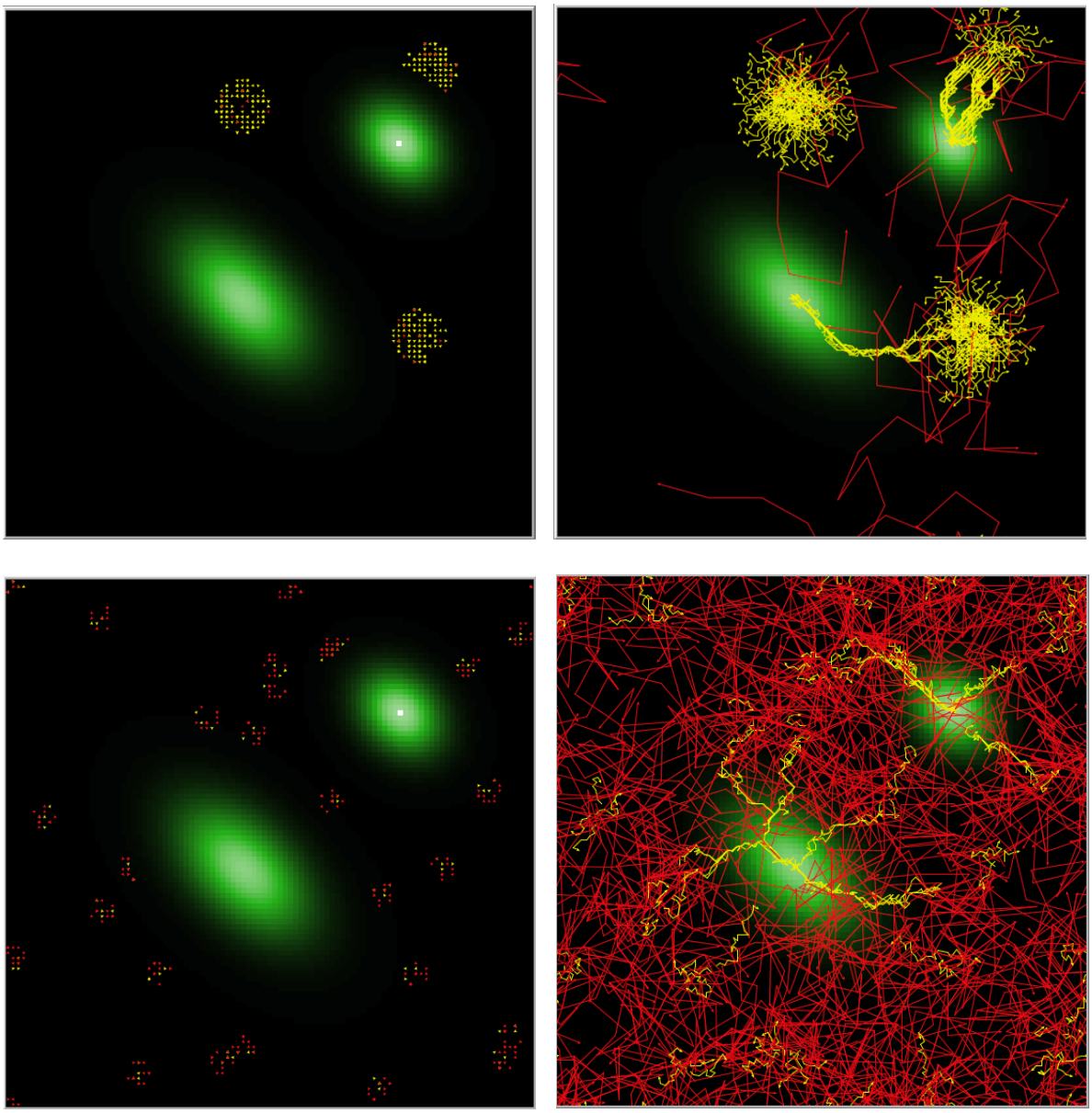


Figure 6. As an illustration of the kind of variations explored in the model, snapshots of two different simulations are presented. The top row presents a simulation of a population initially dispersed across three clusters and made up of 375 conservative scientists (yellow) and 25 creative scientists (red). The bottom row presents a simulation of a population initially dispersed across 27 clusters and made up of 100 conservative scientists (yellow) and 300 creative scientists (red). For each row, the picture on the left is a snapshot of the simulation at cycle 0 (i.e. how the simulation is set up before scientists start exploring the landscape), and the picture on the right is a snapshot of the same simulation at cycle 500 (i.e. the point at which the simulation stops and the epistemic performance of the population is evaluated). The lines represent the paths of the scientists up to that point in time, and the ticks represent the scientists.

For clarity, let me fix some terminology. I will refer to a given unique specification of the model as a *simulation scenario*. Simulation scenarios are distinguished from each other in terms of the combination of parameter specifications of the model variables. So, for example, two simulations which have the same population composition (in terms of how many creative and conservative scientists there are in them), but which start dispersed across a different number of clusters, are instances of different simulation scenarios. Similarly, two simulations dispersed across the same number of clusters but whose the populations have different compositions are still not instances of the same simulation scenario. I refer to simulation scenarios as distinct from simulations. I will use the latter term to refer to individual instances of a specific simulation scenario, since I will run each simulation scenario multiple times.

I will thus consider all combinations of the variations in each of individual-level creativity and initial group-level creativity profiles. Therefore, I will simulate a total of 119 different scenarios[29]. Each simulation scenario is run a total of 500 times[30]. I then aggregate the data from each one of the 500 simulations from each one of the 119 simulation scenarios. This allows me to assess and compare the epistemic performance across different scenarios.

Finally, I will refer to *cluster groups*. A certain cluster group specifies the group of simulations in which agents are initially dispersed across the same number of clusters.

[29] That is the number of variations in individual-level creativity (17) multiplied by the number of variations in initial group-level creativity (7).
[30] For a robustness check, each scenario was simulated 1000 times in parallel, and the results were identical.

So, for example, cluster group 3 refers to the simulations in which agents are initially dispersed across three clusters.

4.2.2. Experiments and results

4.2.2.1. Experiment 1 – three metrics

For the first set of simulations, I run each simulation of each scenario for 500 cycles, at which point the simulation stops and the epistemic performance of the population in that simulation is assessed. The data I consider varies across experiments. In this experiment I consider three distinct metrics aimed at measuring how competent the population is at discovering important facts about the topic under research:

a) *epistemic progress* – Weisberg and Muldoon (2009) define this as the number of visited patches (by the end of the simulation) with positive significance as a proportion of the total number of patches (visited or unvisited) with positive epistemic significance.

b) *epistemic progress (high)* – this is a slight variation of *epistemic progress*, in that it only takes into account patches of high epistemic significance. Thus it gives the number of visited patches (by the end of the simulation) with high significance as a proportion of the total number of patches (visited or unvisited) with high significance. I have chosen an arbitrary epistemic significance value of 500 above which patches are considered to be of high epistemic significance (considering that the patch at the highest peak is of epistemic significance equal to 750). The idea behind this is that some search strategies may be better at finding their way to particularly highly epistemically significant patches than others.

c) *accumulated significance* – given the concern that *epistemic progress* is not sensitive to the fact that some patches are much more significant than others, I also consider a measure of the total amount of significance uncovered by the end of the simulation, which is measured by the sum of the epistemic significance of all visited patches (at the end of the 500 cycles of each simulation). The specific value indicated by the metric of accumulated significance is meaningless, but that is irrelevant since I aim merely to make comparative claims regarding how much better a certain social arrangement is better at uncovering the truth than another one.

The findings regarding this experiment are illuminating. Across all three metrics, an increase in the number of initial clusters is strongly statistically significantly associated with an increase in epistemic performance. For example, a unit increase in the number of clusters (controlling for the number of individually creative scientists in the population) is associated with an increase in accumulated significance of 194.3 units (with a p-value of virtually 0). Although this number is not very meaningful, for a term of comparison note once again that the highest of the two peaks on the landscape is at a height of 750 units.

An increase in the number of creative scientists in the population has more ambiguous consequences. While it is associated with an increase in both *epistemic progress* and *epistemic progress (high)*, it is associated with a decrease in *accumulated significan*ce. These effects, albeit small, are all statistically significant with very low p-values.

In any case, the aggregate data is more illuminating once it is segmented by cluster group. Among scientific communities which are initially relatively conservative at the group-level (i.e. communities which are initially dispersed across a relatively low number of clusters), the general trend is that increases in individual creativity lead to an improvement in epistemic performance. However, when scientific communities are already widely dispersed, increases in individual creativity are not at all clearly associated with a better epistemic performance. In fact, with regards to *epistemic performance (high)* and *accumulated significance*, an increase in the number of creative scientists in the population leads to lower levels of the respective epistemic metric. Moreover, even within the set of simulations in which individual creativity is overall epistemically beneficial, there often are diminishing returns to individual creativity. These trends are apparent in figures 7, 8 and 9 below, each of which refers to one of the three epistemic metrics under consideration. For all three line graphs, the data is obtained by generating the average epistemic performance for each one of the 119 different simulation scenarios. A similar analysis was conducted using median, rather than mean, values, and the results were nearly identical.

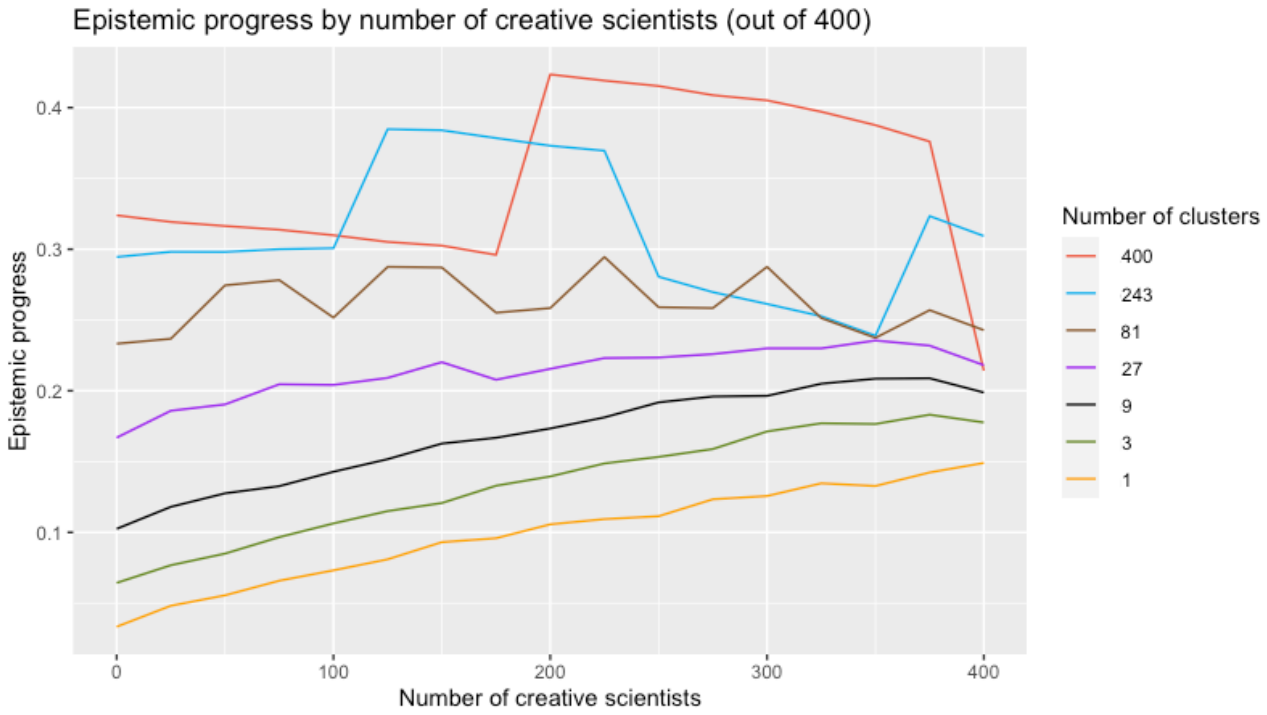


Figure 7. Average *epistemic progress* at time (cycle) = 500 by number of creative scientists and by cluster group. The number of conservative scientists is given by (400 – [number of creative scientists]).

