

INTERDISCIPLINARITY CAN AID THE SPREAD OF BETTER METHODS BETWEEN SCIENTIFIC COMMUNITIES

PAUL E. SMALDINO^{1,2,*} AND CAILIN O'CONNOR³

September 3, 2022

Abstract: Why do bad methods persist in some academic disciplines, even when they have been widely rejected in others? What factors allow good methodological advances to spread across disciplines? In this paper, we investigate some key features determining the success and failure of methodological spread between the sciences. We introduce a formal model that considers factors like methodological competence and reviewer bias towards one's own methods. We show how these self preferential biases can protect poor methodology within scientific communities, and lack of reviewer competence can contribute to failures to adopt better methods. We then use a second model to argue that input from outside disciplines can help break down barriers to methodological improvement. In doing so we illustrate an underappreciated benefit of interdisciplinarity.

Keywords: interdisciplinary; cultural evolution; bias; competence; agent-based model; meta-science

1. INTRODUCTION

Much has been written about the benefits of interdisciplinarity. Prior discussions have focused on two categories of benefit, which we might call “innovation” and “emergence.” Innovation occurs when methods and theories find applications in fields external to those in which they were developed, leading to new insights (Foster et al., 2015; Rzhetsky et al., 2015). Examples include the application of spin glass models from physics by neuroscientists to understand perceptual categorization (Hopfield, 1982) and the application of disease contagion models by social scientists to understand the diffusion of innovations (Bass, 1969; Centola, 2018).¹ Emergence goes beyond knowledge transfer to involve the creation of new amalgamate disciplines, often to great success. Examples include

¹UNIVERSITY OF CALIFORNIA, MERCED

²SANTA FE INSTITUTE

³UNIVERSITY OF CALIFORNIA, IRVINE

E-mail address: psmaldino@ucmerced.edu.

¹Important innovations can also involve the combination of two or more ideas from different disciplines (Uzzi et al., 2013).

the relatively recently formed fields of cognitive science (Goldstone, 2019), network science (Brandes et al., 2013), and cultural evolution (Mesoudi, 2017), while cybernetics (Wiener, 1948) provides an earlier example.²

In this paper, we describe an additional benefit of interdisciplinarity, usually omitted from the conversation, which we might call “methodological competition”. We propose that interdisciplinarity promotes the maintenance and spread of superior scientific methodologies. Superior methods as we consider them need not be innovative, and as such this discussion is orthogonal to the “innovation” benefit. Our main claim is that when inferior methods persist in individual disciplines, interdisciplinary contact can improve the chances that superior methods take over. Our argument is supported by a case study and by mathematical and agent-based modeling grounded in the framework of cultural evolution.

In many cases, scientific communities continue to use inferior methods even after they are abandoned in other disciplines. The replication crisis, for instance, has shed light on problematic practices in areas like social psychology and biomedical research (Open Science Collaboration, 2015; Ebersole et al., 2016). Even as some disciplines adopt reforms like open data, preregistration, and more rigorous statistical training, others remain resistant.³ On the face of it, this pattern seems surprising. We expect scientific communities to share the normative goal of seeking truth, and thus expect that methods promoting this goal should be widely adopted. What is going wrong?

Our first model considers groups of scientists who choose between methods that are more or less likely to yield epistemic successes—that is, to contribute new knowledge or understanding. The research produced by these scientists is reviewed by peers, and those methods that yield publications tend to be adopted by others in the field. When review tracks only epistemic success, there is no problem. Communities adopt superior methods.

We are interested in cases where reviewers are either incompetent to judge between methods, or else show biases for the (possibly inferior) methods that they themselves are already using. We find that these factors can lead to the stable persistence of poor methods. This result extends previous work by Akerlof and Michaillat (2018), who focus on how self preferential bias prevents the spread of superior theoretical paradigms.

Our second model shows how interdisciplinarity can break this pattern. In particular, we focus on two mechanisms by which this can happen: 1) credit assignment and 2) copying between disciplines. First, when reviewers from disciplines with different competences and biases are able to evaluate work within a problematic discipline, and assign credit based on their own standards, better methods can gain traction within the original

²Interdisciplinarity also involves costs. Individual researchers often find it effortful to stay abreast of research in multiple fields simultaneously (Leahy, 2018), and interdisciplinary research correspondingly tends to lead to fewer total publications (Leahy et al., 2017) and reduced success in securing funding (Bromham et al., 2016). Interdisciplinary research has been occasionally characterized as “high risk, high reward” (Uzzi et al., 2013; Foster et al., 2015; Leahy, 2018).

³There is occasionally disagreement regarding the benefit of such practices, such as with preregistration (Szollosi et al., 2019). Such disagreements serve to highlight variation in the norms and methods between related research communities.

discipline. Second, when individuals using poor methods occasionally copy the methods of successful peers in other fields, this increases the likelihood that better methods take over in their own community.

Our models are tailored to scientific communities, but the mechanisms at play may be generalized to other settings. For instance, it has been recognized that sub-optimal practices in a wide variety of organizations can get “locked in” by positive feedback effects (Arthur, 1989). When firms continue to use poor practices (or fail to adopt new innovations), when government bodies do not adopt ideal scientific standards, or when sports leagues do not try out best training regimes, contact with other, similar groups may help. Again, this can happen via outside credit giving, i.e., when feedback from outsiders allows innovative insiders to have more influence. It can also happen when practitioners copy the successful methods of another firm or team. In this way, the right levels of contact between various sorts of groups may improve the performance of all. To put it even more broadly, interconnectivity may improve the collective intelligence of groups by facilitating the adoption of superior problem-solving methods.

The remainder of the paper is organized as follows. In section 2 we present a case study which illustrates the importance of competence, self preferential bias, and interdisciplinarity in determining methodological practice in the sciences. In section 3 we give some relevant background for our model, and discuss previous research looking at the failure of good methods to spread in science. Section 4 outlines our first model, looking at the persistence of poor methods. We present an analysis of this model in section 5. We then move on to discuss interacting communities and interdisciplinarity in section 6. We conclude in section 7 by discussing implications for the optimal structure of scientific communities, and outlining, in more detail, the roles our models can play in reasoning about the real world.

2. MBI AND SPORTS SCIENCE

In this section we discuss a case study from the field of sports science.⁴ Two factors we explore in our models—competence and self preferential bias—seem to have played important roles in this case. And furthermore, interdisciplinary feedback now seems to be playing a key role in disciplinary reform. In the conclusion we will briefly discuss a few additional examples.

Foam rolling involves lying down, face up, with a soft polymer cylinder under one’s back and rolling over it with the aim of relaxing muscles. MacDonald et al. (2014) purport to show that foam rolling reduces muscle soreness, improves range of motion, and even improves performance in activities like vertical jump height when used after exercise. Since publication in 2014, their paper has been cited hundreds of times. But their claims rely on a statistical method called magnitude-based inference (MBI), which statisticians have widely criticized as misleading and incoherent (Sainani, 2018). Over the years, hundreds of published papers in the discipline of sports science have used MBI, and the demonstrated unsoundness of the method throws many of these results

⁴This case study draws heavily on the work of the science journalist Christie Aschwanden (Aschwanden and Nguyen, 2018; Aschwanden, 2018, 2019).

into doubt (Lohse et al., 2020). A re-examination of the data from the foam rolling study using more rigorous statistical testing casts doubts on many of their claims (Lohse et al., 2020).

Magnitude-based inference was introduced by Batterham and Hopkins (2006). Briefly, it compares the risk that an intervention causes harm with the chance that it benefits an athlete.⁵ Rather than providing equations or making documented code available that detailed the workings of MBI, the authors developed and distributed Microsoft Excel spreadsheets that allow readers to implement their method without understanding it.

Since its introduction, statisticians have been raising red flags about MBI and calling for statistical reform in sports science (Barker and Schofield, 2008; Welsh and Knight, 2015; Sainani, 2018; Lohse et al., 2020; Sainani et al., 2020). Biostatistician Andrew Vickers has described MBI as, “a math trick that bears no relation to the real world” (Aschwanden and Nguyen, 2018). Recently there has been movement to suppress the method in sports science, including a ban by the well-respected journal *Medicine and Science in Sports and Exercise* (MSSE). But despite this at the time of writing numerous articles continue to be published using variations of MBI.

Several factors contributed to the uptake and persistence of MBI. First, its originators, who are prominent members of the sports science community, actively promoted the method through a website and at sports science conferences.⁶ Second, the ease of use of the spreadsheets seems to have been attractive to users. Third, many sports scientists (like many other scientists) are relatively unfamiliar with the mathematical details of their statistical methods, and thus unable to adjudicate the quality of such methods themselves (Vigotsky et al., 2020). And last, the method has a high false positive rate. This means that researchers in sports science, who are often looking for small effects in small samples, were able to publish results using MBI that would otherwise have been statistically insignificant.⁷

There are a few things to highlight about this case. First, for the duration of its use statisticians have universally labeled MBI as a misleading methodology. Nevertheless, reviewers *within* sports science, in judging the research of their peers, have approved its use. As mentioned, this seems to be at least partly due of a lack statistical competence by these reviewers. It also seems to be due to the strength of self preferential biases. Once the use of MBI became widespread, researchers who employed it themselves were perfectly willing to accept other papers using it for publication. In editorial on MBI biostatistician Doug Everett writes, “I almost get the sense that this is a cult. The method has a loyal following in the sports and exercise science community, but that’s the only place that’s adopted it” (Aschwanden and Nguyen, 2018).

A second thing to highlight is the role of statisticians in the slow move towards rejecting MBI. This includes their role as reviewers in the field. For example, when the

⁵A detailed description of how MBI works is provided by Sainani (2018).

⁶As of the time of writing, Hopkins continues to actively promote the use of MBI, though Batterham seems to have distanced himself from it.

