Conservatism and the Scientific State of Nature

Erich Kummerfeld Kevin J.S. Zollman*

July 30, 2014

Abstract

Those who comment on modern scientific institutions are often quick to praise institutional structures that leave scientists to their own devices. These comments reveal an underlying presumption that scientists do best when left alone – when they operate in what we call the *scientific state of nature*. Through computer simulation, we challenge this presumption by illustrating an inefficiency that arises in the scientific state of nature. This inefficiency suggests that one cannot simply presume that science is most efficient when institutional control is absent. In some situations actively encouraging unpopular, risky science would improve scientific outcomes.

*The authors would like to thank Liam Bright, P. Kyle Stanford, several anonymous referees, and audiences at conferences in Montreal, Nancy, and Pittsburgh for their helpful comments and suggestions. This work was supported by the National Science Foundation under Grant No. SES 1026586. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

> Scientific progress on a broad front results from the free play of free intellects, working on subjects of their own choice, in the manner dictated by their curiosity for the exploration of the unknown.
>
> Vannevar Bush
>
> *The Endless Frontier*

1 Introduction

Mario Capecchi had a crazy idea. Perhaps one could find a way to replace a single gene in the genome of a mouse, which would allow unprecedented investigation into the effects that genes have on an organism. Playing the role of a dutiful scientist in 1980 he applied to the National Institutes of Health in the U.S. to sponsor this project. The NIH responded that the project was so fanciful as to be "not worthy of pursuit" (Gumbel, 2007). Capecchi ignored the NIH, co-opting money slated for another project to pursue his idea (Harford, 2011). Eventually he succeeded and was awarded the Nobel Prize in Physiology or Medicine for his success.

The history of science is filled with stories of the crazy idea that turned out to be right. We lionize those scientists who bucked their peers to pursue their own ideas, reinforcing the narrative about perfectly functioning science expressed by Vannevar Bush. If only scientists could be left to their own devices – if they could operate in a scientific state of nature – science as a whole would be most effective. Michael Polanyi expresses this view most clearly,[1]

> So long as each scientist keeps making the best contribution of which he is capable, and on which no one could improve (except by abandoning the problem of his own choice and thus causing an overall loss to the advancement of science), we may affirm that the pursuit of science by independent self-co-ordinated initiatives assures the most efficient

[1] A similar sentiment is echoed by Horace Kallen in a dialogue with Otto Neurath published in *PPR* (Kallen, 1946). Somewhat ironically, Kallen argued in this paper against Vennevar Bush. Not because he thought Bush was wrong about the propriety of leaving scientists to their own devices, but because he anticipated – perhaps correctly – that a national research organization would over time begin to exercise an undue homogenizing influence.

> possible organization of scientific progress. And we may add, again, that any authority which would undertake to direct the work of the scientist centrally would bring the progress of science virtually to a standstill (Polanyi, 1962).

In this paper, we argue that the reverence for the scientific state of nature as a endogenous source of diversity expressed by Bush, Polanyi, and others is misplaced. We will show that in the scientific state of nature there exists a conservative influence which will move scientific practice away from what would be best for science as a whole. Much like Hobbes' political state of nature, this conservative influence takes the form of a free-rider problem which encourages scientists to leave the burden of exploration to others. We develop a formal model of scientific practice for a group of communicating scientists and we show that in this model each individual scientist has an incentive to pursue a research project that currently looks most promising and to leave experimenting with new, radical alternatives to others. Since everyone faces the same incentives, each individual will opt to choose her scientific projects conservatively. This can lead to homogeneity in scientific practice where scientists uniformly choose to engage in safe research projects.

While the ultimate conclusion of our work is that leaving science in the state of nature is sub-optimal, one should not make the mistake of presuming that *any* form of intervention would improve science. Undoubtedly many of the recent policies of "oversight" would likely compound the problem rather than remedy it. Instead, we conclude that governments and foundations should reverse the general trend and intervene to *encourage* risky science beyond what would occur without their intervention.

In section 2 we present our model of the scientific state of nature. Utilizing this model, we show in section 3 that there exists a free-rider problem that causes individuals to be unwilling to take risks which would benefit the group as a whole. We consider (but ultimately dismiss) restricting the exchange of information between agents as one potential solution to this problem in section 4 and then conclude our discussion in section 5.