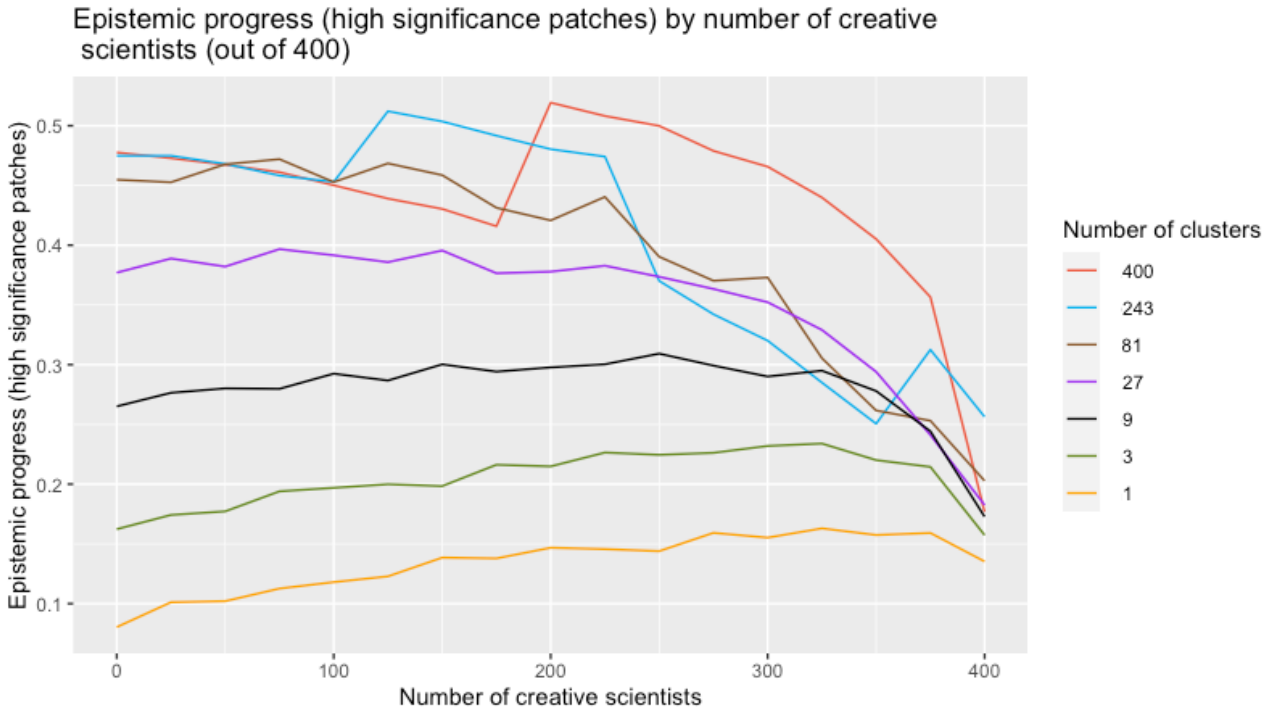


Figure 8. Average *epistemic progress (high)* at time (cycle) = 500 by number of creative scientists and by cluster group. The number of conservative scientists is given by (400 – [number of creative scientists]).

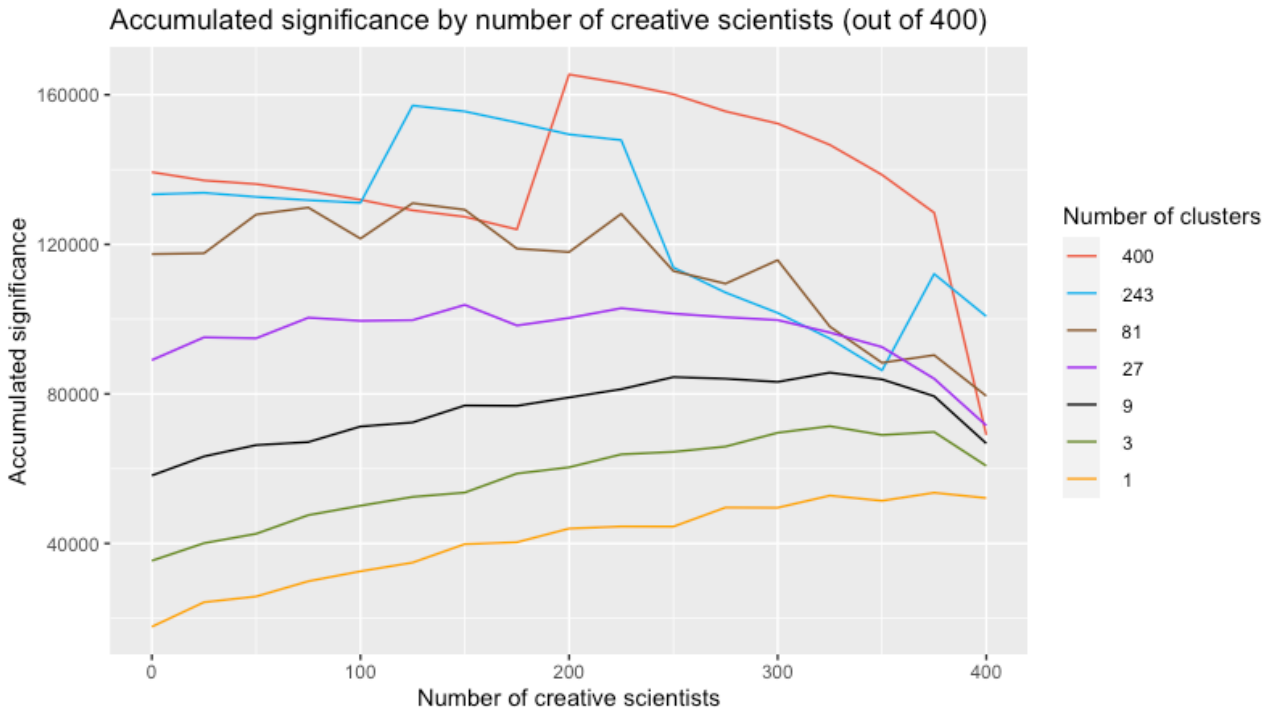


Figure 9. Average *accumulated significance* at time (cycle) = 500 by number of creative scientists and by cluster group. The number of conservative scientists is given by (400 – [number of creative scientists]).

Across all three indicators, I find that an increase in the number of individually creative scientists in the population is only unambiguously beneficial when (a) scientists start off from a relatively conservative disposition at the group-level *and* (b) there is a low number of individually creative scientists already in the population. Moreover, the following statement holds about some populations:

> No increase in the number of creative scientists (for a population of a fixed number of scientists) is as epistemically beneficial as a sufficiently large increase in the extent to which that population is initially dispersed across the landscape.

For example, notice that, with respect to all three metrics considered, a population which is initially clustered around one cluster does worse than a population initially clustered around 27, 81, 243 or 400 clusters, regardless of the composition of the populations in question.

Regarding the two metrics that are sensitive not only to whether visited patches are significant, but also to the *magnitude* of visited patches (i.e. *epistemic significance (high)* and *accumulated significance*), I also find that there is a point at which more (individual) creativity is too much creativity. This is observable by the downward sloping ends of the lines on the graphs.

These results lend at least some support to the claim that individual creativity matters less the more initially creativity populations are at the group-level. In fact, individual creativity can be detrimental to the discovery of significant facts about the topic in question. A certain minimum threshold level of individual conservatism appears to be welcome.

Finally, it should be acknowledged that the graphs seem to suggest that the effects are safer to assess with simulations in which the populations are initially dispersed across lower numbers of clusters. The lines connecting the data regarding clusters 1, 3, 9, 27 (and, to a lesser extent, 81) are generally smoother and have clearer trends than the lines corresponding to clusters 243 and 400. It is unclear to me why this is the case, and more research should be done to understand this.

Before moving on to another experiment, remember that I considered not two, but three different types of individual scientists: creatives, conservatives, and controls. It is useful to have some benchmark against which to analyse the results, and that is why control agents – which do not take into account any social information – are useful. For the control simulation scenario, I consider the same setup in terms of population size (400 scientists) and variation of initial dispersion across the landscape (I consider 1, 3, 9, 27, 81, 243 and 400 clusters). The population, however, is always exclusively made up of control scientists. The results are summarised as below.

Clusters / **Metric**	1	3	9	27	81	243	400
Epistemic progress	0.06	0.11	0.17	0.26	0.33	0.36	0.38
Epistemic progress (high)	0.15	0.27	0.43	0.58	0.66	0.68	0.68
Accumulated significance	39,610	71,344	114,249	158,308	184,676	192,513	196,003

Table 1. Average performance of control population by number of clusters from which the population begins exploring the landscape.

The general trend is that, at the lowest cluster groups, the control population is outperformed by some non-control populations of the corresponding cluster group. However, if the control population is sufficiently widely dispersed at the beginning of the simulation, then it outperforms (on average) any of the considered non-control

populations of the same cluster group. I will consider the potential implications of these results in the next chapter, in which I will discuss the results more broadly.

4.2.2.2. Experiment 2 – time efficiency

Another aspect of interest which epistemic landscape ABMs allow one to explore is the ability of a research community to discover the peaks of the landscape. After all, these represent the local points of maximum epistemic significance. As such, one can investigate how effective and efficient different scientific communities are at finding the most significant and illuminating facts about the topic they are researching.

In this experiment I explore precisely this aspect. I measure how many cycles of the simulation are needed in order for both peaks of the landscape to be visited. If, by the end of an arbitrarily large number of 40 thousand cycles, at least one of the peaks has not yet been visited, then that population is deemed incapable of exploring the landscape widely enough that it would be able to reach both peaks. I thus consider two main aspects. Firstly, I consider what proportion of the simulations run is able to reach both peaks. This gives me an indicator of the probability that a scientific population with a certain individual- and group-level creativity profile is able to visit both peaks. Secondly, I consider the time (in terms of simulation cycles) that it takes for a population to reach those peaks.

This experiment thus adds some new information and overcomes some limitations of the previous experiment. Firstly, it does not depend on an arbitrary cut-off time, at the end of which some indicators are measured. Rather, time is here one of the main

indicators. To be sure, I still establish a cut-off point at which the simulation stops in case the population has not yet reached both peaks, but this cut-off point is set at so large a value that, in the cases in which it is not reached, it means that that population would not be able to reach both peaks anyway, let alone at any realistic time-frame. In order to minimise the influence of these extreme cases on the final results, I consider the median (rather than mean) time taken by a population of a given simulation scenario[31]. Secondly, the focus on the discovery of the peaks captures an important aspect of scientific discovery as being directed towards finding out the most promising and significant approaches to a scientific problem. This element is already captured to some extent by *epistemic significance (high)* and *accumulated significance* in the previous experiment, but a focus on the two peaks permits gauging how competent a community is in climbing the hills of the landscape, which reveals how time efficient that community is in finding ever more significant approaches until it reaches the peak.

Similarly to experiment 1, I consider fixed populations of 400 scientists, I apply the same variations in terms of individual-level creativity and initial community-level creativity, I combine all individual-level creativity profiles with all initial group-level creativity profiles, and I run 500 simulations of each simulation scenario. The results are as follows.

[31] In any case, in the majority of simulation scenarios, the mean and median points are rather close.

The population was able to visit both peaks before the 40-thousand-cycles cut-off in almost all simulations. Only in five out of the 119 simulation scenarios was this not the case.