⁷For work arguing that pressure to publish can lead to the widespread adoption of poor statistical methods in this way, see Smaldino and McElreath (2016).

originators of MBI tried to publish a defense of the method in the journal *MSSE*, reviewers in statistics rejected the paper.⁸ In addition, some sports scientists have worked directly with statisticians to promote better methods and call for reform (Sainani et al., 2020). These forms of contact with statisticians seem to have been crucial in the current turn against the method.

3. BIAS, COMPETENCE, AND INTERDISCIPLINARY CONTACT

In the next section we use models to systematically investigate some of the key features at work in the case of MBI and others like it, and test their relevance to the persistence of poor methodology. Before doing so, let us first address these key features—self preferential bias, competence, and interdisciplinary contact—in greater detail. We will follow this with a discussion of previous, relevant modeling literature.

3.1. Bias and Competence. One focus of our investigation here is researcher *bias* towards one's own methods. There are a cluster of causes for this sort of bias. Many communities have norms for researchers to use the standard methods of the field. And journals often act as gatekeepers standardizing a field's practices. Novel methods must often overcome conservative biases for existing methods, a point noted by Kuhn (1962) and others. Recently, concerns have been raised about the adverse effects of conservative norms in rejecting both novel methodologies and novel questions in the social sciences (Stanford, 2019; Akerlof, 2020; Barrett, 2020a). In addition, a wide body of empirical work has found that scientists show preferences for work like their own during peer review.⁹

Despite the influence of self preferential biases, we expect new methodologies to spread in a scientific community if researchers discern a clear advantage to adopting them. For this reason, *competence* to assess methodological quality is critical. By competence, we mean the ability of researchers within a research tradition to assess the relative quality of methodologies in producing epistemic value. Several studies show low inter-rater reliability among peer reviewers (Cole et al., 1981; Cicchetti, 1991; Mutz et al., 2012; Nicolai et al., 2015) suggesting that competence may sometimes be highly variable. While competence will not be the same for all researchers within a discipline, shared norms and training practices may create systematic differences in competence across

⁸Hopkins and Batterham then submitted to the journal *Sports Medicine*, where reviewers from sports science, who Hopkins describes as having been “groomed” by him, accepted the paper (Aschwanden and Nguyen, 2018; Hopkins and Batterham, 2016).

⁹Mahoney (1977) asked reviewers to rate manuscripts with identical procedures, but different conclusions, and found significant bias towards findings supporting the reviewer's own perspectives. This, of course, shows a preference for a finding, not a method, but still illuminates the sort of self preferential bias we model here. Travis and Collins (1991), in an observational study of reviewers of grant applications, found evidence of bias towards one's own “cognitive community”—those sharing the same interests and assumptions. Likewise Lamont et al. (2009), drawing on interviews with panelists from interdisciplinary funding boards, describes preferences towards the practices of panelists' home disciplines. In a study of over 600,000 publications, Wang et al. (2017) considered papers with unusual combinations of citations, indicating novelty. Although they found that the most highly cited papers were often novel, novel papers were also systematically published in lower-impact journals, indicating a likely bias against new methods.

fields, as in the case of MBI. Exacerbating this effect, in some fields the rewards for adopting new methodologies may be especially hard to evaluate compared to others, particularly when effect sizes are small and data is hard to gather.

We have just discussed self preferential bias and competence to assess methods. Before continuing, it is worth stepping back to clarify how we are using the term “method” here. There are a number of related things that the models we develop can apply to. Paradigmatic applications include cases like the MBI one, where scientists use discrete, easily-labeled practices, which they learn from other scientists. Things like statistical practices (MBI vs standard frequentist methods), experimental paradigms (like induced valuation in economics or looking times in developmental psychology), quality control methods (preregistration or open data), mathematical frameworks (like using Markov chain models to capture genetic drift), etc. fall within these clear cases.

There are other things the models can apply to, but where the mapping between models and reality will be less precise. These might include guiding theories or paradigms like natural selection or relativity theory. These shape practice and spread between scientists, but are embedded in many other theoretical commitments and aspects of practice. In addition, as noted, we might apply the models to practices in non-scientific communities, such as the use of checklists by airlines to prevent fatal accidents, or interval training by cross country coaches. Readers should ideally keep the sorts of cases discussed in the previous paragraph in mind when assessing our models, though we leave open the possibility of wider applications.

3.2. Interdisciplinarity. Many scientific fields are linked via shared interests, forming a loose network of communities (Vilhena et al., 2014). Such links provide a social structure for fruitful interdisciplinary contact. Interdisciplinarity allows different fields or sub-communities to retain their cultural identity and norms, while also receiving and sharing knowledge, ideas, and methods.

There are different ways in which interdisciplinary contact occurs. Journals receiving work that cannot be reviewed within their field may seek input from outsiders. Alternatively, a researcher may submit work for evaluation in adjacent fields. An academic might attend a conference in another field and learn about new methods or theories. Two researchers might collaborate across disciplines and, in doing so, train each other in their different methods and assumptions. A research trained in one discipline might even change fields, bringing methods and ideas from her original discipline to her new one.

There are contextual factors that determine when and whether researchers will benefit from such contact. Some fields have norms discouraging interdisciplinary publication, and in some fields outside publications or collaborations will not support tenure or promotion decisions.¹⁰ Some disciplines have practices that cannot be appropriately critiqued from without, such as when they use specialized, expensive technology that requires extensive training. In still other cases, the costs of training and technology will

¹⁰For example, consider the supremacy of the “top five” journals in economics (Heckman and Moktan, 2020).

mean that members of one discipline are unable to learn good methods from another, even if doing so might be epistemically beneficial.

In other cases, though, researchers can be successful by being relatively interdisciplinary—by publishing in the journals of adjacent fields, receiving grants centered in adjacent disciplines, collaborating with outsiders, or learning methods from adjacent fields. As we will argue, in these cases it may be possible for better methods to gain traction even in communities marked by strong self preferential bias or low competence. In this way contact between disciplines, or other sorts of groups, can potentially improve the performance of all. In the MBI case, for example, interdisciplinary contact of these sorts seems to be helping to eliminate poor statistical practice.

3.3. Previous Work. In developing our models, we (fittingly) draw on work from several disciplines. We follow previous authors in assuming that scientists are part of a “credit economy”. That is, scientists strive for publications—and the citations, talk invitations, job offers, grant money etc. that ensue—in the same way that normal people strive for wealth or happiness.

Many credit economy models focus on scientists as rational credit seekers (see, for example, Kitcher (1990); Bright (2017)). Our model, in contrast, falls in line with previous work on “selection” models, where scientists are part of a population where certain behaviors are selected by dint of their success. In particular, as will become clear, we assume that methods which generate credit tend to spread. This could be because prominent role models tend to be imitated, as with MBI in sports science. It could be due to conscious choices to copy methods that generate credit. Or it could stem from differential success of students whose advisors use credit-producing methods. Previous models have shown that these sorts of selection processes can help explain failures of methodology and discovery in science (Smaldino and McElreath, 2016; Smaldino et al., 2019; Holman and Bruner, 2017; O’Connor, 2019; Stewart and Plotkin, 2021; Tiokhin et al., 2021).

Such models are related to biological models looking at the selection of beneficial behavioral traits. In particular, the multi-group model we present in section 6 reflects a cultural evolutionary model developed by Boyd and Richerson (2002). They consider a population where different cultural groups adopt cultural variants via imitation. They assume, as we do, that imitation tracks payoff success, and show how beneficial variants can spread between sub-groups.¹¹ We, likewise, are interested in cases where beneficial epistemic practices can spread to other sub-fields. Our models, however, are specifically designed to represent details of scientific communities. In addition, when it comes to the spread of methods, we focus both on imitation between groups and the possibility of interdisciplinary credit allocation.

Akerlof and Michaillat (2018) model the persistence of “false paradigms.” That is, they investigate why a scientific community might continue to adhere to a set of sub-optimal guiding theories. They consider the possibility that scientists are biased towards their own paradigms in tenuring younger faculty. As they show, such a bias can stabilize

¹¹Boyd and Richerson are interested in the spread of strategically relevant variants, like coordination strategies, not payoff-neutral variants, like enjoying Nicholas Cage films.

poor paradigms, especially when tenure judgments are less sensitive to the quality of the work performed. While we make some different modeling assumptions, our findings concerning self preferential bias are very similar to theirs, adding robustness to this idea. Even with different modeling choices, self preferential bias can stabilize poor practice, particularly in the face of incompetence. Our second model further departs from Akerlof and Michaillat (2018) by investigating the role of outside influences in methodological change. They propose that paradigm change is largely contingent on improvements in the competence of community members. This may be the right focus for paradigms that are strictly discipline specific. As we show, however, methods which are used across disciplines can spread through contact between communities.

4. THE ONE-COMMUNITY MODEL

We consider a community where each scientist employs one of two characteristic methods: an “adequate” method, (A), or a “superior” method, (S). We assume that the quality of methodology is relevant to the epistemic products of these scientists. The expected epistemic value produced by a researcher using the adequate method is a baseline of $x_A = 1$, while a researcher using the superior method produces an expected epistemic value of $x_S = 1 + \delta$. A higher value of δ corresponds to a greater distinction between the quality of the methods. In the sports science case, this difference would track the gap between MBI, which produces low quality results that are poor guides to action in the world, and more standard statistical methods. As another example, we might compare randomized control trials to trials without a control. The former method is better at eliminating spurious correlational findings, and thus is better at adding to existing bodies of knowledge.

The first question we ask is: under what conditions do these methods yield more or less credit for their users? On the assumption that high-credit methods tend to be imitated, answering this question will allow us to predict whether communities will adopt one method or the other.