2 Scientists and Bandits

Consider the following scenario. A scientist faces two potential research projects but only has time to work on one. One project is very likely to produce a small intellectual advance, the other might produce some significant intellectual advances but might also totally flop (or perhaps some outcome that lies in between these two extremes). The scientist thinks that the safe alternative is more likely to produce a meaningful intellectual advance, but recognizes the possibility that she might be wrong about the uncertain project. How should she choose which project to pursue?

Although not all scientific decisions are analogous to this one, we believe this story represents an important decision faced by many scientists. Decisions of this sort are best modeled by a statistical problem known as a bandit problem. Originally used in statistics (Berry and Fristedt, 1985) and later in economics (Bala and Goyal, 1998; Bolton and Harris, 1999), bandit problems have recently been employed in the philosophy of science to model some aspects of scientific practice. Unlike the traditional models used in philosophy of science, bandit problems capture the fact that scientists are not passive observers of evidence but are active in determining the type of evidence that arrives.

Bandit problems are sequential decision-making problems where a player is confronted with a fixed set of options repeatedly. Imagine a gambler who must repeatedly choose which of a fixed set of slot machines to play. Each slot machine provides a payoff to the player drawn from a probability distribution that can vary across machines (but stays constant over time) and are unknown to the player. The player knows the outcome of all her previous choices before being forced to choose again. The goal of the player is to maximize her accumulated payoffs over the course of this process, which may have a finite number of rounds or be unending.

As a model for scientific practice, the different options (or "arms") represent different potential general research projects (like different theoretical commitments, paradigms, research methodologies, treatment strategies in medicine, etc.). A single pull of an arm represents an attempt by a researcher to apply that theory, method, etc. to a new, untried problem. The payoff represents the degree to which that attempted application succeeds. The strategy used by the player represents how the scientist chooses different subsequent research projects based on her past experiences with success and failures of general research strategies. This is not an accurate model of situations with a definitive experiment (such as the historical Capecchi example), however a number of examples of this model in contemporary science have been suggested (Huttegger, 2011; Mayo-Wilson et al., 2011; Zollman, 2007, 2010).

For example, Zollman (2010) advocated this as a model for research on the cause of and treatment for peptic ulcer disease. Before the discovery of *Helicobacter pylori*, there were two theories under consideration: that ulcers were caused by bacteria and that ulcers were caused by excess acid production in the stomach. Scientists worked on treatment strategies that were founded on one of these hypotheses. Each hypothesis is represented by an arm of the bandit problem and each individual attempt to treat peptic ulcer disease represents a pull of that arm. The payoff represents the degree to which that treatment strategy succeeds or fails in curing the disease. Ultimately, scientists desire to find the most effective treatment for peptic ulcer disease – to pull the arm which, on average, provides the highest payoff.

Mayo-Wilson et al. (2011) argue that research in psychology follows a similar pattern. For example, in research on concept formation there are a number of broad theories postulating different cognitive mechanisms for categorization. Each broad theory is represented by an arm. Individual experiments will rely on one or another broad theory; this is a single pull of the arm. These experiments can be illuminating (provide a high payoff) or can be flops (provide a low payoff). Ultimately, scientists want to employ the broad theory which will be most epistemically illuminating – to pull the arm which, on average, provides the highest payoff.

Bandit problems capture an important aspect of many real-world learning problems where there is a trade-off between *exploiting* the action which appears best (based on the payoffs that one has observed) and *exploring* other options. For example, if a researcher was interested in developing a more effective treatment for peptic ulcer disease, the evidence available from 1954 until at least the mid 1980's suggested that the best course of action would be to focus on improving acid suppression either through surgery, drugs, or lifestyle and diet changes. Each member of that scientific community individually chose to pursue the option that seemed best, the excess acid hypothesis, and left experimenting with the bacterial hypothesis to others. As a result, the entirety of scientific research on the topic was focused on acid suppression, and essentially no exploration of the bacterial hypothesis occurred. Only after two relatively unknown scientists explored a wild idea was it discovered that the bacterial hypothesis was correct. In this particular historical setting scientists were exploring too little and exploiting too much (cf. Zollman, 2010).

To illustrate this tension more abstractly, consider two different strategies for approaching a bandit problem. Suppose on each round one person, call her *Exploiter*, chooses the arm which has done best on average up until that round. Consider another person, *Explorer* who alternates between the available arms on every round.