Clusters / **No. of creatives**	1	3	9	27	81	243	400
0	**1.2%**	0%	0%	0%	0%	0%	0%
400	0%	**0.2%**	**0.2%**	**0.2%**	0%	0%	**0.2%**

Table 2. Failure rate by number of creative scientists in the population across cluster groups. I define the failure rate as the number of simulations as a proportion of the total of 500 runs of each simulation scenario in which the population failed to visit both peaks of the landscape by the end of 40,000 cycles. The rows corresponding to simulations with other numbers of creative scientists are not listed since in those simulations the failure rate was 0% across all clusters. The cells corresponding to non-zero failure rates are highlighted in bold.

The virtual entirety of simulations was successful in reaching both peaks. However, the time taken for this to be achieved varied substantially across groups, as the figure below reveals.

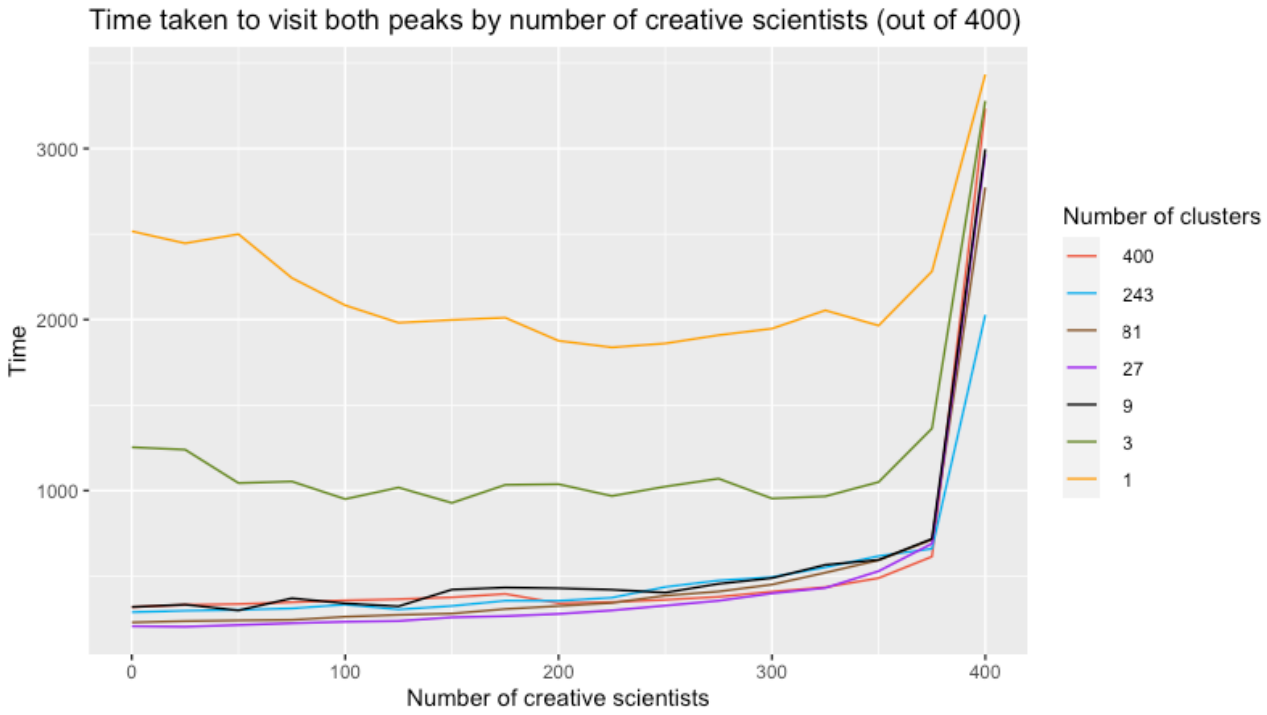


Figure 10. Median number of cycles (as a measure of time) taken to visit both peaks of the landscape by a scientific community, by number of creative scientists and by number of initial clusters.

This data thus suggests some relevant patterns. Firstly, a maximally conservative community (400 conservative scientists initially gathered around one single cluster) has the highest failure rate in reaching both peaks compared to the other scenarios simulated. This lends support to the idea that creativity is (or can be) epistemically beneficial. However, it also becomes apparent that increases in the number of individually-creative agents in the population most of the time does not do much to reduce the time taken to reach both peaks. The only visible benefits of such increases are in populations which begin exploring the landscape from either one or three clusters, and which have relatively low numbers of creative scientists in them.

Even these benefits are quite modest when compared with the benefit that comes with increasing the number of initial clusters. Notice, for example, how much more effective an increase in the number of initial clusters from one to three is in decreasing the amount of time taken to reach both peaks than an increase in the number of individually creative scientists in *any* cluster group. To be sure, when the community is sufficiently dispersed, further increasing the number of clusters does not seem to help much. But in these cases, neither does it help to increase the number of creative scientists in the population. In fact, there can be 'too many' individually creative scientists in a population, especially at the point where there are no conservative scientists left. This can be observed by the steep increase in the time taken at the end of all lines in the graph.

This experiment, too, seems to lend support to the hypothesis that increases in the number of individually creative scientists in the population are not particularly effective relative to increases in dispersion. Once a given population is dispersed widely enough, increases in the number of creative scientists in the population bring no benefit and become, at some high enough point, detrimental to the discovery of both peaks in the landscape.

As in the previous experiment, let us consider the aggregate data for the control population (as before, the control population is entirely made up of control scientists). Below in table 3 is the summarised data.

Clusters / **Metric**	1	3	9	27	81	243	400
Failure rate	0.2%	0%	0%	0%	0%	0%	0%
Time taken (median)	7,339	2574	578	279	222	220	213

Table 3. Performance of control population by number of clusters from which the population begins exploring the landscape.

While the control population virtually never failed to find both peaks, its median times are in most cluster groups within the performance range of the non-control populations of the corresponding cluster group. However, when it comes to especially low initial numbers of clusters, controls do significantly worse than all or almost all non-control populations of the corresponding cluster group. For example, in terms of median values, the performance of the control population when it starts exploring the landscape gathered around one cluster only (7,339 cycles) is more than double the

time taken by the population made up of 400 creative scientists initially gathered around one cluster (under 3,500 cycles), which is the worst performing population of that cluster group. As in the previous experiment, I will leave further reflections to the next chapter.

4.2.2.3. Experiment 3 – depletable landscapes

I mentioned previously that one question that is raised by modelling scientific research as the exploration of an epistemic landscape is the level of generality at which a research topic and the approaches to study that topic are represented by the landscape and the patches of the landscape. For example, perhaps the landscape represents a rather narrow topic, such as the topic for a special issue of a scientific journal. Or maybe it more aptly represents a much broader conception of 'research topic', for example an entire discipline. As the number of a patches making up the landscape is fixed, this also affects what one think about what one patch represents. If one's conception of the landscape as a research topic is narrower, patches will be distinguished from each other along more finely grained lines too.

This, in turn, affects the plausibility of other modelling assumptions. In particular, the assumption that the epistemic significance of approaches remains constant regardless of how many times that approach has been tried in the past might be a plausible approximation when approaches are broadly conceived. As mentioned in section 3.2.1, if approaches are broad enough, then the fact that a given scientist adopted it and reached a certain discovery does not mean that that is the only discovery of that level of significance that that approach enables to find. However, if approaches are more

narrowly defined, then it becomes more plausible to assume that each approach gives access to relatively few significant discoveries. The extreme case would be one in which each patch corresponds to one and only one scientific paper. In these narrower cases, since significant facts cannot be discovered twice and only the first time has impact, it is more plausible to assume that patches are somewhat 'depletable'. That is, the potential of a specific approach for uncovering significant facts about the topic in question in expectation decreases for each time that that approach has been tried.

In order to attend to this consideration, I consider different depletion rates: 1.2, 1.4, 1.8 and a more extreme rate of 5. A depletion rate of, e.g., 1.4 means that the epistemic significance of a certain patch is divided by 1.4 each time that patch is visited. Suppose that a given patch on the landscape is originally of epistemic significance 280. If the depletion rate is of 1.4, this means that after being visited, the epistemic significance associated with that patch becomes 200 (280/1.4 = 200). If that patch is visited a second time, then the epistemic significance associated with that patch becomes approximately 143, and so on and so forth. For each depletion rate scenario, I set up the model as in previous experiments. That is, I consider a population of 400 scientists, and I vary its composition in terms of how many creative scientists there are out of those 400 scientists. I also vary the number of clusters across which scientists are initially dispersed in the usual way.

In order to compare the epistemic performance of different populations in different scenarios, I use *epistemic progress* as the metric of interest. As I have pointed out, this measure is not sensitive to the fact that epistemically significant approaches may

vary in terms of the extent to which they are significant. However, since in this experiment the epistemic significance of patches is itself variable, *epistemic progress* allows for a simpler, more straightforward comparison which will suffice for my purposes.

Instead of fixing one exact interpretation to the epistemic landscape and patches, this variation in depletion rates allows me to test the extent to which the patterns observed in previous experiments are stable across different interpretations of the landscape. I find that, controlling for the number of creative scientists in the population, an increase in the number of clusters is robustly associated with an increase in the epistemic progress at the end of 500 cycles of the simulation across all depletion scenarios. The magnitude of this effect, moreover, is rather stable too, decreasing only slightly as the depletion rate increases. The lesson is intuitive – the benefits of having a scientific community adopting a diversity of approaches rather than focusing on a narrow set of options hold at the narrower as well as at the broader interpretations. On the other hand, the effect of increasing the number of creative scientists in the population while controlling for the number of initial clusters is more volatile across different depletion scenarios. While it is positive in some, it is negative in others. The magnitude of the effect of a marginal increase in the number of individually creative scientists, moreover, is relatively small (both when it is positive and when it is negative) when compared with the magnitude of the marginal effect of increasing the number of clusters.

Across the four depletion scenarios, the closest the marginal effect of increasing the number of individually creative scientists in the population gets to the marginal effect of increasing the number of initial clusters is when the landscape is highly depletable (depletion rate = 5). Even in this scenario, the marginal benefit of individual creativity is still twelve times lower than the marginal benefit of community-level creativity. If experiment 1 (where the depletion rate equals 1) is included in the depletion scenarios under consideration, the marginal benefits get closer, but even in this scenario the benefit of marginally increasing the number of initial clusters is four times higher the benefit of marginally increasing the number of creative scientists. Table 4 provides an illustration of how the results of this experiment were extracted and analysed with reference to one specific depletion scenario. Similar analyses were conducted for the remaining depletion scenarios.