Scientists employ their methods to generate results, and are then assigned credit by reviewers in their field. Reviewers here could represent those who literally review work for publication. But reviewers could also represent other credit assignment mechanisms in science, including, for instance, those who read papers and cite them, or invite the author to apply for a job, or to speak at a conference. Reviewers are characterized by two field-specific properties: competence and self preferential bias. Competence, $\omega \in [0, 1]$, is the ability to discern the relative value of another scientist’s methods. At the extremes, when $\omega = 1$, reviewers are perfectly accurate in their assessment, and when $\omega = 0$ reviewers cannot distinguish the value of a method from the average method used in their field. The competence-based credit rating assigned to scientist i is given by:

$$(1) \quad K_i = \omega x_i + (1 - \omega)\bar{x},$$

where x_i is the quality of the method used by the focal individual and \bar{x} is the mean epistemic quality of the methods used in the population. If we let p be the proportion

of the population using superior methods, this mean methodological quality is given by

$$(2) \quad \bar{x} = (1 - p)(1) + p(1 + \delta) = 1 + p\delta.$$

Notice that when competence is very low (i.e., $\omega \approx 0$), reviewers rate all papers as having the same quality. In computational versions of the model, detailed below, we can make the model more realistic by adding an error term where reviewer judgments are noisy. In such cases, different papers will receive different quality ratings, but incompetent reviewers will draw all these ratings from the same distribution.¹²

Self preferential bias, α , is the extent to which reviewers prefer research that uses similar methods to their own. When $\alpha = 1$, reviewers assign credit entirely based on similarity to their own methods, whereas when $\alpha = 0$, they view methodology solely in terms of its (perceived) objective merit.

The total credit given to research produced by scientist i using method x_i by reviewer j is thus:

$$(3) \quad C_i = (1 - \alpha)K_i + \alpha B_j(x_i),$$

where K_i , again, is the contribution based on reviewer competence, and $B_j(x_i)$ is the contribution based on the bias of reviewer j towards their own method when assessing method i . This term will be 1 if scientist i uses the same method as the reviewer j , and 0 otherwise.

To review, when α —the weight determining bias versus competence in judgments—is very low and ω —the actual competence of reviewers—is very high, scientists receive a payoff commensurate with the quality of their methods. When ω decreases, reviewers are not biased to a particular method, but are unable to distinguish the underlying quality of the two methods. When α is higher, self preferential bias plays a role in review. The credit a researcher can expect to receive becomes dependent on the current distribution of methods in the field, such that more prominently used methods tend to get higher payoffs because more reviewers are familiar with them and prefer them.

The model we have presented to this point is a simple and highly idealized one. The goal is to investigate how just a few key factors in a community can impact the adoption of methods. In addition, this simplicity will allow us to naturally extend the model to the case of interdisciplinarity.

5. PAYOFFS AND INVASION IN A SINGLE COMMUNITY

With our model defined by the equations above, we can calculate the expected credit that will be assigned to scientists using either of the two methods. Again, let p be the proportion of scientists using the superior method. The expected credit contribution from reviewer bias, B_i , is $1 - p$ for an adequate scientist and p for a superior scientist. A scientist using method A should thus expect to receive credit of

$$(4) \quad C_A = (1 - \alpha) [\omega + (1 - \omega)(1 + p\delta)] + \alpha(1 - p),$$

¹²However, we find that as long as this noise is symmetrical, the qualitative nature of our results is unchanged when noise is added.

while a scientist using method S should expect to receive credit of

$$(5) \quad C_B = (1 - \alpha)[\omega(1 + \delta) + (1 - \omega)(1 + p\delta)] + \alpha p.$$

If we assume cultural evolutionary dynamics such that scientists who receive more credit will be better positioned to transmit their methods (c.f. McElreath and Boyd, 2007), then method S will increase in frequency whenever $C_S > C_A$. This will occur whenever

$$(6) \quad \delta > \frac{\alpha(1 - 2p)}{(1 - \alpha)\omega}.$$

In other words, method S spreads whenever the epistemic advantage to adopting the superior method (δ) is large enough to overcome limitations from the competence and bias of reviewers. Greater competence (ω) lowers this threshold. Greater bias (α) increases it.

The distribution of methods in the community (p) also matters. If a substantial number of individuals have already adopted method S, bias can work in favor of the superior method. In particular, there is a threshold frequency of users of method S which, if reached for whatever reason, will allow the superior method to take over. We can compute this threshold by solving for the minimum proportion of scientists using method S necessary for it to increase in frequency, p^* . This is the value of p for which $C_S > C_A$:

$$(7) \quad p^* = \frac{1}{2} - \frac{(1 - \alpha)\omega\delta}{2\alpha}.$$

If decisions are made entirely based on bias ($\alpha = 1$), then whichever method is more common will spread ($p^* = \frac{1}{2}$). As bias goes to zero, the better method is increasingly guaranteed to spread at lower threshold frequencies. For intermediate values of bias, increased competence can move the critical threshold lower, so that a better method held by the minority can still spread even in the presence of bias. Similar logic holds for when competence increases from zero. Figure 1 shows threshold values of p^* under several parameter values, along with the results of agent-based simulations of the model (described below). Note the close match between analytic and agent-based results.

Let us focus on the case in which method A is firmly entrenched in a scientific community, so that almost everyone is using method A. This case is perhaps the most interesting because it allows us to ask: under what conditions can an objectively better method increase in frequency when it is rare (or “invade” in the language of evolutionary game theory)? We can calculate the criteria for invasion by setting $p \approx 0$. The following inequality shows the minimum epistemic value advantage for method S to invade:

$$(8) \quad \delta > \frac{\alpha}{(1 - \alpha)\omega}$$

Figure 2 illustrates how competence and bias both contribute to the spread of better methods into a community that is currently adopting poor ones. This figure shows the minimum competence, ω , needed for method S to spread when rare as a function of

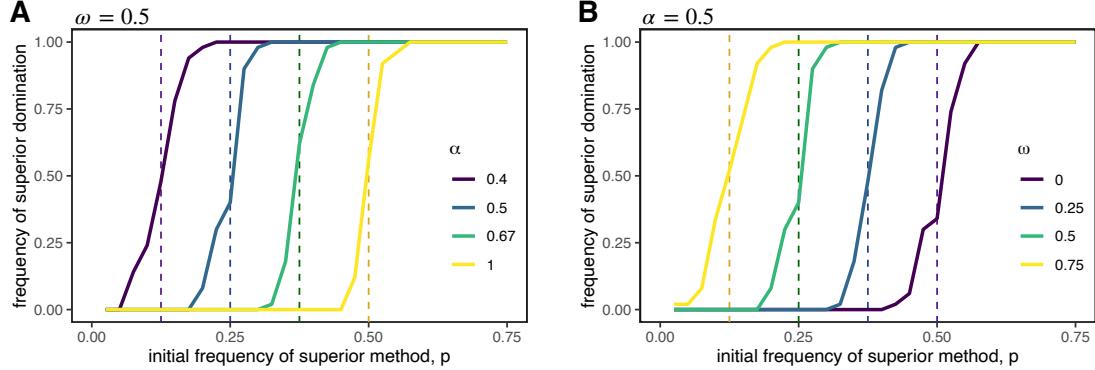


FIGURE 1. The evolution of superior methods as a function of their initial frequency, p . The left graph (A) varies self preferential bias, α , while holding competence fixed at $\omega = 0.5$. The right graph (B) varies competence, ω , while holding bias constant at $\alpha = 0.5$. Dashed lines are analytical calculations from Equation 7. Solid lines are the proportion of 100 runs of an agent-based model in which superior methods spread to fixation.

bias, for several values of the method's epistemic advantage¹³, δ . The graph could also be interpreted as the maximum bias allowable for the invasion of superior methods for a given level of competence. Either way, the greater the extent to which reviewers are biased towards their own methods, the more competence they must possess to accurately identify superior methods. The smaller the improvement of the method, the less likely superior methods are to spread when reviewers are even somewhat biased and imperfectly competent. This indicates that there may be a wide range of conditions under which novel superior methods fail to spread in a community of scientists.

5.1. Agent-based simulations. Agent-based simulations allow us to add noise and error to the model, to explore finite populations, and to add additional community structure. We find that they support the findings of the analytic model just presented. The simulation model described here will be extended in the subsequent section to allow us to examine multiple interacting communities.

In these simulations, we generate a population of $N = 100$ agents. We assume that self preferential bias and competence are homogeneous across the group. In each round each agent produces some epistemic product using their method. We calculate the credit an agent expects to receive for this work given the make-up of their community, just as in

¹³Note that we consider a wide range of values for δ ranging from 0.2 to 25. It is not out of the question that a new method may be an order of magnitude or more better than a previous method. Consider, for example, how innovations like PCR or fMRI revolutionized the biosciences.

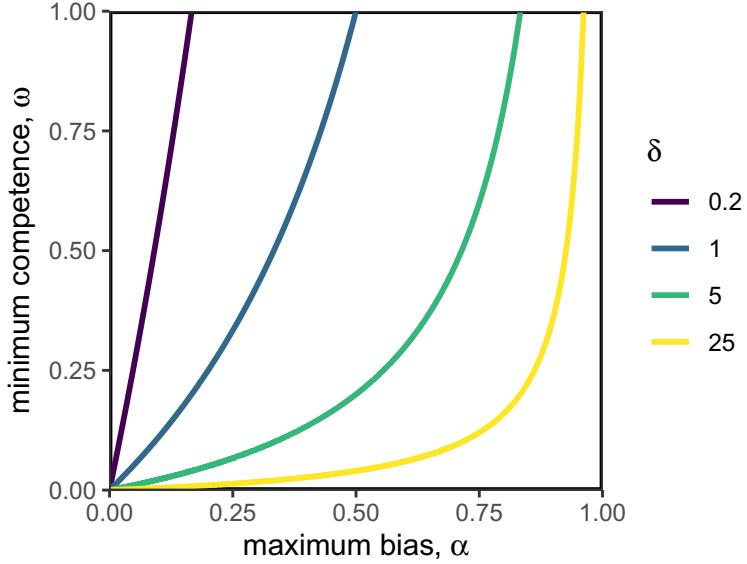


FIGURE 2. Minimum competence, ω , necessary for superior methods to invade a population of adequate methods as a function of bias, α . The shape of the curve depends on the epistemic advantage of better methods, δ .

the analytical model.¹⁴ Credit assignment reflects self preferential bias and competence, but might also involve the introduction of noise, which we will describe below.