Suppose that, unbeknownst to Exploiter, she is confronted by two arms, one of which always pays \$1 and the other pays \$0 half the time and \$3 the other half (making the second arm superior). For simplicity, assume that she starts by pulling each arm once, so that she has a single data point to begin estimating the expected payoff of each arm. With a 0.5 probability, Exploiter received \$0 when she pulled the second arm and so will choose the first arm (which is suboptimal) for the remainder of her pulls. Even if she received \$3 from the second arm on the first pull, she might switch if the second arm's average payoff drops below \$1. Once she switches to the sub-optimal arm, she will never switch back.

Now consider Explorer. Explorer will alternate between choosing arm one and arm two throughout his trials. By the end of the trials, Explorer will have the best estimate of the value of the two arms, but he will have failed to take advantage of the information. Suppose that he is unknowingly confronted by a relatively easy two arm bandit. One arm pays \$1 on every pull, the other pays \$0 on every pull. Explorer will pull the second arm half the time, and thus receive an average payoff of \$0.50. One might be able to do better by learning over time to avoid the second arm.

Exploiter does badly because she does not investigate sufficiently so that she can determine which is the best arm, while Explorer does badly because he does not use the information he collects. Thought of as big choices in scientific investigation, Exploiter has abandoned a superior scientific theory too quickly because it had an early failure. (This is essentially what happened to the scientific community studying peptic ulcer disease.) Explorer continued to apply an inferior theory, e.g. geocentric models of the solar system, long after it was shown to be inferior to another theory. Neither extreme is desirable; the optimal strategy will involve a combination of exploring and exploiting.

In this paper, we focus on a set of strategies for tackling bandit problems known as ϵ-greedy strategies (Sutton and Barto, 1998; Mayo-Wilson et al., 2011; Huttegger, 2011). These strategies allow for an explicit representation of a scientist's willingness to explore apparently worse alternatives: high ϵ's correspond to high exploration rates, and low ϵ's correspond to low exploration rates. When utilizing an ϵ-greedy strategy, an agent will pursue what she considers to be the best line of research (pull the arm which has given the highest average payoff so far) with probability $1 - \epsilon$ and will pursue a less proven line of research (apparently suboptimal arm) with probability

ϵ.[2] One can show that, in the limit, players will converge to a strategy that plays the optimal arm with probability $(1 - \epsilon)$, this guarantees that regardless of how they communicate in the long run any number of players will come to agree about the best arm. Considering only the limit analysis smaller ϵ's (that are not zero) are better than larger ones. But in the short- and medium- run these results do not hold; groups may not agree, different patterns of communication may affect the reliability of the group, and there will be an optimal ϵ which is greater than zero (Sutton and Barto, 1998; Zollman, 2007).

By representing scientists as choosing an arm using an ϵ-greedy strategy, we are making a number of assumptions which should be made explicit. First, we are assuming that scientists in the state of nature only care about their own intellectual contributions and not the contributions of others: the utility gained is the scientist's quantitative assessment of every way in which she benefited from her experiment. In this respect, we are presuming a particular form for the motivations of scientists. This might come about either from some intrinsic source of motivation or by a pre-existing institutional structure like a professional organization that only rewards individual success. This is represented by modeling scientists as maximizing *their own* payoff and not the payoffs of others. While perhaps an extreme this represents a reasonable starting point for modeling the motivation of scientists. Should scientists be only partially motivated by a desire to assist in others' discoveries we believe something like the problem we discuss will remain. We return to this issue in the conclusion.

Second, by modeling scientists as ϵ-greedy agents, we are taking into account their intelligence but also their natural restrictions insofar as they are not perfectly rational beings. In past studies, scientists have been modeled as purely greedy agents (Zollman, 2007, 2009, 2010) who always choose the option that they believe is best. We believe that these models are too simple; actual scientists may be aware of the epistemic benefit of exploring other options and may not completely ignore untested waters. But we do not go so far as to model scientists as full expected utility maximizers, who would calculate the optimal lever to pull at each round of play. Computing optimal play in many bandit problems is sufficiently difficult so as to be prohibitive in many realistic situations. ϵ-greedy strategies are sophisticated enough to offer significant improvement over purely greedy

[2] This is a slight modification to the standard definition of ϵ-greedy. We do this for simplicity only: our qualitative results hold equally well for standard ϵ-greedy agents.

behavior, but tractable enough to be implemented by actual people.[3] Thus, ϵ-greedy strategies are a plausible idealization of the actual strategies employed by scientists.