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.810e-01  6.406e-04  282.48   <2e-16 ***
clusters           4.024e-04  2.174e-06  185.08   <2e-16 ***
number_of_creatives 3.380e-05  2.537e-06   13.32   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0758 on 59497 degrees of freedom
Multiple R-squared:  0.3666,    Adjusted R-squared:  0.3666
F-statistic: 1.722e+04 on 2 and 59497 DF,  p-value: < 2.2e-16
```

Table 4. A table summarising the regression results for depletion scenario where depletion rate is 5. The model on the basis of which the results were extracted is a linear model which controls for both initial number of clusters and number of creative scientists in the population. The estimates inside the highlighted rectangle indicate the estimated effect of marginally increasing the variable on the left, while controlling for the other variable. As mentioned, the estimate of the effect of marginally increasing the number of clusters is roughly 12 times higher than the estimate of the effect of marginally increasing the number of creative scientists.

These patterns lend some further confidence to the overall picture which I have been suggesting throughout this thesis – while creativity is an important and beneficial trait for an epistemic community to have, it may not be very beneficial if brought about via individual behaviour. An increase in the level of dispersion (regardless of the composition of the population) yields more significant and more reliably beneficial epistemic effects. This seems to be a relatively stable finding across different assumptions about the depletion of epistemic landscapes and their patches.

4.2.2.4. Experiment 4 – landscape size

Varying the size of the epistemic landscape is interesting for two main reasons. On the one hand, it allows one to test the robustness of the findings regarding the 'standard' size used in the previous experiments; on the other, it enables exploring hypotheses regarding aspects of scientific research that may be captured by the landscape size. In particular, I suggest that variations in the landscape size (in which the size of each approach remains the same and the proportion of significant patches to total number of patches diminishes with size) pick up the fact that some research topics might be more difficult than others. A very large landscape with a proportionally smaller number of significant patches represents a topic in which many of the potential approaches to the topic are sterile. A relatively smaller landscape, on the other hand, is one in which it is more likely that a scientist randomly latches onto something relevant for the discipline.

This, I suggest, relates to the point made by Currie (2019) which I highlighted in the previous chapter. I mentioned that Currie's argument for creativity is more nuanced

than usual. He does not claim that creativity is always beneficial. Rather, he suggests that creativity may be particularly helpful in contexts in which there is very scant evidence and high uncertainty about a certain topic. This has some relation to the difficulty of the topic. If there is weak evidence about a certain research topic, scientists and scientific communities are more ignorant about which specifications of the dimensions that determine an approach are potentially useful for understanding the topic – for example, how many background theories are potentially fruitful, or whether certain experimental methods are applicable or not. In difficult topics, then, it is natural to assume that scientists try (and fail) more – thus the number of potential approaches is larger, and the proportion of significant approaches to the total number of approaches is lower.

For time and computational limitations, I did not run all 119 different simulation scenarios across different landscape sizes. Instead, I considered two populations of scientists. One is made up of 350 conservative and 50 creative scientists, and the other is made up of 50 conservative and 350 creative scientists. I call the former the *moderately conservative population*, and I call the latter the *moderately creative population*, with reference to the composition of the population in terms of individual creativity. I vary the number of clusters across which each population is distributed in the usual way. Finally, for each population and each cluster distribution, I vary the landscape size, by considering square landscapes the sides of which are size 60, 80, 100 (the original size), 120 and 140 patches long. As the size increases, the proportion of significant patches to the total number of patches decreases. Each simulation is run 500 times. At the end of 500 cycles of each simulation, I consider the usual three

epistemic metrics: *epistemic progress*, *epistemic progress (high)* and *accumulated significance*. The comparison between only two different population compositions means that this experiment is less complete than the previous ones. However, the motivation is that each of the two populations may be more generally representative of the trends among populations tendentially made up of conservative scientists and tendentially made up of creative scientists, respectively. In future research, though, a more complete investigation is welcome to verify the stability of the patterns observed in this experiment.

Two main patterns emerge from the experiment, and they hold regarding all three metrics considered. As an illustration, I include the results regarding *epistemic progress (high)*, but the qualitative results are very similar concerning the two other metrics.

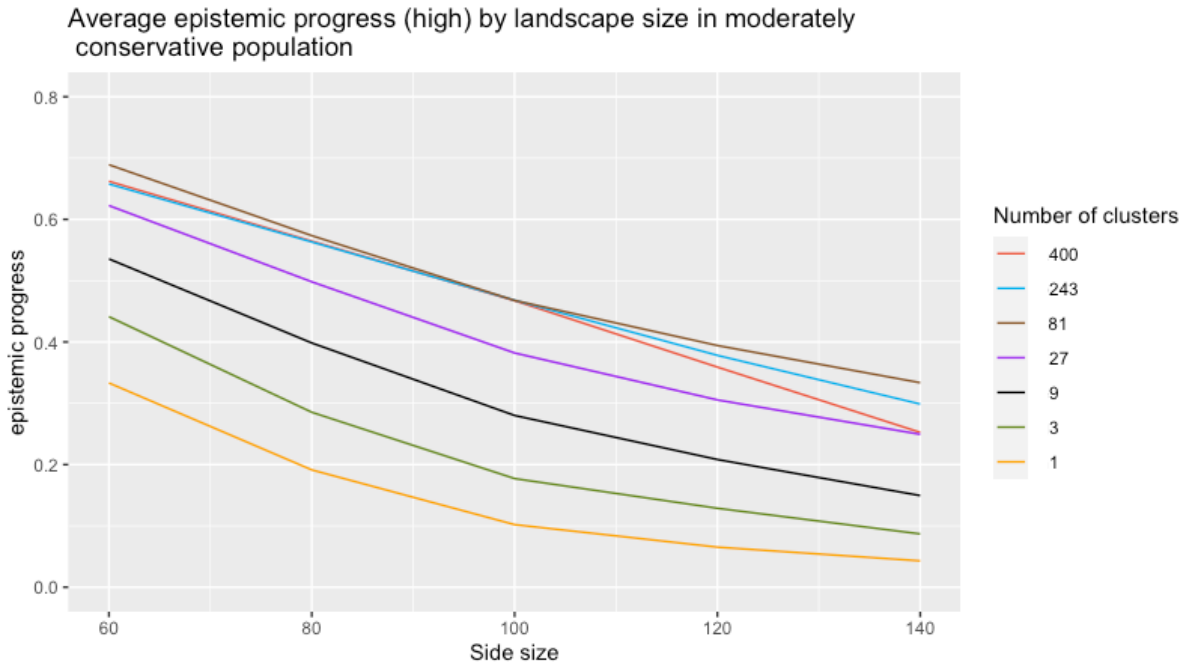


Figure 11. Average epistemic progress (high) for the moderately conservative population by landscape size and cluster group.

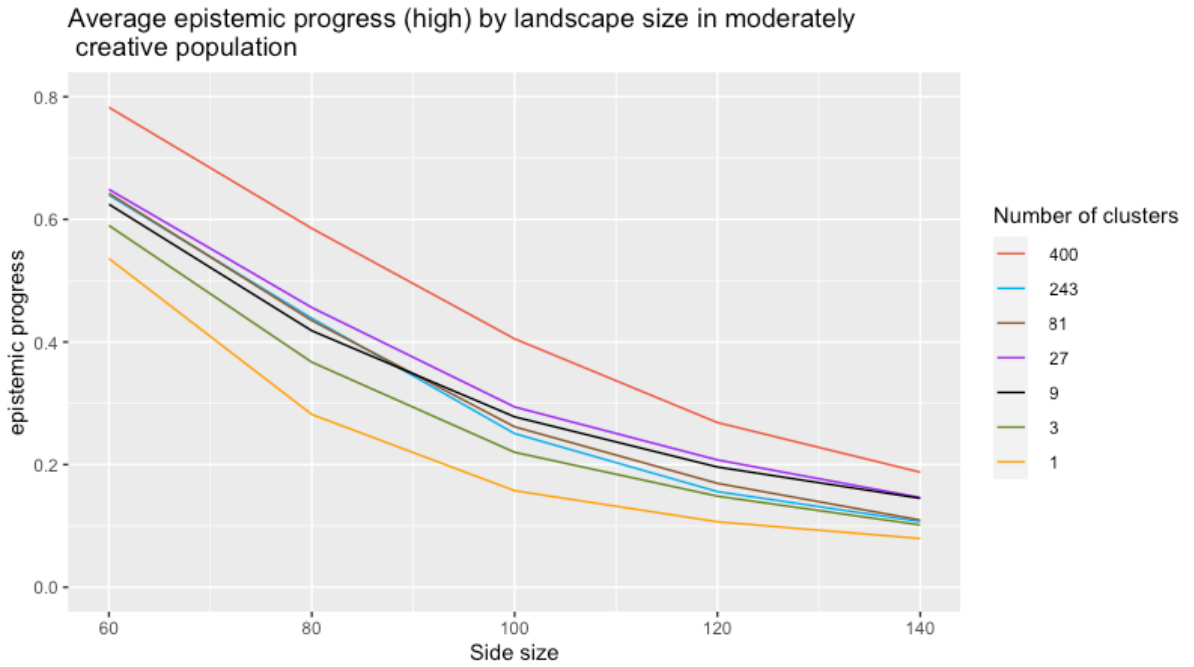


Figure 12. Average epistemic progress (high) for the moderately creative population by landscape size and cluster group.

Let me then present the two patterns that can be visualised above. Firstly, the general finding that an increase in the number of initial clusters is epistemically beneficial, especially at low levels of initial dispersion, generally holds across landscape sizes, both among the moderately conservative and the moderately creative populations. Once the initial dispersion is sufficiently high – which tends to be at cluster groups 27 or 81 –, one sees that further increases in the initial dispersion may not be effective in further facilitating scientific progress. This is in line with the findings of the previous experiments, and in that sense reveals that the results are robust. While figure 12 actually shows that the cluster group 400 considerably outperforms the other groups, it should be remembered that the higher cluster groups display more unstable patterns as the population composition changes, and so these specific results regarding cluster group 400 may not be very significant. Moreover, experiment 2 shows that in terms of time efficiency, the cluster group 400 is generally not the top performing. Cluster

groups of 27 or 81 clusters, however, tend to perform rather well and their trends seem to be more reliable, as made clear especially in experiments 1 and 2.

The second pattern observed is that, controlling for the number of clusters, an increase in the landscape size has a more negative impact in the epistemic performance of the moderately creative population than of the moderately conservative one. This is true of all three epistemic metrics, and it holds when assuming both a linear as well as a decreasing logarithmic model of the simulation data. As a result of this, while most cluster groups of the moderately creative populations may perform better than their moderately conservative counterparts among the smaller sized landscapes, the pattern is reversed among the relatively larger landscapes. For example, note from figures 11 and 12 that when the landscape side is 140 patches long (the largest considered), the moderately creative population only performs clearly better than the moderately conservative one among the simulations of cluster group 1. The performances of cluster groups 3 and 9 are very similar across both populations, and by cluster group 27 the performance of the moderately conservative population is clearly superior to that of the moderately creative one.

Linking the results back to the intended interpretation, it becomes clear that Currie's thesis that creativity is particularly beneficial in evidentially weak epistemic contexts is cast into doubt as far as *individual* creativity in concerned. While an increase in the initial community-level creativity is generally beneficial across all landscape sizes, individual-level creativity does not seem to clearly have this effect. As the landscape size increases, the performance of the moderately creative population declines (in

absolute and relative terms) more than that of the moderately conservative one, and in the larger landscapes the moderately conservative population generally performs at least as well as the corresponding cluster group of the moderately creative one (with one exception being cluster group 1, as highlighted). As such, the results suggest (even if only on a first approximation) that provided communities are sufficiently dispersed across the landscape, individual-level creativity may actually be an epistemically harmful trait for scientists to display as the topics that they research get more difficult. In the next chapter, I will suggest one explanation for this phenomenon.

Chapter 5. Discussion, limitations, and ways forward

Throughout the thesis, I have mentioned and discussed some aspects of the simulations I have run, but I have also left some of that discussion out. In this chapter, I will discuss the results more broadly, draw the main lessons, and address potential issues with the model I have developed and the simulations I have run. I will also consider some of the policy and institutional implications of my findings.

5.1. Main lessons

At this point, let me draw the main findings of all experiments presented and ask how they fit the picture I sketched for the potentially optimal arrangement of a predominantly conservative scientific community at the individual level but sufficiently widely dispersed across the landscape.

I return to the question in the title of section 4.1. Does science need creativity? What one sees is that an initial wide dispersion of scientists across the landscape is robustly associated with a better epistemic performance. The trends are rather reliable and the overall mean and median effects of increasing initial dispersion across the landscape reflect a relatively large, positive difference to the performance of the population. This is the case considering all four major metrics I considered – epistemic progress, epistemic progress (high), accumulated significance and time efficiency to reach both peaks. Moreover, these results (at least those concerning the first three metrics) generally hold in different modelling scenarios. As per experiment 3, they hold across various conceptions of how depletable the landscape is, and, by experiment 4, they hold in different landscape sizes.

In this sense, the model quite conclusively suggests that creativity often is, indeed, epistemically beneficial. But those benefits come about more reliably from a wide dispersion of the population across the landscape (regardless of population composition) than from a community that is significantly composed of individually creative scientists. When it comes to individual-level creativity, the basis for arguing that it is epistemically beneficial is much weaker. Indeed, as long as the community is sufficiently widely distributed, individual creativity has very little to improve, and may actually at some levels impede the efficient discovery of significant scientific facts. Consider, for example, experiment 2 regarding how time efficient different populations are at reaching both peaks of the landscape. In this set of simulations, increases in the number of individually creative scientists in the population only lead to speedier discovery times in quite a small subset of the simulation scenarios. And even in those scenarios, a marginal increase in the initial number of clusters would have a greater beneficial effect than that a marginal increase in the number of individually creative scientists.

To be clear, this does not mean that increasing individual-level creativity is always futile. It does seem to have a considerable, positive epistemic impact particularly when communities are initially highly concentrated in the epistemic landscape. But it is also a stable pattern that this benefit is not as impactful as the benefit brought about by a moderately small increase in the number of clusters across which the community is distributed (for example, from 1 to 3 or 9). This fact gains particular relevance when one considers the main reasons that motivate the investigation into this topic.