We then use a process similar to the Moran process from evolutionary biology to determine how methods change over time (Moran, 1958). An individual i is randomly chosen to adopt a new method. They do so by randomly selecting another researcher j as a target for possible imitation. Agent i copies the method of agent j with a probability given by the following sigmoid function:

$$(9) \quad \Pr(i \text{ copies } j) = \frac{1}{1 + e^{-\gamma(C_j - C_i)}}$$

where γ is a constant that reflects the strength of selection. For simplicity, we used $\gamma = 10$ in all our simulations. This means that agents are more likely to copy targets with greater credit payoffs than themselves (if $C_j > C_i$) and rarely copy those with less credit, but on occasion they explore and make the non-credit-maximizing choice. As noted, this tendency to copy could represent several possibilities. Scientist who receive more credit tend to be more prominent—their papers are placed in better venues, they hold higher prestige jobs, they give more talks, etc. Their methods might be copied

¹⁴Note that this assumes no stochasticity based on random reviewer assignments. Of course in reality papers in a discipline are assigned to just a few reviewers, which means the same paper could receive different assessments. Since we include noise elsewhere in the credit-assignment process, we do not add extra noise in this way.

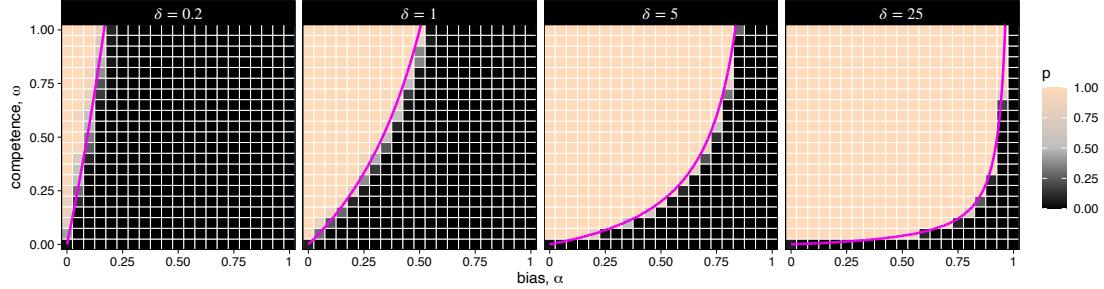


FIGURE 3. Proportion of simulation runs in which superior methods successfully invaded a population of adequate methods, as a function of competence, ω , and self preferential bias, α , for several values of epistemic difference, δ . Colored cells indicate the proportion of runs in which methods S dominated after 10^5 time steps, averaged across 100 runs for each combination of parameter values. Magenta curves are predictions from the analytical model seen in Figure 2.

simply because they are more visible to peers. In addition, scientists might be sensitive to which peers are receiving more credit, and actively replicate their methods. And last, prominent scientists might train and place more graduate students, so that “copying” represents the disproportionate replacement of retiring academics.

We ran simulations long enough that they reached stable states (where one of the two methods reached fixation) to see whether the population evolved to use the superior method over the course of simulation. A full description of the agent-based model, including all parameters used and a link to the source code, is provided in the Appendix.

Figure 3 shows the proportion of simulation runs in which method S successfully invaded and went to fixation. For these results we initialize the population with 95% of agents using method A, and with 5% using method S. As is evident, the analytic predictions (pictured as magenta lines) correspond very well to the outcomes of these simulations.¹⁵ Bias and competence trade-off in determining whether the better method can invade. And the epistemic difference between methods, δ , shifts the degree of competence necessary to offset bias.

As any academic knows, the peer review process is not always a consistent one. For this reason we also ran simulations in which error was introduced into credit assignment. Under this assumption, the credit assigned to scientist i was:

$$(10) \quad C_i = (1 - \alpha)K_i + \alpha B_i + \epsilon,$$

where ϵ is a random draw from a normal distribution with a mean of 0 and a standard deviation of σ . This error term captures two sources of noise in the evaluation process. First, the same method may produce research of higher or lower epistemic value based on the details of the research question and noise inherent in any complex process. Second,

¹⁵The match to analytic results indicates that our 5% frequency served as a good approximation of invasion in a finite population.

Model Parameters		
Parameter	Description	Values Tested
N	size of each community	100
δ	advantage of superior method	.02, 1, 2, 5, 25
γ	strength of selection	10
σ	noise in credit assignment	0, 0.1, 0.2
$\alpha_{1,2}$	bias in communities	[0, 1]
$\omega_{1,2}$	competence in communities	[0, 1]
c	out-group credit assignment	[0, 0.4]
m	out-group copying	[0, 0.4]

TABLE 1. Parameters used in the multi-community model

reviewers may be influenced by stochastic factors in addition to general competence and bias, ranging from hunger to prejudice. We found that the model was largely unresponsive to this sort of symmetric noise. Noise only created some increased stochasticity for parameters near the critical threshold. Because the plots of runs with noise up to at least $\sigma = 0.2$ are barely distinguishable from those with $\sigma = 0$, we have not included them here. All remaining results will use $\sigma = 0$.

6. INTERACTIONS BETWEEN COMMUNITIES

We have seen that if bias for existing methods is strong or if competence to assess the quality of new methods is low, inferior methods can persist and superior methods cannot spread. However, this conclusion so far holds only for an isolated community, in which community members always evaluate one another and in which copying always occurs within the community. In this section, we consider what happens when *two* scientific communities interact. This model is similar to the one-community version, with two exceptions (see Appendix for more details). First, when a scientist is evaluated for credit, the reviewer is chosen from the other community with probability c , which we call *out-group credit assignment*. Second, when a target is chosen for the potential transmission of methods, the target is chosen from the other community with probability m , which we call *out-group copying*. These represent two types of interdisciplinarity within a research area. The full two-community model is illustrated in Figure 4, and Table 1 summarizes parameter values we consider.

In particular, we are interested in cases where one community has adopted the superior method, and another the worse one. We consider two communities of size $N = 100$. We initialize community 1 as before, with 95% of agents using method A and only 5% using the superior method. We choose parameters such that in the absence of inter-group interaction ($c = m = 0$), the superior method will not spread in community 1. And we initialize community 2 with 95% of agents using method S, and 5% A, so that the superior method S is stable in the absence of interdisciplinarity. Our analyses show that a moderate amount of interdisciplinary contact can lead to the spread of the superior

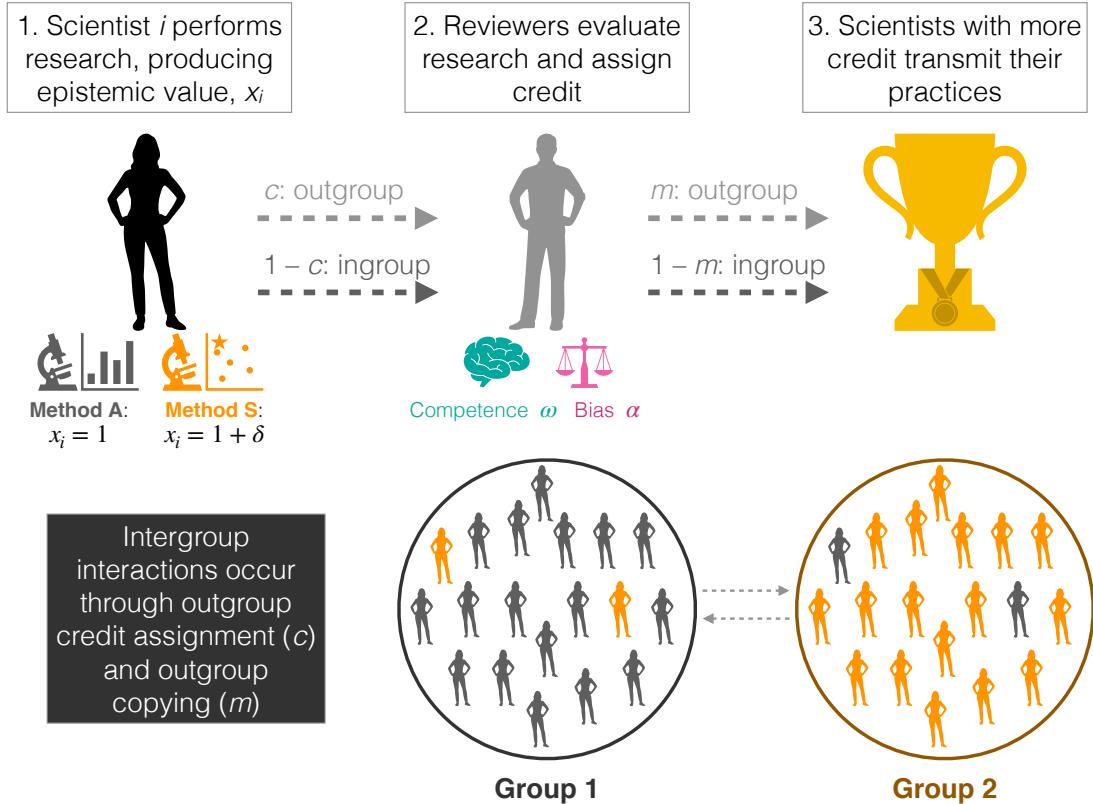


FIGURE 4. Model diagram.

method in community 1. The rest of this section will elaborate this finding, and point to a few cases where it does not always hold.

We begin by focusing on cases where method A is common in a community with either (1) low competence or (2) high self preferential bias, such that method S fails to spread when rare. The parameters of bias and competence used here are illustrated in figure 5. The pink horizontal axis represents the case in which superior methods fail to spread because of high bias. The two communities have the same competence, $\omega = .5$, but differ in bias with $\alpha_1 = \{.8\}$ and $\alpha_2 = \{.2\}$. The turquoise vertical axis represents the case where one community has relatively low competence. Our two communities have the same bias, $\alpha = .5$, but differ in competence with $\omega_1 = \{.2\}$ and $\omega_2 = \{.8\}$.