Third, we assume that the scientists have the same utility function. That is, the probability distribution over payoffs from a given lever does not vary between scientists. As such, it isn't the case that a lever could be best for one scientist but not best for another; the best lever is best for all scientists. This is a substantive assumption which we will return to in the conclusion.

Given the complexity of this problem, we have chosen to focus on a particular bandit problem that captures a classic scientific conflict between projects that are likely to produce small intellectual advances (akin to Kuhnian "Normal Science") and projects that have much more uncertain prospects (akin to employing a new theoretical apparatus). It may be helpful to consider a fictionalized example, that of an astronomer circa 1550 interested in predicting the locations of heavenly bodies. The astronomer is tasked with predicting first the location of Mars at a certain date, later that of Venus, and so forth. Since our astronomer is mortal, she will only make a finite number of predictions. She is aware of both the Ptolemaic and Copernican paradigms, but has at her disposal only enough resources to use one or the other each time she makes a prediction. She wants to make the best predictions she can, but doesn't know which paradigm will serve her better. Further, the accuracy of any prediction is prone to some degree of randomness from many sources, e.g. imperfections in the measurement instruments or mistakes made by the astronomer in applying the paradigm.

When our astronomer makes a prediction, she communicates the results to her peers in astronomy. These peers are also making predictions, are aware of both paradigms, and communicate their results to each other and to her as well. Our astronomer can look back at the successes and failures of past predictions made by herself and by her peers in order to decide which paradigm she will use to make her next prediction. Since our astronomer is more familiar with how to apply the Ptolemaic paradigm to make predictions, the average quality of the Ptolemaic predictions can be treated as a baseline from which the Copernican predictions are evaluated. This simplifies the issue for the astronomer somewhat to evaluating whether the Copernican paradigm gives, on average, better or worse predictions than the Ptolemaic paradigm.

[3] We could have allowed our scientists to decrease their ϵ values as they gain more data, but this would add more parameters to the model with relatively little benefit, as our agents only take 50 pulls in our simulations.

We have very roughly quantified this example with the following bandit problem, which we will analyze later in this paper. The agents (astronomers) are allowed 50 pulls each on a two-armed bandit where one arm always returns a value of 0 (the Ptolemaic paradigm, the safe choice) and the other arm returns a value drawn from a Normal distribution with mean 1 and variance 9 (the Copernican paradigm, the risky choice).[4] We believe that our results are robust across a wide range of numerical assumptions; they do not depend on the fact that one arm has 0 variance, as opposed to a smaller variance than the other, or that the agents pull exactly 50 times, and so on.

3 Choosing an ϵ

Suppose that we have a single scientist acting in isolation who will use an ϵ-greedy strategy. What ϵ should she use? How willing should she be to explore apparently bad options? This question can be answered, usually with the aid of computer simulations, given that we know something about what type of bandits the scientist might face (Sutton and Barto, 1998).

This problem becomes more complex if our scientist is not a lone investigator but inhabits a community of other scientists who are themselves confronted by the same bandits and who share their results after each pull. In this situation there is some incentive to leave the experimentation to others. As an extreme example, suppose that one knows there is another member of the community who will be maximally experimental (that is who has an $\epsilon = 0.5$). If one is able to see that person's failures and successes, one does best by reducing one's own experimentation and "free-riding" on the experimentation of the other.[5]

It is this possibility that will concern us here. When scientists are left to their own devices to choose their level of experimentation, will they choose a level of experimentation that makes the group as a whole epistemically best? That is, in the scientific state of nature will a community of scientists self-organize in a way that maximizes the sum of the agents' utilities (will it be socially optimal)? Or will they choose to free-ride on others' experimentation and make the community as a whole worse off by decreasing the sum of the agents' utilities?

[4]The reader should not take '0' to mark a substantive assumption about the effectiveness of the program. This is an arbitrary choice for fixing a baseline utility. The particular values in all our examples can be freely rescaled using any positive affine transformation without altering the results.