As I have pointed out, many claim that science is too conservative, and that this excessive conservatism prevents scientific communities from performing at their best epistemically speaking. The results of the simulations presented support the second half of the previous sentence. What of the first half – that science is too conservative? We must first ask if it is meant that science is too conservative at the individual level or, rather, at the community level. If the point is that, individually, scientists are too conservative even if they may on aggregate try various different approaches to learning about a topic, then my model suggests that this thesis is misguided. When controlling for cluster group (that is, when controlling for the extent to which scientists try different approaches between each other), note that an increase in the number of individually creative agents in the population is associated with worse epistemic outcomes according to at least some of the metrics considered. These negative effects are especially strong when the population is already relatively widely spread across the landscape.

The alternative interpretation is that the worry with extreme conservatism is taken from a community-level standpoint. In this case, the worry is that scientists as a community are too concentrated around a narrow set of approaches, that they follow each other too much and do not venture into new, different ways of studying a topic[32]. If this is what is meant by the assertion that science is too conservative, then, indeed, an increase in the number of individually creative scientists in the community may make a positive difference. The simulations suggest that an increase in individual level

[32] As I had mentioned in chapter 3, note that, on my model, if community-level persists in the long run this implies that the population is mostly composed of individually conservative scientists.

creativity has a relatively stable and robust, positive impact in the epistemic performance of scientific communities which are concentrated around a low number of patches in the landscape and which are made up of a relatively high number of conservative scientists. But it is also among these simulation scenarios that a relatively modest increase in the dispersion of scientists across the landscape has a greater (positive) effect in the epistemic performance of the population than any increase in the number of individually creative scientists in the population. And as the dispersion increases, the positive effect of increasing the number of individually creative scientists wanes. Take figure 8 (from experiment 1), for example. The best performing population of cluster group 1 is roughly as good as the worst performing population of group 3. In other words, no change in the composition of a scientific population grouped around one cluster is as epistemically beneficial (in terms of epistemic progress (high)) as an increase in the dispersion of the population from 1 to 3 initial clusters. And, once the population is dispersed across 3 clusters, the benefit of marginally increasing the number of individually creative scientists wanes (compared to the magnitude among cluster group 1). For example, if one considers epistemic progress (high), the estimated benefit of marginally increasing the number of creative scientists in the populations among cluster group 3 is roughly 43% lower than the estimated benefit among cluster group 1.

As such, both interpretations suggest that a focus on community-level creativity would be a more effective route to improve the epistemic performance of an overly conservative epistemic community. This supports my general conclusion that the answer to the question of whether science benefits from creativity is more nuanced

than the usual arguments presented in favour of creativity. In particular, the focus on creativity as an advantageous individual characteristic is somewhat misguided. Either individual creativity is actually not beneficial or, in the scenarios where it is potentially beneficial, it is not the most effective way of achieving such goals. One might suggest, however, that increasing the number of individually creative scientists in the population is the only *feasible* way of achieving creativity at the group-level; that the 'pockets of conservatism' model mentioned in section 3.3.2., whereby community-level creativity is achieved even though individuals behave (mostly) conservatively, is just overwhelmingly more difficult to achieve. Of course, if one aims to argue for this, a convincing story explaining why this is the case should be provided. No such story is yet available, in part because the distinction between community- and individual-level creativity is not prominent in the literature. For this reason, I will not engage directly with this hypothesis. However, I will consider in the next section some practical factors that may indirectly bear on this issue.

5.2. Towards an optimal arrangement – institutional reality check

In this section, I wish to be a bit more specific about what is suggested by the considerations just made in conjunction with attention for the institutional setup of science. In particular, my claim is that, all-things-considered, the 'pockets of conservatism' model mentioned approximates the socio-epistemic ideal which should be striven for. As a reminder, what I mean by 'pockets of conservatism' is an arrangement whereby the scientific community is made up of a moderately high number of subcommunities, and where these subcommunities are made up (or overwhelmingly made up) of conservative scientists. Thus each subcommunity (each

cluster) explores the epistemic landscape conservatively, attempting a narrow set of similar approaches and slowly trying to gradually improve them. However, the approaches adopted by each of these different subgroups are substantially different from one another, which renders the whole community creative as long as there are enough such subcommunites.

I will argue for this arrangement in two parts, which will explain, respectively, why the optimal arrangement is one in which the population is divided into 'pockets', and why these pockets are made up of conservative scientists.

5.2.1. Why pockets

Why does an optimal arrangement likely approximate one of 'pockets' rather than one whereby agents are maximally or near-maximally dispersed across the landscape? In order to answer this question, note that across the experiments, there seems to be a point (often at cluster group 27 or 81) beyond which, while further increasing the dispersion of the population may not be harmful, it does not bring many benefits either. Experiment 2 is particularly illuminating in this regard. Note, for example, that the quickest median performance of all simulation scenarios (including the control simulations) in experiment 2 is that achieved by a population initially dispersed across 27 clusters only. Thus scientific communities may not require a maximal initial dispersion across the landscape (in this case represented by the cluster group of 400) in order to achieve optimal or near-optimal epistemic performance. This implies that each cluster in the landscape will be composed of a few scientists, rather than just

one or two, thus constituting a 'pocket' of scientists and, as such, representing a subcommunity.

Surely enough, when it comes to the other experiments, the populations of cluster groups 27 or 81 are (on average) outperformed by a significant number of the populations of higher cluster groups. However, as I have highlighted, the results at those higher cluster groups do not follow a clear a trend and for that reason are perhaps not especially reliable. These dynamics at the higher cluster groups need to be understood in greater detail, but they perhaps reveal that at those higher levels of dispersion the epistemic benefits of creativity are more volatile. On the other hand, at moderately high initial dispersion levels (as characteristic of simulations like those of cluster group 27), the patterns are quite stable and the performance rather positive across all metrics. And, indeed, at such a cluster group increases in the number of individually creative scientists make little positive difference in the best of cases. The strong performance of such cluster groups is observed in the other experiments, and it is highly robust to changes in the composition of the population too.

5.2.2. Why conservatism

What benefit comes from having pockets of *conservative* scientists, specifically?

One important consideration concerns the incentive-structure of science. Science is often conceived as being governed by an incentive-structure based on credit (see Merton 1957). That is, scientists are credit-seekers, where credit is realised via peer-recognition, publications in prestigious journals, participations in important

conferences, etc. And, in particular, the attribution of credit is thought to be strongly linked to the impact of one's research (Heesen 20129). In the social epistemology of science, important results have been presented which point to the epistemic benefits of this alignment between the significance of a scientist's findings and the credit that is assigned to that scientist (see, especially, Kitcher 1990; Strevens 2003; Strevens 2006; Heesen 2017). While I am not going to analyse them closely, these are all independent results which offer support to the thesis that having credit directly proportional to epistemic significance promotes efficient epistemic outcomes in science. Yet, if individual creativity is to be promoted, such incentive structure would be changed. For example, it is plausible to assume that, in order to promote individual creativity, credit would have to go from being directly proportional to epistemic significance to, e.g., being exponentially related to epistemic significance. That would incentivise scientists more strongly towards seeking significant results, and, on approximation, that would incentivise them to forego the safety of conservative searches for the promise of a highly impactful finding. While it is not crucial for my argument that this assumption holds, the important point is that, in order to get more scientists to be creative, the structure of individual incentives would likely need to be significantly changed. But, as Heesen (2019) argues, it is not clear how it would be possible to achieve this without undermining the role played by the current incentive structure of promoting other efficient epistemic outcomes. Something would have to be sacrificed.

The good news is that individual creativity does not tend to play a very beneficial role among high and moderately high cluster groups. As I mentioned in the previous

chapter, more dispersed cluster groups tend to do better than less dispersed ones. But at moderately high levels of dispersion (such as cluster groups 27 or 81), individual creativity tends not to bring epistemic benefits, and in some circumstances is actually harmful for the discovery of significant facts. As such, we may have our cake and eat it: as far as individual-level creativity is concerned, it need (and perhaps should) not be promoted, and hence the structure that governs individual incentives need not be changed, which, in turn, allows for the efficient epistemic role played by the current incentive-structure to be preserved.

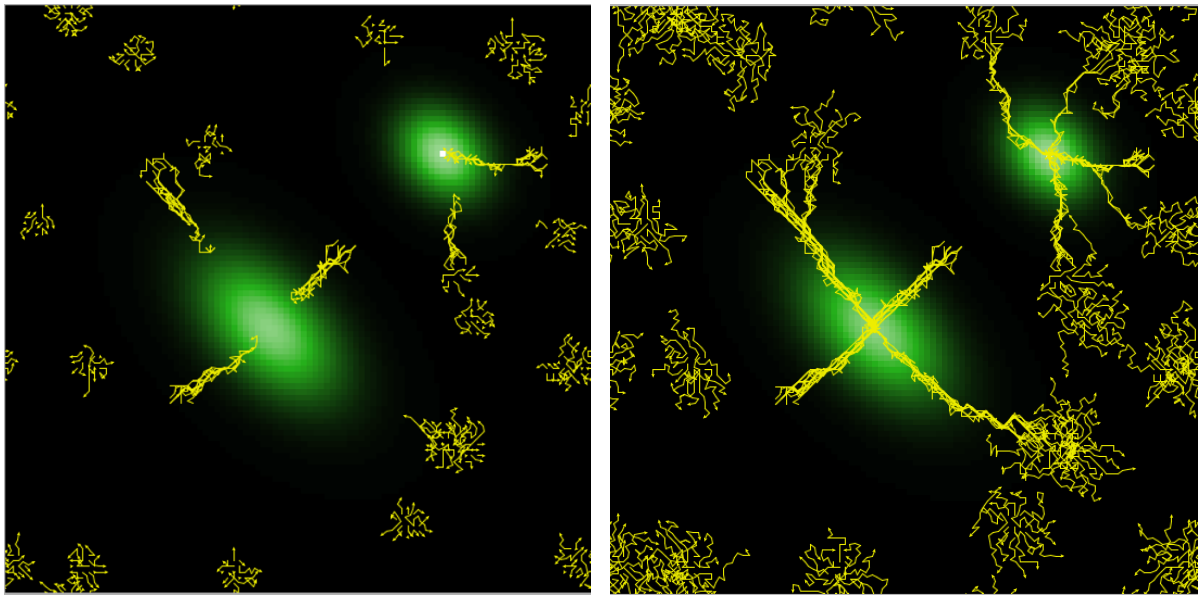


Figure 13. An instance of the 'pockets of conservatism' model: a population of 400 conservative scientists clustered around 27 subgroups after 100 cycles (left) and 500 cycles (right) of a simulation.

We thus end up with an ideal image of science which combines both community-level creativity and individual-level conservatism. In my model, such an ideal resembles the simulation represented in figure 13 above. Indeed, the fact that the 'pockets of conservatism' arrangement combines creativity and conservatism is likely what explains the epistemic success of the populations corresponding to this arrangement across the experiments of the previous chapter. That is because such populations are well-positioned to reap the benefits of community-level creativity while also taking

advantage of the benefits of conservatism, as hinted at the end of section 4.1. By being moderately highly dispersed across the landscape, these populations explore the landscape widely. This, as I have noted, represents communities which, collectively, try a wide variety of different approaches to studying a given topic of research. This allows such populations to uncover highly significant facts which they might otherwise not have discovered if they were more concentrated around a narrow set of approaches. However, the fact that each cluster – each subcommunity – is mainly constituted by conservative scientists means that the members of these subcommunities use the knowledge about current best scientific practice in order to adapt and improve their approaches in an efficient way. Deviations from this model would give up on at least one of the advantages. If the population is too concentrated over a small number of patches, this means that the scientific community is not considering a diverse enough set of potentially interesting approaches to understanding a topic. And if the population is heavily constituted by creative scientists, then useful knowledge regarding which approaches are fruitful for studying a certain topic would largely be wasted.