If a small but sufficient number of credit assignments and/or copying observations are made between the communities, the superior method can spread within both the high-bias and the low-competence communities. The greater the level of out-group copying and/or out-group credit assignment, the greater the likelihood that the superior

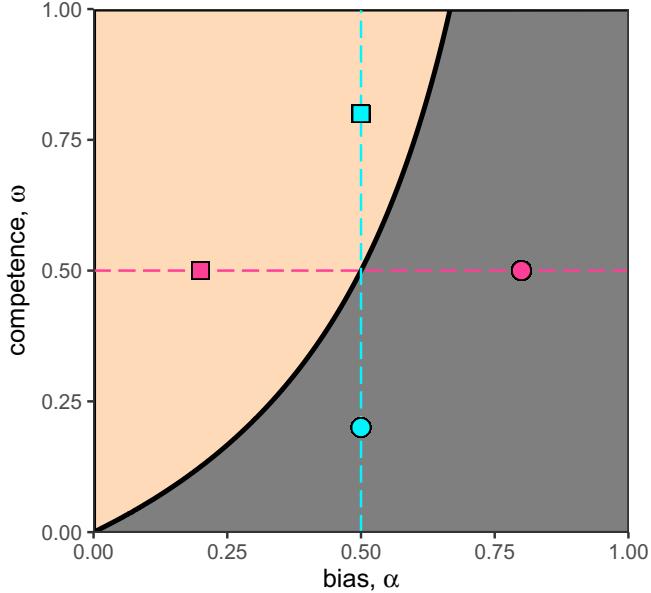


FIGURE 5. The relation between competence and bias for the invasion of better methods within a community, for $\delta = 2$. The gray area denotes a region where superior methods cannot invade, the peach-colored area denotes a region where they can. The points show the scenarios comparing adequate (circles) and superior (squares) communities, with the turquoise points on the vertical line differing on competence ($\omega = \{0.2, 0.8\}$) and the pink points on the horizontal line differing on bias ($\alpha = \{0.2, 0.8\}$).

methods ends up spreading through community 1. In addition, these two sources of interdisciplinarity interact additively. Figure 6 and figure 7 A,B display these results.¹⁶

Out-group credit assignment from community 2 means that those in community 1 using the superior method receive more credit than their colleagues using the adequate method, even when the latter method is more common in their community. This leads to greater prominence for these individuals, and more copying of their methods. The effect of out-group copying is even stronger. Credit assignment is accumulated over multiple interactions, only some of which involve the out-group. Out-group copying, on the other hand, has the potential to introduce superior methods into a community in one fell swoop. Notice that these effects interact with the epistemic distinction between

¹⁶Note that unlike the one-population model, we did not run these simulations until they reached stable outcomes. This is because with stochastic out-group credit assignment and copying, rare methods could persist at low levels for an arbitrarily long time, meaning that full stability was not always reached. Instead, we display results after $t = 10^5$ time steps, which was sufficient to approximate the long-term behavior of the model under various parameter values.

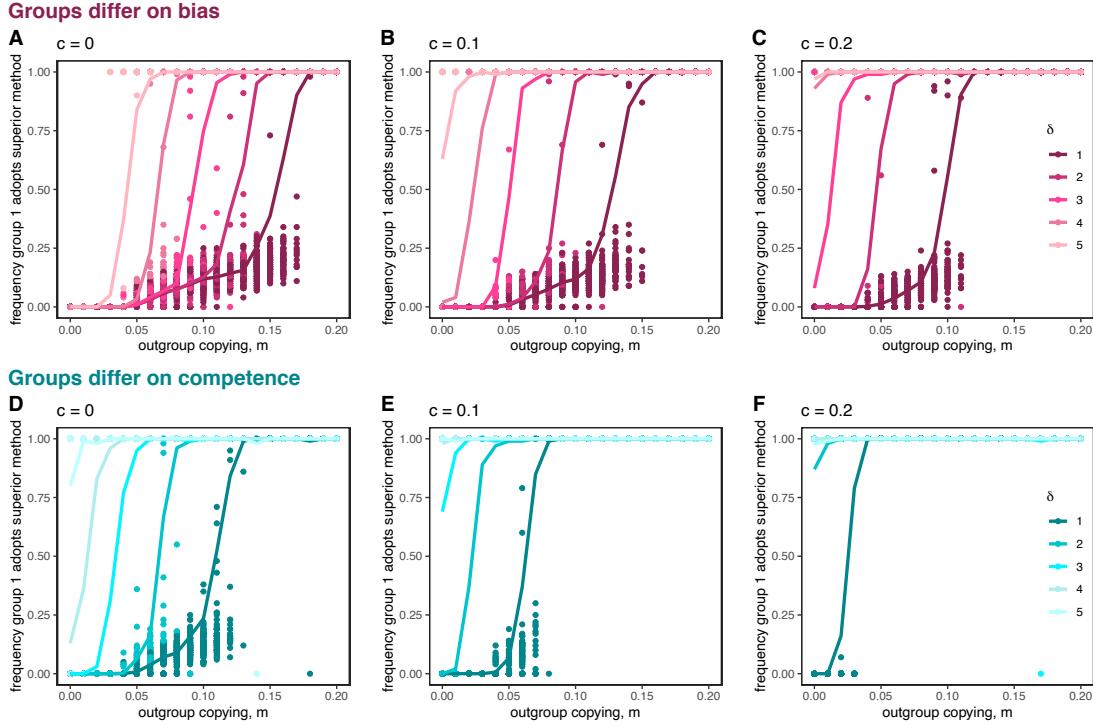


FIGURE 6. Interdisciplinarity allows the spread of better methods into a group that would otherwise not adopt them. Circles represent the proportion of agents in community 1 (the “adequate” community) who have adopted Method S at $t = 10^5$ time steps on individual runs of the model, for several values of δ . Lines are the averages across 100 runs for each set of parameter values. Top row (A–C): Superior methods spread to a highly biased group when there is sufficiently high out-group copying, m , and/or out-group credit assignment, c . Here $\alpha_1 = 0.8$, $\alpha_2 = 0.2$, $\omega_1 = \omega_2 = 0.5$. Bottom row (D–F): Superior methods spread to a low-competence group when there is sufficiently high out-group copying, m , and/or out-group credit assignment, c . Here $\omega_1 = 0.2$, $\omega_2 = 0.8$, $\alpha_1 = \alpha_2 = 0.5$.

the methods, δ . The greater the epistemic benefit of the superior method, the less interdisciplinarity is needed for its spread.

Strikingly, the spread of better methods by interdisciplinarity does not actually require that community 2 be particularly competent or unbiased. Instead, if a community adopts superior methods by *any* means, perhaps through some accident of history or through influence from another discipline, they can still pass on these beneficial practices. Figure 7 shows heat maps indicating the spread of superior methods both for the conditions in which communities differ, and for the conditions in which both communities are

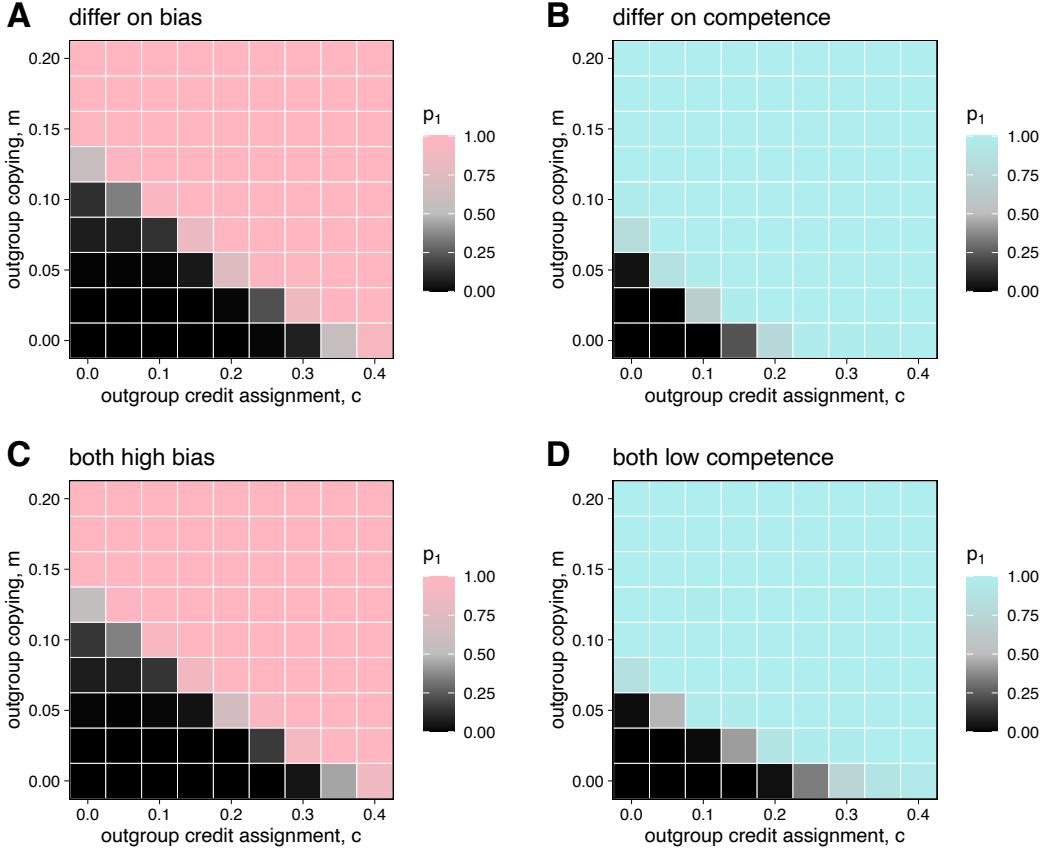


FIGURE 7. The positive effects of out-group credit assignment and out-group copying are additive. Heat maps display the mean frequency of agents using of method S in group 1, p_1 , after 10^5 time steps, averaged over 100 runs for each combination of parameter values. This effect holds when groups differ on bias (A: $\alpha_1 = 0.8$, $\alpha_2 = 0.2$, $\omega_1 = \omega_2 = 0.5$) or competence (B: $\omega_1 = 0.2$, $\omega_2 = 0.8$, $\alpha_1 = \alpha_2 = 0.5$). However, interdisciplinarity can also aid the spread of superior methods even when both communities are plagued by high bias (C: $\alpha_1 = \alpha_2 = 0.8$, $\omega_1 = \omega_2 = 0.5$) or low competence (D: $\omega_1 = \omega_2 = 0.2$, $\alpha_1 = \alpha_2 = 0.5$). For all cases, $\delta = 2$.

low in competence or high in self preferential bias. As we see in Figure 7C and D, even if both communities are incompetent or biased, superior methods still spread through interdisciplinary contact. This indicates that superior methods could potentially spread throughout *many* communities via a cascade of interdisciplinary contact.