[5]The possibility for experimental free-riding in bandit problems has already been noted by Bolton and Harris (1999), although they consider a different type of bandit problem and a different set of strategies.

We have investigated these questions by analyzing what we call the Choose-Your-ϵ game. In this game, each player's goal is to maximize expected payoff from a finite-round bandit problem with some number of other players. Each player is restricted to the set of strategies consisting only of ϵ-greedy strategies – each player must choose, before any arms are pulled, what ϵ to use. Players cannot change their ϵ's after this initial choice, and they make this choice without knowledge of the ϵ's of other players (it is a simultaneous-move game). With their ϵ's chosen, the players then play a finite bandit problem where they observe the outcomes of their own actions and the outcomes of (some) other players after each pull. The value of a particular choice of ϵ against other choices is based on how, on average, an individual utilizing that ϵ would fare in this finite bandit problem.

Why should the scientists be constrained to choose their ϵ only at the beginning? We believe a scientist's ϵ represents a personality trait, it is a measure of how prone the scientist is to pursuing risky lines of research. Such characteristics change slowly with respect to individuals, so assuming that the ϵ remains fixed for the duration of play is a plausible simplification. Furthermore, it is unclear what a reasonable way to change one's ϵ would be. It seems unlikely that one could condition one's choice of experimentation on the choices of others, because determining what is another's choice of ϵ is difficult. While one can (sometimes) determine whether another person has experimented with an inferior arm, it would require further inference to estimate her ϵ. Given that in our game there are only a small number of choices made by the scientists, this estimate would be very noisy.

We are not presuming that an individual's ϵ is fixed entirely randomly by some outside force. If it was fixed in this way, nothing could be gained from a game theoretic analysis which models the ϵ's as either the subject of deliberation or the result of a selective process. Instead, we conceive an individual's ϵ as malleable by the individual, but only on a relatively long time scale. Individual scientists will not change their willingness to experiment from day to day, but they might change them over their entire careers. Or even if they are fixed for an individual scientist's lifetime this willingness might be passed down from supervisor to student in a kind of cultural evolutionary process.

Because of its complexity, we have analyzed this game with a combination of computer simulations and analytic game theory. Computer simulations were used to estimate the expected payoffs for various ϵ-greedy strategies. We used these data points to infer two primary characteristics of this game. First, we estimated the ϵ which, when adopted by every individual member, provides the entire community of scientists with the largest overall payoff, i.e. the ϵ that, when universally adopted, maximizes the sum of utilities of the scientists. This represents what the scientists might collectively agree to if they had some sort of enforcement mechanism.

Second, we inferred the symmetric *Nash equilibrium* for this game. The Nash equilibrium is a solution concept in game theory. It is widely used in economics as a method for identifying expected outcomes of games. Intuitively, Nash equilibria are stable situations where no player has any reason to alter their strategy. If the players are not in a Nash equilibrium, then at least one player has reason to change her strategy and would be expected to do so.

We inferred a Nash equilibrium for this game by first estimating a best response function for an individual agent, called the *hero*, when the rest of the community of players, called the *group*, play a specified ϵ.[6] Said another way, we ask what ϵ a hero would choose if she knew that everyone else in her community was playing a different ϵ'. If the hero would choose the same ϵ as that played by the entire group, then that ϵ represents a symmetric Nash equilibrium of the game. We employed simulations to estimate the payoff for the hero for particular ϵ's played by the hero and by the group, respectively (the details are later in this section). For each group ϵ we identified the ϵ which provided the highest payoff to the hero.[7] By doing this for a range of different group ϵ's we collected a number of data points of the form $\langle$group ϵ, approximate hero's best response$\rangle$. We then regressed the approximate best responses on the group ϵ's (see Figures 1 and 4). The result of this regression is an estimate of the best response function for the hero.

This estimated best response function is used to calculate a symmetric Nash equilibrium for the game, i.e., one in which all players choose the same ϵ. We find the symmetric Nash equilibria by finding points where the estimated best response function intersects the $x = y$ line.[8] In our case, there is a unique such point for each game, as these best response functions have positive intercept and negative slope. As such, there is reason to expect scientists to play that ϵ. We have identified this Nash equilibrium, and found that the ϵ chosen by communicating scientists in the symmetric

[6] It should be noted that this is an inferential technique, used to identify the symmetric Nash equilibria of the game, and not a substantive empirical assumption about the homogeneity of the scientific community.