As I noted, this makes sense of the fact that the populations which match this model perform very well in relative terms across all experiments. Moreover, it might be particularly useful in understanding the pattern seen in experiment 4 that, as the landscape gets larger, and provided communities are sufficiently dispersed across the landscape, individual-level creativity is actually an epistemically disadvantageous trait for scientists to display. The intuition is that, as topics get more and more difficult, community-level creativity gets more and more beneficial, as it mitigates the difficulty

to predict which approaches are going to yield promising results. However, precisely because so many attempted approaches may be fruitless, it is also crucial for scientists to make good use of the existing knowledge regarding which approaches are productive. Individual-conservatism thus becomes more and more important as it guarantees that, once a significant approach is tried, others can take advantage of such an achievement and methodically tweak the approach to get more and more significant results. When landscapes are smaller, the hot searches of creative scientists might be more beneficial than when landscapes are larger, as the likelihood of randomly latching onto something promising is higher.

The conclusion that the optimal socio-epistemic arrangement approximates a 'pockets of conservatism' model has an intuitive appeal, too. It represents a community which is subdivided into smaller intra-collaborative groups, and as such it makes sense of the intuition that it is advantageous that individuals are inserted in subcommunities with other scientists working in closely related ways and who they communicate with and learn from primarily. Despite Goddard's inspirational story, scientists are not usually lone wolves, and, it turns out, this is for good reason.

5.2.3. Anti-conformity arguments revisited

Let me now review some of the arguments regarding why science is too conformist and how that is bad for the progress of science in light of this conclusion. For the sake of scope, it will be useful to focus on the accounts offered by Stanford and Currie presented in the last chapter. In particular, I highlighted three main factors. Firstly, peer-review and related publication norms which may filter out papers (thus rendering

them unsuccessful) based on the approach employed by those papers. Secondly, the hierarchical and centralised structured of science (in great part brought about by the phenomenon of Big Science). And, finally, the existence of strict disciplinary boundaries. My contention is that the identification of at least some of these factors as detrimental to the progress of science because they promote conservatism may be misguided. The 'pockets of conservatism' image will help me convey this point.

Take peer-review for a start. Peer-review, it is argued, disincentivises scientists from attempting relatively different approaches. But this is compatible with there being a community which, at the group-level, is creative. Provided there are enough pockets of conservatism, the solution space of a given topic broadly construed is still likely to be widely explored, even if, within each pocket, there are quite substantial conforming pressures. For example, some scientists working in enzymology (the study of enzymes' function and structure) may try and learn about the topic from a computational approach based on biophysical insights in quantum and molecular mechanics (Roston and Cui 2016), while others may approach it from a biochemical perspective, with a focus on real chemical reactions (Mitić and de Vries 2012). These are quite different approaches, in terms of the theoretical level of description on which the approaches are based, the experimental tools used, and the specific questions asked. The level of pre-publication scrutiny in each area is likely high, and detailed knowledge is required to estimate which experimental setups are promising and which are not. This might indeed lead to some conformity within each subgroup of scientists. Yet this fact is not too worrying given that the conformity effect is quite local. That is, it only applies within the different pockets of conservatism and it does not necessarily prevent new

subcommunities from being formed. A more problematic scenario would be if particular subfields would be prevented from flourishing due to the peer-review-led gatekeeping from other subfields – if work in real-life laboratory experiments regarding the mechanisms of enzyme catalyst were filtered out by reviewers working in computational enzymology. But it is not at all clear that such phenomena occur; and even if they did, the point is that as long as there are mechanisms which prevent or counteract such phenomena to be realised, peer-review as a local phenomenon is not as pernicious as it is sometimes purported to be.

Similar considerations apply to Currie's point about strict disciplinary boundaries and subsequent lack of interdisciplinary work. His argument is, mainly, that if interdisciplinary approaches are ignored, potentially promising areas of the solution spaced are left unexplored. Hard disciplinary boundaries may therefore prevent epistemically significant facts about the topic in question from being uncovered. Once again, I believe the results presented in this thesis suggest a more nuanced hypothesis. Extremely static subdisciplines and subcommunities which take no insights from other approaches into their research are indeed potentially problematic, in that they prevent the scientists from navigating the landscape even if only gradually. However, that is a very extreme picture of conservatism. A more plausible picture, suggested by the 'pockets of conservatism' model, is one in which the members of subcommunities are indeed individually conservative, but this does not imply that there are only a few subdisciplines producing research output. Interdisciplinary disciplines are themselves sub-fields which, in their interdisciplinarity, may incorporate high levels of individual conservatism. Take, for example, game theoretic political

economy. This subdiscipline incorporates today a highly institutionalised cluster of approaches. It features prominently in very reputable, academic journals such as the *Journal of Political Economy*, the articles of which are peer-reviewed and subject to a lot of pre-publication scrutiny. The people who adopt such approaches form communities whose members influence each other's approaches via articles, conferences, etc. It is, in this sense, an area which is likely subject to the conformist tendencies just Currie and Stanford worry about. And yet it is also highly interdisciplinary, drawing on insights from political theory, political science, economics, mathematics (game theory), sociology and behavioural psychology (see, e.g., Landa 2015). As such, communities may consist of a great number of conservative scientists and still reap the benefits of multidisciplinarity.

Perhaps Currie's suggestion is that current institutions of science are preventing such interdisciplinary pockets of conservatism from being formed in the first place. This is, in the end, a matter for empirical appreciation. Nevertheless, the more general takeaway from my findings is that the fact that the scientific subcommunity to which one belongs is relatively conservative does not imply that no interdisciplinary work can be done. Some such pockets of conservatism might just be those exploring interdisciplinary approaches. Moreover, from a more time-dynamic perspective, new interdisciplinary areas of the epistemic landscape need not be uncovered by individually creative scientists. Rather, through slow, gradual research, an initially small number of scientists may uncover promising paths towards interdisciplinary approaches, paths which are then followed by other scientists which might lead to the formation of relatively stable, locally conservative subcommunities. Indeed, this seems

to be a good approximation to how many of the interdisciplinary areas of research gained prominence.

From this perspective, more threatening to achieving community-level creativity might be the centralising and hierarchical structures fostered by Big Science. If a small minority of the scientific community is able to control where most of the resources necessary to conduct scientific research are directed, then this might indeed concentrate the scientific community around rather small subsets of all possible approaches to a question. The metaphor used by biologist and writer Aaron Hirsh in a *New York Times* opinion column, which Stanford cites (2019: 3923), is instructive here. He writes:

> 'There's something disturbingly hierarchical about the new architecture of the scientific community: what was before something like a network of small villages is today more like an urban high-rise, with big offices at the top and a lot of cubicles down below' (Hirsh 2009).

Such a hierarchy does indeed pose the danger of concentrating the scientific community around a narrow set of approaches. If those on top of the hierarchy have a very specific and limited idea of which approaches are epistemically promising, then this may result in a community with not enough pockets of conservatism. As such, the design of institutions and incentives which counteract the conforming effects of the structure of Big Science structures are likely desirable if one is ultimately concerned with the creativity of a given scientific community.

It appears, then, that at least some of the factors that have been pointed out as roots of an excessive, epistemically sub-optimal level of conservatism may not be as damaging as suggested. A main reason for this is that factors such as peer-review and relatively consolidated subcommunities and subdisciplines may result in individual conservatism, but do not necessarily result in community-level conservatism. However, the hierarchical and centralising dynamics of Big Science, because of their potential impact on a more macro level, pose greater threats for the achievement of a wide dispersion of scientists across the epistemic landscape. In general, then, the policy measures directed at promoting the progress of science should focus on increasing the variety of approaches that are adopted by the community as a whole, rather than fostering individual creativity itself.

The question of what specific incentives and institutions are able to promote such a 'pockets of conservatism' social arrangement is too broad to be addressed in this essay. Still, this thesis makes clear which interventions may *not* have a desirable effect. More generally, it becomes clear that the findings of this thesis have important policy implications. Note that if the 'pockets of conservatism' model is indeed a near-optimal arrangement, some institutions that may be deemed pernicious might not be so, and they might actually promote the progress of science. For example, peer-review might be a good mechanism to promote individual conservatism, while being compatible with a wide dispersion of different subcommunities across the landscape. Ultimately, this reveals that the effectiveness of science policy is highly dependent on the accuracy of one's interpretation of the socio-epistemic dynamics in science.

5.3. Further reflections

The conclusions and discussion I have laid out here are, of course, strongly based on the model I have presented. As such, one might try and dispute at least some of my findings via a criticism of the model which I have set. Let me, then, address some potential worries regarding the model and the simulations run.

Firstly, I return to the findings regarding the control populations in experiments 1 and 2. Remember that what sets control agents apart from non-control agents is the fact that the former do not engage in social learning, as they do not consider whether the patches of the landscape have been previously visited, nor do they, consequently, consider the elevation of previously visited paths. I have noted that the control population outperforms non-control populations in most simulation scenarios in experiment 1. In this experiment, it is only among the two lowest cluster groups (1 and 3) that there are non-control populations which on average outperform the control population in the same cluster group. In experiment 2, the control population is also usually rather competent, often comparable to the best-performing non-control populations. These are surprising results, in that they suggest that social learning is not epistemically advantageous in most scenarios. Indeed, this result is similar to Pöyhönen's (2017) own findings that social learning is not beneficial in smooth epistemic landscapes such as the one I have used in this thesis.