To summarize, interdisciplinary contact, in the form of credit giving between disciplines and out-group copying, can unseat poor methods and replace them with better ones. This finding holds under a wide range of conditions. In addition to the parameters

featured in our figures, we ran additional simulations with varying levels of competence, bias, epistemic advantage, and population size, and found that our results are robust across a very wide range of parameter values. In particular, the precise levels of self preferential bias and competence seem to matter little as long as method S is stable in group 2. However, while superior methods are usually stable once common, there remain some rare cases in which a backslide is possible. We explore these in the next section.

6.1. Can worse methods ever replace superior methods? We have shown how superior methods can spread from a community in which they are common to a community in which they are rare. But can poor methods also spread due to interdisciplinary contact? Our analyses suggest this outcome is unlikely but not impossible. Consider a scenario in which there is widespread lack of competence to compare two different methodologies. That is, scientists in both communities cannot tell if one method is better than another. Instead they rely entirely on self preferential bias in making credit assignments. If both communities are strongly biased toward their existing methods, then nothing should change. But what if bias is strongest in the community that has adopted worse methods, and weaker in community that has adopted superior methods?

We analyze this “worst case” scenario by again considering two communities in which the superior method is initially rare in community 1 and common in community 2. Both communities lack any competence to compare methods ($\omega_1 = \omega_2 = 0$), and members of community 1 are strongly biased while members of community 2 are completely open minded ($\alpha_1 = 1, \alpha_2 = 0$). We find that under these circumstances, the worse method can spread to group 2, and that out-group credit assignment is actually the primary driver of this backslide (Figure 8, top left square). This is because reviewers in group 2 rate all scientific products equally, while reviewers in group 1 give strong ratings only to research using method A.

Further analysis, however, shows that increasing competence and self preferential bias in group 2 prevent this backsliding (Figure 8). In other words, worse methods are unlikely to spread between communities as long as those communities have at least some competence or self preferential bias. This points to a surprising possibility for why such a bias might sometimes be beneficial. In addition, some out-group copying also helps prevent backsliding to worse methods, as it helps both groups to maintain a population of individuals using the superior norm.

6.1.1. Productivity Advantages for Worse Methods. Our models suppose that good methods are no more difficult to employ than bad ones. But in many cases, rigorous methods are costly and time-consuming, and part of the draw of non-rigorous ones is that scientists using them can publish more quickly and easily (Smaldino and McElreath, 2016; Heesen, 2018). This was almost certainly part of the draw in the case of MBI in sports science. We might ask: in cases where poorer methods are highly productive, do they instead spread via interdisciplinary contact? The answer is, in certain cases, yes.

In our model, the parameter δ determines the degree to which the superior method is, in fact, superior. Suppose, however, that the adequate method increases productivity relative to the superior method. We can imagine adding a productivity boost, ρ , to the output for this method. In this regime scientists employing the worst method tend

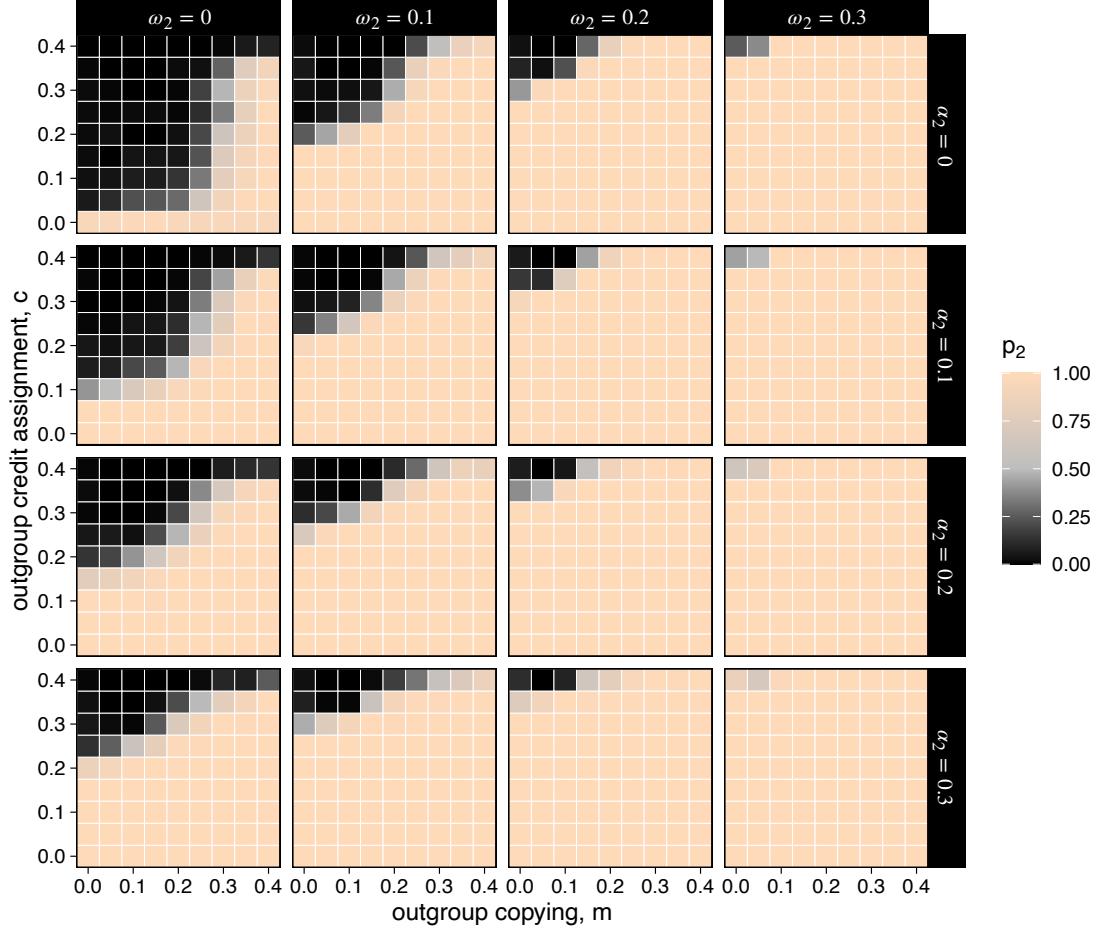


FIGURE 8. Frequency of the superior method in group 2 after 10^5 time steps under the “worst case” scenario. In all cases, $\omega_1 = 0$, $\alpha_1 = 1$, and $\delta = 2$. Each cell represents the average over 100 simulation runs for that combination of parameter values, and each box shows the long-term frequency of method S as a function of out-group credit assignment, c , and out-group copying, m . Each column represents a different value of ω_2 and each row represents a different value of α_2 .

to publish more papers per unit time as captured by ρ . And so although each paper might not be as epistemically important, researchers could still potentially gain a credit advantage by publishing more. The addition of ρ will tend to close the epistemic gap between the methods. This will, in turn, shift the threshold levels of bias and competence for which the better method will take over within one community, as well as the threshold level of community contact for the better method to spread to a new discipline. If ρ is high enough, the worse method will actually yield *more* epistemic output than the

better one. In such a regime, our results indicate that the worse method will instead be the one that spreads when communities come in contact.

Notice that for this to happen in our model, it must be the case that the worse method is actually producing greater net epistemic value on average, even if each individual paper is of worse quality. In model-world, the spread of the worse method in this case actually increases the output of the discipline. Thus, there is some question about whether the spread of the “worse method” counts as an epistemically poor outcome here. Presumably this regime does not track a case like MBI where, despite its high productivity, the method is so poor that it produces results that are highly questionable. Indeed, it may often be the case that output produced using low-quality methods is worse than producing nothing at all. Imagine a case, though, where a scientist develops off-the-shelf statistical packages that are easily employed and fairly dependable. These packages might replace a practice by medical researchers of hiring trained statisticians to analyze their experimental data. Clearly the trained statisticians will produce more dependable, accurate analyses and thus more trustworthy and useful epistemic outputs. But using these packages might greatly improve speed of production, and allow researchers to direct grant money to new experiments rather than data analysis. For this reason, the use of these easy packages might replace a “superior” practice, but simultaneously increase output in the discipline in such a way that is not clearly epistemically worse. In this case, the tradeoffs between short- and long-term costs and benefits to science must be considered in ways that are beyond the scope of the present analysis.