Pöyhönen suggests that this is indeed the result of modelling the epistemic landscape as a smooth surface, since the pattern vanishes once one (arguably more plausibly)

considers a rugged landscape. This finding is relevant for explaining my own control results, and, as such, it may ease the potential worries regarding the relatively good performance of controls. Moreover, I would highlight the curious fact that, unlike Pöyhönen, I do find exceptions to this pattern in some simulation scenarios, namely those in which the community is (at least initially) highly concentrated in the landscape. With Pöyhönen's results in mind, then, these exceptions invite the conclusion that social learning is especially beneficial in strongly group-level conservative populations. This is, of course, under the assumption that modelling the landscape as a smooth surface affects the results concerning the benefits of social learning in the same direction across *all* scenarios. While Pöyhönen offers some support towards this assumption, more research should be done to grant it. The result is nevertheless an intriguing one, and further research would be needed in order to understand its roots and dynamics.

There might be further, more general worries regarding the application of ABMs to the topic of this thesis. I have motivated the use of ABMs in general and in the social epistemology of science more specifically. Science is a social enterprise which displays behavioural heterogeneity at the individual level. These are two main factors that render ABMs promising tools to study the epistemic dynamics involved in scientific research (see Payette 2012). Moreover, computer simulations are a sophisticated form of thought experiment, and in that sense they can play an important exploratory role in the study of a philosophical topic (Mayo-Wilson and Zollman 2021). But, of course, the model I have presented is highly idealised and one might therefore be suspicious of the conclusions that it supports. Indeed, recently there has been a thriving literature

on how ABMs of scientific communities can enable one to learn about the models' target systems (e.g. Frey and Šešelja 2018; Kuorikoski and Reijula 2019; Aydinonat et al. 2021; Kuhlmann 2021; Šešelja 2021)[33]. As Aydinonat et al. (2021) highlight, the highly idealised natured of such models raises two main concerns – firstly, whether their representational adequacy is empirically established (Martini and Pinto 2017; Reijula and Kuorikoski 2019) and, secondly, whether the findings promoted by the models are robust (Grimm and Berger 2016; Rosenstock et al. 2017; Frey and Šešelja 2018; Pinto and Pinto 2018; Frey and Šešelja 2020). The scope of this thesis does not allow me to empirically test the adequacy of my model. However, I will address some concerns regarding the second worry.

As with all models, some implementation decisions run the risk of being arbitrary. A major concern may be the way in which I specified the individual moving strategies. This particular aspect followed closely from Currie's semi-formal account of creativity and conservatism in science, and in chapter 3 I devoted some effort to motivating as well as refining the different research strategies in the model. Still, the implementation depended on some arbitrary choices (such as the fact that a creative jump is ten patches long), and these parameters should be wiggled to test the robustness of the results. While such a sensitivity analysis is welcome, I should highlight the measures I took to mitigate robustness-related worries.

[33] See also Frigg and Reiss (2009) for a critical assessment of the philosophical discussions regarding simulation-based models more generally.

Some features of the main model I used to run simulations were borrowed from previous projects employing ABMs to study the social epistemology of scientific creativity, such as the fact that scientific communities were made up of 400 agents, or the fact that in some experiments, the outcomes of interest were measured after 500 cycles of each simulation were run (e.g. Weisberg and Muldoon 2009; Thoma 2015). But one might still reasonably worry about the arbitrariness of these and other modelling decisions. However, these worries should be assuaged by a few factors. Firstly, the fact that I ran simulations measuring different outcomes in different scenarios served to test the robustness of the main results. Indeed, my main conclusions were those that relied on the most robust of the results across all simulations of all experiments. Moreover, while it is true that, for example, I relied on modelling a community of 400 scientists, the arbitrariness of this decision was mitigated by the fact that I considered different landscape sizes. This allowed to consider different relative sizes of a scientific community, holding all else equal. The use of time as an outcome variable in experiment 2 also helped to mitigate the arbitrariness of measuring outcomes after 500 cycles in the other experiments. Moreover, the use of different depletion scenarios in experiment 4 allowed me to test the robustness of the results to different conceptions of the extent to which neighbouring approaches are distinct from each other (coarse-grained vs fine-grained distinction).

Of course, it was not possible to test for the robustness of the results with respect to all modelling decisions, and in that sense my investigation welcomes further investigation.

Chapter 6. Conclusion

It would be difficult to argue that creativity does not contribute to the progress of science. But, to be clear, this is *not* what I have argued in this thesis. Instead, I have raised the point that creative scientific communities need not be populated by creative individuals. And, indeed, perhaps they should not, for the sake of the progress of science.

By combining some considerations about the institutions of science with a set of results based on my epistemic landscape agent-based model of science, I suggested that an optimal arrangement approximates one in which a scientific community explores a given topic of research by having subcommunities conservatively exploring sets of approaches which differ substantially across these subcommunities. This arrangement combines the benefits of conservatism with the benefits of creativity. The focus on significantly overhauling the incentives to which scientists are subject in order to promote more creativity is therefore misguided. Rather, the efforts should be geared towards promoting a wide enough array of subcommunities adopting sufficiently distinct sets of approaches to the ones that are already adopted. However, contrary to some suggestions, it is beneficial, not alarming, that the members of these subcommunities tend to behave conservatively.

I would like to end this thesis by alluding to an anecdotal counterpoint to the example I gave of Robert Goddard in the Introduction. It refers us to another scientist from the turn of the 20th century, physicist Max Planck. In a letter (cited in Hermann and Nash

2004) written by Planck to fellow physicist Robert Wood, in which the former accounts for his own discovery of the quantum hypothesis, Planck writes:

> "Briefly summarized, what I did can be described as simply an act of desperation. By nature I am peacefully inclined and reject all doubtful adventures. But by then I had been wrestling unsuccessfully for six years with the problem of the equilibrium between radiation and matter and I knew that the problem was of fundamental importance to physics (…). A theoretical interpretation therefore had to be found at any cost, no matter how high".

The discipline of physics, of course, was revolutionised by Planck's discovery. And yet, Planck, the *reluctant revolutionary* (Kragh 2000), matches rather closely, by his own account, the model of the conservative scientist, '[rejecting] all doubtful adventures', preferring the well explored theories and approaches, exploring the space of possible solutions coldly and slowly. While Goddard and the 'lone-wolf', creative type of scientist he represents are certainly often admirable and impressive (and indeed may play an important epistemic role when the wider scientific community is too conservative at the group-level), it seems that science can well thrive without them.

Bibliography

Alexander, J. M. (2013). Preferential attachment and the search for successful theories. *Philosophy of Science*, *80*(5), 769-782.

Alexander, J. M., Himmelreich, J., & Thompson, C. (2015). Epistemic landscapes, optimal search, and the division of cognitive labor. *Philosophy of Science*, *82*(3), 424-453.

Avin, S. (2019). Mavericks and lotteries. *Studies in History and Philosophy of Science Part A*, *76*, 13-23.

Avin, S. (2020). Centralized funding and epistemic exploration. *The British Journal for the Philosophy of Science*.

Axelrod, R. (1997) *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton Univ. Press, Princeton, NJ.

Aydinonat, N. E., Reijula, S., & Ylikoski, P. (2021). Argumentative landscapes: the function of models in social epistemology. *Synthese*, *199*(1), 369-395.

Bala, V., & Goyal, S. (1998). Learning from neighbors. *Review of Economic Studies*, *65*(3), 595–621.

Baumgaertner, B. (2018). Models of Opinion Dynamics and Mill-Style Arguments for Opinion Diversity. *Historical Social Research/Historische Sozialforschung*, *43*(1 (163), 210-233.

Bishop, M. A. (2005). The autonomy of social epistemology. *Episteme*, *2*(1), 65-78.

Boland, T., & Griffin, R. (Eds.). (2015). *The sociology of unemployment*. Manchester University Press.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, *99*(suppl 3), 7280-7287.

Bright, L. K. (2017). On fraud. *Philosophical Studies*, *174*(2), 291-310.

Bruner, J. P. (2013). Policing epistemic communities. *Episteme*, *10*(4), 403-416.

Bedessem, B. (2019). The division of cognitive labor: two missing dimensions of the debate. *European Journal for Philosophy of Science*, *9*(1), 1-16.

Bedessem, B. (2021). Two conceptions of the sources of conservatism in scientific research. *Synthese*, *198*(7), 6597-6614.

Borg, A., Frey, D., Šešelja, D., & Straßer, C. (2019). Theory-choice, transient diversity and the efficiency of scientific inquiry. *European Journal for Philosophy of Science*, *9*(2), 1-25.

Brezis, E. S. (2007). Focal randomisation: An optimal mechanism for the evaluation of R&D projects. *Science and Public Policy*, *34*(10), 691-698.

Clary, D. A. (2003). *Rocket man: Robert H. Goddard and the birth of the space age*. Hachette UK.

Currie, A. (2015). Marsupial lions and methodological omnivory: function, success and reconstruction in paleobiology. *Biology & Philosophy*, *30*(2), 187-209.

Currie, A. (2018). *Rock, bone, and ruin: An optimist's guide to the historical sciences*. MIT Press.

Currie, A. (2019a). Creativity, conservativeness & the social epistemology of science. *Studies in History and Philosophy of Science Part A*, *76*, 39-48.

Currie, A. (2019b). Existential risk, creativity & well-adapted science. *Studies in History and Philosophy of Science Part A*, *76*, 39-48.

Currie, A. & Avin, S. (2019). Method pluralism, method mismatch, & method bias. *Philosophers' Imprint, 19*(13), 1.

Gillies, D. (2014). Selecting applications for funding: why random choice is better than peer review. *RT. A Journal on research policy and evaluation*, *2*(1).

De Langhe, R. (2014a). To specialize or to innovate? An internalist account of pluralistic ignorance in economics. *Synthese*, *191*(11), 2499-2511.

De Langhe, R. (2014b). A unified model of the division of cognitive labor. *Philosophy of Science*, *81*(3), 444-459

De Langhe, R. (2018). An Agent-Based Model of Thomas Kuhn's The Structure of Scientific Revolutions. *Historical Social Research/Historische Sozialforschung*, *43*(1 (163), 28-47.

De Langhe, R., & Greiff, M. (2010). Standards and the distribution of cognitive labour: A model of the dynamics of scientific activity. *Logic Journal of the IGPL*, *18*(2), 278-293.

De Marchi, S., & Page, S. E. (2014). Agent-based models. *Annual Review of political science*, *17*, 1-20.

Devezer, B., Nardin, L. G., Baumgaertner, B., & Buzbas, E. O. (2019). Scientific discovery in a model-centric framework: Reproducibility, innovation, and epistemic diversity. *PloS one*, *14*(5), e0216125.

Dilaver, Ö., Calvert Jump, R., & Levine, P. (2018). Agent-Based Macroeconomics and Dynamic Stochastic General Equilibrium Models: Where Do we Go from Here?. *Journal of Economic Surveys*, *32*(4), 1134-1159.

Dollinger, S. J. (2007). Creativity and conservatism. *Personality and Individual Differences*, *43*(5), 1025-1035

Epstein, J. (2006). *Generative social science [electronic resource] : Studies in agent-based computational modeling* (Princeton studies in complexity). Princeton: Princeton University Press.

Fazelpour, S., & Steel, D. (2022). Diversity, Trust and Conformity: A Simulation Study. *Philosophy of Science*, 1-29