7. DISCUSSION

Strong bias for current methods or low competence to assess new methods can impede scientific progress. The case of MBI in sports science fits well with the analysis provided here, and there are many other cases that do as well. For instance, the discipline of evolutionary psychology has been widely criticized for sometimes publishing findings that are unsupported and even irresponsible. A central part of these criticisms involves the development of ultimately speculative narratives about the evolution of the human mind. As critics argue, these narratives are too unconstrained, and thus are unlikely to accurately capture facts about human evolution (Gould, 1991; Lloyd, 1999; Lloyd and Feldman, 2002; Gannon, 2002). When such work is published in journals edited and assessed by evolutionary psychologists, both self preferential bias and lack of competence can contribute to the maintenance of these uncritical applications of evolutionary theory. Competing traditions of more rigorous work in the evolutionary human sciences, including in human behavioral ecology, gene-culture co-evolution theory, and also work by more rigorous evolutionary psychologists (c.f. Gurven, 2018; Amir and McAuliffe, 2020; Barrett, 2020b), point to promising avenues for future reform. Other candidate examples include the past use of factor analysis in marketing research (Stewart, 1981) and applications of unit root econometrics in macroeconomics (Campbell and Perron, 1991).

A final example worth highlighting is the field of social psychology, which has recently undergone serious methodological changes. Retrospective studies found that, particularly pre-2011, the field was rife with poorly designed and improperly analyzed data. Many attempts to replicate high profile findings in the field failed (Open Science Collaboration, 2015; Ebersole et al., 2016). Some problematic practices like small sample sizes, lack of credible priors, and use of invalid measurement instruments may have persisted for a long time in this discipline because of the types of factors we identify here.

This field, though, also illustrates that not all problematic methods are well-captured by the models we develop. Consider HARKing (hypothesizing after results are known) and p-hacking (selectively reporting or manipulating data to produce desirable inferential statistics). Both of these are what we might call *implicit methodologies*. They are not typically reported in the details of published work. Reviewers thus do not typically know whether manuscripts they review involve p-hacking or HARKing, and so their self preferential biases and competence to assess these methods does not bear on their judgments. Instead, the issue stemmed from the competence of the researchers themselves, who likely did not have the statistical and/or philosophical training to understand the problems with these practices.¹⁷ Subsequent improvements in social psychology have benefited from interdisciplinary contact and feedback, but appear to have been driven largely by critiques from within the field. This is to say that the mechanisms we focus on throughout this paper are important, but not obligatory, for explaining the persistence, and eventual improvement, of poor methodology.

Our analysis suggests that interdisciplinarity should be promoted—especially contact related to methodological practice—as such contact may facilitate the spread of better methods to new disciplines. We have already talked about some keys mechanisms for this contact—the use of competent reviewers from neighboring disciplines; academics receiving grants from, or publishing in, other disciplines; academics attending conferences outside their discipline; and cross-disciplinary collaboration. In addition, the advent of social media platforms has led to increased contact between those in different disciplines, and especially to opportunities for feedback between these groups. This may help to lessen the silo-ing of academic fields. Furthermore, online platforms may improve opportunities for interdisciplinary credit giving when academics share or laud cross-disciplinary work. Likewise explicitly interdisciplinary conferences, publishing outlets, and institutes may play an important role in the spread of good methods. Such institutions bring academics from different disciplines into contact, increasing chances of credit-giving through citations and invitations, and of methodological copying.

Other communities outside academic science, such as governmental bodies and firms, may benefit from similar sorts of contact. Arthur (1989, 1990) describes how accidents of history can lead to lock-in effects in industry where once technologies or practices become common, they are not easily replaced by superior options. Fostering a diversity of practice in different communities, and then bringing these communities into contact, may help. This could involve regular conferences where workers discuss and share their

¹⁷Indeed, incentives to publish positive and exciting results may have *selected* for misunderstandings of methodological details (Smaldino and McElreath, 2016).

methodologies, presentations about the internal workings of these organizations to other groups, or collaborative efforts between organizations. These sorts of events may improve function at a wider level.

We might also ask: what stands in the way of sufficient interdisciplinarity in science? There are norms against interdisciplinarity in some fields. For instance, some fields consider publication in top insider journals a requirement for promotion. This limits the possibility of interdisciplinary credit-giving. In some cases these norms might arise from in-group favoritism and biases against the out-group. In other cases, they might arise out of a desire to preserve the special status of a discipline.¹⁸ In such disciplines, moves to change these norms may have positive epistemic impacts. Other fields may be silo-ed as a result of an inability to understand or engage with outside disciplines. In these cases, improved training may help, such as required graduate courses in methods from nearby fields.

One might conclude from our discussion that the best structure for scientific communities is a flat one, without disciplinary boundaries. But our results do not actually support this conclusion. A single unified scientific community would suffer from the problem indicated by our baseline model: the risk that new and improved methods fail to spread. Many have pointed to the benefits of certain types of diversity in academia. Longino (1990) in particular has advocated for critique across diverse views as important for rooting out poor assumptions and practices in science. Feminist philosophers of science like Longino (1990) and Okruhlik (1994) have lauded personal identity as an important source of beneficial cognitive diversity. Another such source stems from a diversity of educational regimes. However, several forces act to decrease diversity of practice within close-knit communities like academic disciplines, including human tendencies towards conformity, norm following, and practices of indoctrination. Some disciplinary structure may be important in preserving diversity of methods and assumptions. The aim, though, is to have enough contact between disciplines so that this diversity can prove beneficial to science as a whole.¹⁹

Before concluding, we want to say a little more about the usefulness of our models in informing real world systems. These models are obviously highly simplified, and, as such, may fail to incorporate relevant features of actual scientific communities. This means that in real cases interdisciplinary contact may not work just as our results suggest. The models, however, can play several epistemic roles in reasoning about interdisciplinary contact. First, they play a “how-possibly” role. At minimum we see that it is in principle possible for interdisciplinary contact to play a direct causal role in improving poor scientific methodology. More importantly our models are useful in directing our

¹⁸Maintaining such a special status might even be tied to explicit financial incentives, as with economics vis-à-vis the other social sciences.

¹⁹See Grossmann (2021) for similar arguments about the social sciences specifically. There is also a connection here to research indicating an intermediate amount of contact between groups, rather than full connectivity, is optimal for solving some types of complex problems (Lazer and Friedman, 2007; Derex and Boyd, 2016), and to work suggesting that an intermediate amount of communication is optimal for scientific theory change because it ensures transient diversity of beliefs in science (Zollman, 2010; Wu and O'Connor, 2022).

attention to real world interventions that have real promise vis-à-vis methodological improvement. Our models give good reasons for researchers to pay further attention to, and to further investigate, how and whether interdisciplinary contact can reform areas of the sciences that have fallen behind methodologically. Of course, a number of other researchers have already argued for various benefits to interdisciplinarity. But the models here outline specific causal pathways involving interdisciplinary credit-giving and copying that lead to the spread of superior methods. The viability of such pathways should be a topic for further research.

ACKNOWLEDGMENTS

Many thanks to Christie Aschwanden, Kristin Sainani, and Aaron Caldwell for communications about methods in sports science. Thanks to Andrew Vigotsky for sharing unpublished work on sports science. For comments on previous drafts of this manuscript, we thank Liam K. Bright, Remco Heesen, Colin Holbrook, Aaron Lukaszewski, Leo Tiokhin, and Pete Richerson. Thanks to Scott Page and several anonymous referees for feedback. And finally, thanks to the organizers of the Metascience 2019 conference, which provided the interdisciplinary setting at which this paper was first conceived.

DECLARATION OF CONFLICTING INTEREST

The Authors declare that there is no conflict of interest.

REFERENCES

Akerlof, G. A. (2020). Sins of omission and the practice of economics. *Journal of Economic Literature*, 58(2):405–18.

Akerlof, G. A. and Michaillat, P. (2018). Persistence of false paradigms in low-power sciences. *Proceedings of the National Academy of Sciences*, 115(52):13228–13233.

Amir, D. and McAuliffe, K. (2020). Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behavior*.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394):116–131.

Arthur, W. B. (1990). Positive feedbacks in the economy. *Scientific American*, 262(2):92–99.

Aschwanden, C. (2018). A flawed statistical method was just banned from a major sports science journal.

Aschwanden, C. (2019). Sports science is finally talking about its methodology problems.

Aschwanden, C. and Nguyen, M. (2018). How shoddy statistics found a home in sports research.

Barker, R. J. and Schofield, M. R. (2008). Inference about magnitudes of effects. *International journal of sports physiology and performance*, 3(4):547–557.

Barrett, H. C. (2020a). Deciding what to observe: Thoughts for a post-WEIRD generation. *Evolution and Human Behavior*.

Barrett, H. C. (2020b). Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends in Cognitive Sciences*.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*, 15(5):215–227.

Batterham, A. M. and Hopkins, W. G. (2006). Making meaningful inferences about magnitudes. *International journal of sports physiology and performance*, 1(1):50–57.

Boyd, R. and Richerson, P. J. (2002). Group beneficial norms can spread rapidly in a structured population. *Journal of theoretical biology*, 215(3):287–296.

Brandes, U., Robins, G., McCranie, A., and Wasserman, S. (2013). What is network science? *Network Science*, 1(1):1–15.

Bright, L. K. (2017). Decision theoretic model of the productivity gap. *Erkenntnis*, 82(2):421–442.

Bromham, L., Dinnage, R., and Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, 534(7609):684–687.

Campbell, J. Y. and Perron, P. (1991). Pitfalls and opportunities: What macroeconomists should know about unit roots. *NBER Macroeconomics Annual*, 6:141–201.

Centola, D. (2018). *How behavior spreads*. Princeton University Press.

Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, 14(1):119–135.

Cole, S., Cole, J. R., and Simon, G. A. (1981). Chance and consensus in peer review. *Science*, 214(4523):881–886.

Derex, M. and Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, 113(11):2982–2987.

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B., Boucher, L., et al. (2016). Many labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, 67:68–82.

Foster, J. G., Rzhetsky, A., and Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *American Sociological Review*, 80(5):875–908.

Gannon, L. (2002). A critique of evolutionary psychology. *Psychology, Evolution & Gender*, 4(2):173–218.

Goldstone, R. L. (2019). Becoming cognitive science. *Topics in Cognitive Science*, 11(4):902–913.

Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. *Journal of Social Issues*, 47(3):43–65.