Frankenhuis, W. E., & Nettle, D. (2018). Open science is liberating and can foster creativity. *Perspectives on Psychological Science*, *13*(4), 439-447.

Frey, D., & Šešelja, D. (2018). What is the epistemic function of highly idealized agent-based models of scientific inquiry?. *Philosophy of the Social Sciences*, *48*(4), 407-433.

Frey, D., & Šešelja, D. (2020). Robustness and idealizations in agent-based models of scientific interaction. *The British Journal for the Philosophy of Science*.

Frigg, R., & Reiss, J. (2009). The philosophy of simulation: hot new issues or same old stew?. *Synthese*, *169*(3), 593-613.

Galison, P., & Hevly, B. (1992). *Big science: The growth of large-scale research*. Stanford, CA: Stanford University Press.

Garner, R. (2015). Goddard Space Flight Center. https://www.nasa.gov/goddard

Gilbert, N. (2007). *Agent-Based Models* (Quantitative Applications in the Social Sciences). Los Angeles: SAGE Publications.

Godfrey-Smith, P. (2008). Recurrent transient underdetermination and the glass half full. *Philosophical Studies*, *137*(1), 141-148.

Goldin-Meadow, S. (2016, September). Why preregistration makes me nervous. *Observer*. Retrieved from http://www.psychologicalscience.org/observer/why-preregistration-makes-me-nervous

Grim, P., Singer, D. J., Fisher, S., Bramson, A., Berger, W. J., Reade, C., ... & Sales, A. (2013). Scientific networks on data landscapes: Question difficulty, epistemic success, and convergence. *Episteme*, *10*(4), 441-464.

Grim, P., Singer, D. J., Bramson, A., Holman, B., McGeehan, S., & Berger, W. J. (2019). Diversity, ability, and expertise in epistemic communities. *Philosophy of Science*, *86*(1), 98-123.

Grimm, V., & Berger, U. (2016). Robustness analysis: Deconstructing computational models for ecological theory and applications. *Ecological modelling*, *326*, 162-167.

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

Haufe, C. (2013). Why do funding agencies favor hypothesis testing?. *Studies in History and Philosophy of Science Part A*, *44*(3), 363-374.

Heesen, R. (2017). Communism and the Incentive to Share in Science. *Philosophy of Science*, *84*(4), 698-716.

Heesen, R. (2018). Why the reward structure of science makes reproducibility problems inevitable. *The journal of philosophy*, *115*(12), 661-674.

Heesen, R. (2019). The credit incentive to be a maverick. *Studies in History and Philosophy of Science Part A*, *76*, 5-12.

Heesen, Remco. (2021). Cumulative Advantage and the Incentive to Commit Fraud in Science. *The British Journal for the Philosophy of Science,* The British journal for the philosophy of science, 2021-07.

Heesen, R., & Bright, L. K. (2020). Is peer review a good idea?. *The British Journal for the Philosophy of Science*.

Hermann, A., & Nash, C. W. (2004). *The genesis of quantum theory (1899-1913)*. Cambridge: Mit Press.

Hirsh, A. E. (2009). 'Guest column: A new kind of big science', *New York Times* (Opinionator: Exclusive Online Commentary from the Times), published Jan 13. Available from opinionator.blogs.nytimes.com/2009/01/13/guest-column-a-new-kind-of-big-science/. Accessed 20 May 2022.

Holman, B., & Bruner, J. (2017). Experimentation by industrial selection. *Philosophy of Science*, *84*(5), 1008-1019.

Holman, B., & Elliott, K. C. (2018). Teaching and learning guide for: The promise and perils of industry-funded research. *Philosophy Compass*, *13*(11), e12549.

Hull, D. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science* (Science and its conceptual foundations). Chicago; London: University of Chicago Press.

Kaaronen, R. O. (2018). Exploration and Exploitation in Scientific Inquiry: Towards a Society of Explorers. [Preprint].

Kant, J. D., Ballot, G., & Goudet, O. (2020). WorkSim: An Agent-Based Model of Labor Markets. *Journal of Artificial Societies and Social Simulation*, *23*(4).

Kitcher, P. (1990). The division of cognitive labor. *The journal of philosophy*, *87*(1), 5-22.

Kitcher, P. (1993). *The advancement of science : Science without legend, objectivity without illusions*. New York ; Oxford: Oxford University Press.

Klein, D., Marx, J., & Fischbach, K. (2018). Agent-based modeling in social science, history, and philosophy. An introduction. *Historical Social Research/Historische Sozialforschung*, *43*(1 (163), 7-27.

Kragh, H. (2000). Max Planck: the reluctant revolutionary. *Physics World*, *13*(12), 31.

Kuhlmann, M. (2021). On the Exploratory Function of Agent-Based Modeling. *Perspectives on Science*, *29*(4), 510-536.

Kuhn, T. (1962). *The structure of scientific revolutions* (International encyclopedia of unified science; v. 2, no. 2). Chicago; London: University of Chicago Press.

Kummerfeld, E., & Zollman, K. J. (2016). Conservatism and the Scientific State of Nature. *British Journal for the Philosophy of Science*, *67*(4).

Kuorikoski, J., & Reijula, S. (2019). Making it count. An inferentialist account of computer simulation. [Preprint]

Landa, D. (2015). Behavioral political economy, argumentation, and democratic theory. *The Good Society*, *24*(1), 86-97.

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative science quarterly*, *52*(4), 667-694.

Levy, N., & Alfano, M. (2020). Knowledge from vice: Deeply social epistemology. *Mind*, *129*(515), 887-915.

Longino, H. E. (2002). *The fate of knowledge*. Princeton, Oxford: Princeton University Press.

Loye, D. (2007). Telling the new story: Darwin, evolution, and creativity versus conformity in science. In *Everyday creativity and new views of human nature: Psychological, social, and spiritual perspectives.* (pp. 153-173). American Psychological Association.

Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation*, *21*(1), 48-60.
Manning, A. (2013). Monopsony in motion. In *Monopsony in Motion*. Princeton University Press.

Martini, C., & Pinto, M. F. (2017). Modeling the social organization of science. *European Journal for Philosophy of Science*, *7*(2), 221-238.

Mayo-Wilson, C. (2014). Reliability of testimonial norms in scientific communities. *Synthese*, *191*(1), 55-78.

Mayo-Wilson, C., & Zollman, K. J. (2021). The computational philosophy: simulation as a core philosophical method. *Synthese*, *199*(1), 3647-3673.

Mayo-Wilson, C., Zollman, K. J., & Danks, D. (2011). The independence thesis: When individual and social epistemology diverge. *Philosophy of Science*, *78*(4), 653-677.

Mayo-Wilson, C., Zollman, K., & Danks, D. (2013). Wisdom of crowds versus groupthink: learning in groups and in isolation. *International Journal of Game Theory*, *42*(3), 695-723.

Merton, R. K. (1957). Priorities in scientific discovery: a chapter in the sociology of science. *American sociological review*, *22*(6), 635-659

Mitić, S., & de Vries, S. (2012). Rapid mixing techniques for the study of enzyme catalysis. *Ultrafast tangential micro-mixers for the study of biochemical reactions on the microsecond time scale*, 19.

Muldoon, R. (2013). Diversity and the division of cognitive labor. *Philosophy Compass*, *8*(2), 117-125.

New York Times Editorial. (1920). A Strain on Credulity. *The New York Times* (13 January 1920). https://www.nytimes.com/1920/01/13/archives/topics-of-the-times.html

O'Connor, C. (2019). The natural selection of conservative science. *Studies in History and Philosophy of Science Part A*, *76*, 24-29.

O'Connor, C., & Bruner, J. (2019). Dynamics and diversity in epistemic communities. *Erkenntnis*, *84*(1), 101-119.

O'Connor, C., & Weatherall, J. O. (2018). Scientific polarization. *European Journal for Philosophy of Science*, *8*(3), 855-875.

Payette, N. (2012). Agent-based models of science. In *Models of science dynamics* (pp. 127-157). Springer, Berlin, Heidelberg.

Pinto, M. F., & Pinto, D. F. (2018). Epistemic landscapes reloaded: An examination of agent-based models in social epistemology. *Historical Social Research/Historische Sozialforschung*, *43*(1 (163), 48-71.

Polanyi, Michael. (2000). The Republic of Science: Its Political and Economic Theory. *Minerva (London), 38*(1), 1-21.

Popova, N., Moiseenko, Y., & Beavitt, T. (2017). Conformity in modern science: An engine of societal transformation?. *Changing Societies & Personalities. 2017. Vol. 1. Iss. 3*, 237-258.

Popper, K. (1963). *Conjectures and refutations : The growth of scientific knowledge* (Ebook central). Oxfordshire, England ; New York, New York.

Pöyhönen, S. (2017). Value of cognitive diversity in science. *Synthese*, *194*(11), 4519-4540.

Psillos, S. (1999). *Scientific realism : How science tracks truth* (Philosophical issues in science). London: Routledge.

Reijula, H. S., & Kuorikoski, J. E. (2019). Modeling epistemic communities. In *The Routledge Handbook of Social Epistemology* (pp. 240-249). Routledge.

Reijula, S., & Kuorikoski, J. (2021). The diversity-ability trade-off in scientific problem solving. *Philosophy of Science*, *88*(5), 894-905.

Romero, F. (2017). Novelty versus replicability: Virtues and vices in the reward system of science. *Philosophy of Science*, *84*(5), 1031-1043.

Romero, F. (2020). The Division of Replication Labor. *Philosophy of Science*, *87*(5), 1014-1025.

Rosenstock, S., Bruner, J., & O'Connor, C. (2017). In epistemic networks, is less really more?. *Philosophy of Science*, *84*(2), 234-252.

Roston, D., & Cui, Q. (2016). QM/MM analysis of transition states and transition state analogues in metalloenzymes. In *Methods in enzymology* (Vol. 577, pp. 213-250). Academic Press.

Santana, C. (2021). Let's not agree to disagree: the role of strategic disagreement in science. *Synthese*, *198*(25), 6159-6177.

Santana, C. (2022). Why citizen review might beat peer review at identifying pursuitworthy scientific research. *Studies in History and Philosophy of Science*, *92*, 20-26.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, *1*(2), 143-186.

Schneider, M. D. (2021). Creativity in the social epistemology of science. *Philosophy of Science*, *88*(5), 882-893.

Šešelja, D. (2021). Exploring Scientific Inquiry via Agent-Based Modelling. *Perspectives on Science*, *29*(4), 537-557.

Smaldino, P. E., & O'Connor, C. (2020). Interdisciplinarity can aid the spread of better methods between scientific communities. [Preprint]

Smith, T., Husbands, P., & O'Shea, M. (2002). Fitness landscapes and evolvability. *Evolutionary computation*, *10*(1), 1-34.

Stanford, P. K. (2001). Refusing the devil's bargain: What kind of underdetermination should we take seriously?. *Philosophy of Science*, *68*(S3), S1-S12.

Stanford, P. K. (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives*. New York: Oxford University Press.

Stanford, P. K. (2019). Unconceived alternatives and conservatism in science: the impact of professionalization, peer-review, and Big Science. *Synthese*, *196*(10), 3915-3932.

Strevens, M. (2003). The role of the priority rule in science. *The Journal of philosophy*, *100*(2), 55-79.

Strevens, M. (2006). The role of the Matthew effect in science. *Studies in History and Philosophy of Science Part A*, *37*(2), 159-170.

Strevens, M. (2013). Herding and the quest for credit. *Journal of Economic Methodology*, *20*(1), 19-34.

Thoma, J. (2015). The epistemic division of labor revisited. *Philosophy of Science*, *82*(3), 454-472.

Weatherall, J. O., & O'Connor, C. (2021). Conformity in scientific networks. *Synthese*, *198*(8), 7257-7278.

Weatherall, J. O., O'Connor, C., & Bruner, J. P. (2020). How to beat science and influence people: policymakers and propaganda in epistemic networks. *The British Journal for the Philosophy of Science*.

Weisberg, M. (2012). *Simulation and similarity: Using models to understand the world*. Oxford University Press.

Weisberg, M. (2013). Modeling herding behavior and its risks. *Journal of Economic Methodology*, *20*(1), 6-18.

Weisberg, M., & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of science*, *76*(2), 225-252.

Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. http://ccl.northwestern.edu/netlogo/

Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Mit Press.

Williamson, T. (2017). Model-Building in Philosophy. *Philosophy's future: The problem of philosophical Progress*, 159-171.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the sixth international congress on genetic*s (pp. 355–366).

Zamzami, N., & Schiffauerova, A. (2017). The impact of individual collaborative activities on knowledge creation and transmission. *Scientometrics*, *111*(3), 1385-1413.

Zollman, K. J. (2007). The communication structure of epistemic communities. *Philosophy of science*, *74*(5), 574-587.

Zollman, K. J. (2009). Optimal publishing strategies. *Episteme*, *6*(2), 185-199.

Zollman, K. J. (2010). The epistemic benefit of transient diversity. *Erkenntnis*, *72*(1), 17-35.

Zollman, K. J. (2013). Network epistemology: Communication in epistemic communities. *Philosophy Compass*, *8*(1), 15-27.

Zollman, K. J. (2018). The credit economy and the economic rationality of science. *The Journal of Philosophy*, *115*(1), 5-33.

Zollman, Kevin J.S. (2019) The Scientific Ponzi Scheme. [Preprint]