Grossmann, M. (2021). *How Social Science Got Better: Overcoming Bias with More Evidence, Diversity, and Self-reflection*. Oxford University Press.

Gurven, M. D. (2018). Broadening horizons: Sample diversity and socioecological theory are essential to the future of psychological science. *Proceedings of the National Academy of Sciences*, 115(45):11420–11427.

Heckman, J. J. and Moktan, S. (2020). Publishing and promotion in economics: The tyranny of the top five. *Journal of Economic Literature*, 58(2):419–70.

Heesen, R. (2018). Why the reward structure of science makes reproducibility problems inevitable. *The Journal of Philosophy*, 115(12):661–674.

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

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558.

Hopkins, W. G. and Batterham, A. M. (2016). Error rates, decisive outcomes and publication bias with several inferential methods. *Sports Medicine*, 46(10):1563–1573.

Kitcher, P. (1990). The division of cognitive labor. *The Journal of Philosophy*, 87(1):5–22.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Lamont, M. et al. (2009). *How professors think*. Harvard University Press.

Lazer, D. and Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, 52(4):667–694.

Leahy, E. (2018). The perks and perils of interdisciplinary research. *European Review*, 26(S2):S55–S67.

Leahy, E., Beckman, C. M., and Stanko, T. L. (2017). Prominent but less productive: The impact of interdisciplinarity on scientists' research. *Administrative Science Quarterly*, 62(1):105–139.

Lloyd, E. A. (1999). Evolutionary psychology: The burdens of proof. *Biology and Philosophy*, 14(2):211–233.

Lloyd, E. A. and Feldman, M. W. (2002). Evolutionary psychology: A view from evolutionary biology. *Psychological Inquiry*, 13(2):150–156.

Lohse, K. R., Sainani, K. L., Taylor, J. A., Butson, M. L., Knight, E. J., and Vickers, A. J. (2020). Systematic review of the use of “magnitude-based inference” in sports science and medicine. *PLOS ONE*, 15(6):e0235318.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

MacDonald, G. Z., Button, D. C., Drinkwater, E. J., and Behm, D. G. (2014). Foam rolling as a recovery tool after an intense bout of physical activity. *Medicine and Science in Sports and Exercise*, 46(1):131–142.

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2):161–175.

McElreath, R. and Boyd, R. (2007). *Mathematical models of social evolution: A guide for the perplexed*. University of Chicago Press.

Mesoudi, A. (2017). Pursuing darwin's curious parallel: Prospects for a science of cultural evolution. *Proceedings of the National Academy of Sciences*, 114(30):7853–7860.

Moran, P. A. P. (1958). Random processes in genetics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 54, pages 60–71. Cambridge University Press.

Mutz, R., Bornmann, L., and Daniel, H.-D. (2012). Heterogeneity of inter-rater reliabilities of grant peer reviews and its determinants: A general estimating equations approach. *PLOS ONE*, 7(10):e48509.

Nicolai, A. T., Schmal, S., and Schuster, C. L. (2015). Interrater reliability of the peer review process in management journals. In *Incentives and Performance*, pages 107–119. Springer.

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

Okruhlik, K. (1994). Gender and the biological sciences. *Canadian Journal of Philosophy*, 24(sup1):21–42.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716.

Rzhetsky, A., Foster, J. G., Foster, I. T., and Evans, J. A. (2015). Choosing experiments to accelerate collective discovery. *Proceedings of the National Academy of Sciences*, 112(47):14569–14574.

Sainani, K. L. (2018). The problem with “magnitude-based inference”. *Medicine and Science in Sports and Exercise*, 50(10):2166–2176.

Sainani, K. L., Borg, D. N., Caldwell, A. R., Butson, M. L., Tenan, M. S., Vickers, A. J., Vigotsky, A. D., Warmenhoven, J., Nguyen, R., Lohse, K. R., et al. (2020). Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. *British Journal of Sports Medicine*.

Smaldino, P. E. and McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9):160384.

Smaldino, P. E., Turner, M. A., and Contreras Kallens, P. A. (2019). Open science and modified funding lotteries can impede the natural selection of bad science. *Royal Society Open Science*, 6(7):190194.

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

Stewart, A. J. and Plotkin, J. B. (2021). The natural selection of good science. *Nature Human Behaviour*.

Stewart, D. W. (1981). The application and misapplication of factor analysis in marketing research. *Journal of marketing research*, 18(1):51–62.

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T., and Donkin, C. (2019). Is preregistration worthwhile? *Trends in Cognitive Sciences*, 24(2):94–95.

Tiokhin, L., Yan, M., and Morgan, T. (2021). Competition for priority harms the reliability of science, but reforms can help. *Nature Human Behaviour*.

Travis, G. D. L. and Collins, H. M. (1991). New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology, & Human Values*, 16(3):322–341.

Uzzi, B., Mukherjee, S., Stringer, M., and Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157):468–472.

Vigotsky, A. D., Halperin, I., Gal, D., and McShan, B. B. (2020). Decreasing dichotomization for the wrong reasons: Interpretation of evidence in sports science and medicine researchers. unpublished MS.

Vilhena, D. A., Foster, J. G., Rosvall, M., West, J. D., Evans, J., and Bergstrom, C. T. (2014). Finding cultural holes: How structure and culture diverge in networks of

scholarly communication. *Sociological Science*, 1:221.

Wang, J., Veugelers, R., and Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8):1416–1436.

Welsh, A. H. and Knight, E. J. (2015). “Magnitude-based inference”: A statistical review. *Medicine and Science in Sports and Exercise*, 47(4):874.

Wiener, N. (1948). *Cybernetics: or Control and Communication in the Animal and the Machine*. MIT press.

Wu, J. and O'Connor, C. (2022). How should we promote transient diversity in science?

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

APPENDIX A. AGENT-BASED MODEL DESCRIPTION

Consider a population consisting of communities of scientists. We considered cases in which there was always either one or two communities. Each community is made up of $N = 100$ scientists, each of whom keeps track of their group identity and is characterized by a method $x_i \in \{A, S\}$. The true epistemic value of method A is set at 1 without loss of generalization; the true epistemic value of method S is set at $1 + \delta$, so that δ represents the epistemic advantage of method S. Each community is initially characterized by a dominant method, which is used by 95% of its population. The remaining 5% of scientists use the non-dominant method. Each community k is further defined by levels of bias, α_k and competence, ω_k .

The dynamics of the model proceed in discrete time steps, each of which consists of two stages: *Science* and *Evolution*.

A.1. Science. In the Science stage, each scientist in group i performs research using their characteristic method, producing an expected value. Each scientist is then assigned an expected value of credit, on the assumption that their research is assessed by a large number of reviewers. A proportion $1 - c$ of the reviewers are drawn from the agent's own group, i , and the remaining proportion c of the reviewers are drawn from the other group, j (the community to which scientist in group i does *not* belong). When $c = 1$, each community is completely evaluated by the other community, an unrealistic scenario that nevertheless creates conditions for conformity. When c is small but nonzero, we have conditions for interdisciplinarity, in which scientists are occasionally judged by the standards of other communities. If there is only one community, $c = 0$ and all reviewers are naturally drawn from the agent's own community.

To calculate an agent's expected credit payoff, we must know the epistemic advantage of the superior method, the bias, competence, and frequency of the superior method for each group, and the overall level of out-group credit assignment in the research community as a whole. The credit assigned to an agent i using the adequate method is therefore:

$$(11) \quad C_{Ai} = (1 - c) [(1 - \alpha_i)(\omega_i + (1 - \omega_i)(1 + p_i\delta)) + \alpha_i(1 - p_i)] \\ + c [(1 - \alpha_j)(\omega_j + (1 - \omega_j)(1 + p_j\delta)) + \alpha_j(1 - p_j)],$$

while the credit assigned to an agent i using the superior method is:

$$(12) \quad C_{Si} = (1 - c) [(1 - \alpha_i)(\omega_i(1 + \delta) + (1 - \omega_i)(1 + p_i\delta)) + \alpha_i p_i] \\ + c [(1 - \alpha_j)(\omega_j(1 + \delta) + (1 - \omega_j)(1 + p_j\delta)) + \alpha_j p_j]$$

Under conditions where noise is added to agents' credit score, each agent in either group i receives an adjusted credit score:

$$(13) \quad C'_i = C_i + \epsilon,$$

where ϵ is a random draw from a normal distribution with a mean of zero and a standard deviation of σ . Each agent's adjusted credit score is truncated to be non-negative. This stage continues until all scientists in all communities have been assigned credit.

A.2. Evolution. In this stage we use the logic of cultural evolution similar to that used in Smaldino and McElreath (2016), whereby individuals with more credit are more likely to reproduce their methods. This reflects greater success in attracting and placing grad students and postdocs, as well as in influencing other researchers. In each community, one scientist is chosen at random to "die." This does not have to represent literal death, it could also represent retirement or a change of career. Equivalently, we can think of this as a scientist choosing to learn by imitating a high-prestige scientist. Either way, a spot is now open for a new scientist to join the community. The new researcher may be replaced by a student or acolyte of the retiring researcher, or they may be someone whose methods are drawn from another scientist inside or outside of the community. Practically, the algorithm works like this. An individual i is randomly chosen to adopt a new method. They do so by randomly selecting another researcher j as a target for possible imitation. Agent i copies the method of agent j with a probability given by the following sigmoid function:

$$(14) \quad \Pr(i \text{ copies } j) = \frac{1}{1 + \exp[-\gamma(C_j - C_i)]},$$

where γ is a constant that reflects the strength of selection. For simplicity, we used $\gamma = 10$ in all our simulations.

A.3. Analysis. Simulations were run for 10^5 time steps or until one of the two methods reached complete saturation across both communities. The proportions of agents in each community using either method were recorded. The agent-based model was coded in both Java and NetLogo by both authors to confirm the results. The results shown here represent NetLogo simulations. NetLogo code is available at <https://www.comses.net/codebase-release/74086cf3-3398-44b3-bf9c-ada0d70e1d86/>.