[7] We calculate the optimal ϵ by estimating the payoff for a few initial values of ϵ and then, using a decreasing step size, continue to calculate ϵ's that outperform the initial best guess for the optimal. Once the step size reaches $(1/2)^7$ we stop and take the current best guess as the optimal ϵ.

[8] Because the game is symmetric, and because the payoffs are continuous in the choice of ϵ, we can be assured that there is a symmetric Nash equilibrium of the game.

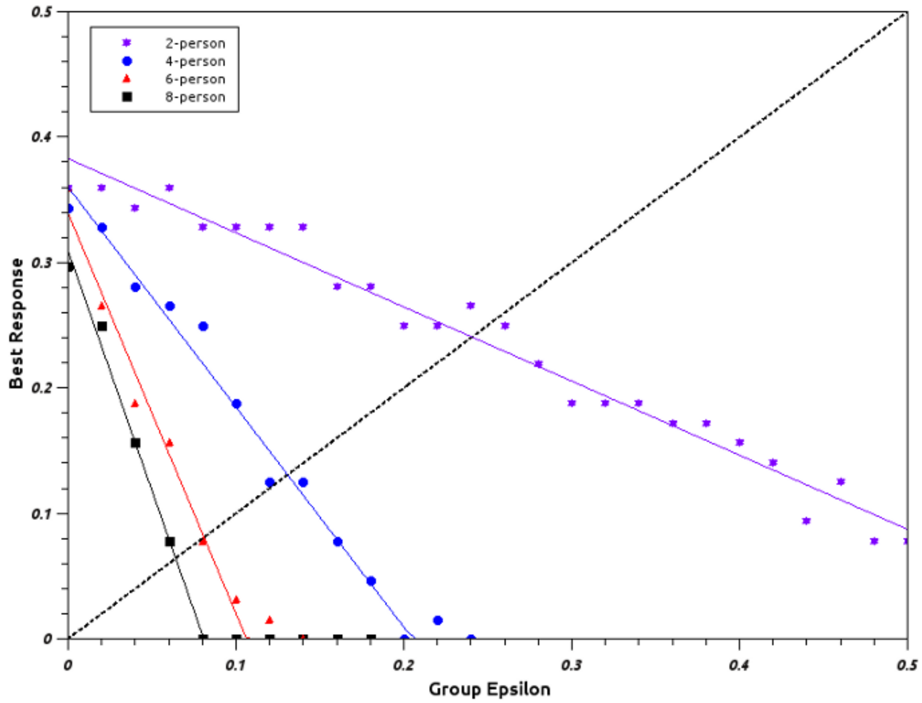

Figure 1: Simulated best response values for hero ϵ in communities of various sizes. The lines represent the best fit lines for each size to the simulated points. We estimate the symmetric Nash equilibrium of this game by finding where the best fit lines cross the line $x = y$ (also pictured).

Nash equilibrium is far lower than the optimal ϵ for the community. In other words, the state of nature of scientific practice is suboptimal.[9]

We will first consider the case where all scientists communicate with all other scientists (in the next section, we will consider the possibility that scientists only share their results with a small number of their peers). The data points in Figure 1 represent the best responses for a hero agent, for each group size, and each group ϵ, as determined by these simulations. The lines are linear regressions for each group size.[10] Although an analytic solution to the game is intractable, these

[9]Because of the complexity of the game, we were only able to infer the *symmetric* Nash equilibria – those equilibria where everyone is playing the same strategy. Because the best response curve in the 2-player case is not the curve $y = (1 - x)$ there cannot be an asymmetric Nash equilibrium in this case. (We thank an anonymous reviewer for pointing this out.) There may be asymmetric Nash equilibria in situations with more than two players which we do not identify. While it is conceivable that there exist asymmetric equilibria that do not feature the free riding problem that we uncover, we believe this is an unlikely result.

[10]We ran simulations for groups of size 2, 4, 6, and 8. For each size, we ran simulations with the group choosing ϵ's ranging from 0 to .5 in intervals of .02. For each group size and each group ϵ, we ran 10 million simulations for each of the hero's possible response ϵ's, and identified her best response to two decimal points of precision.

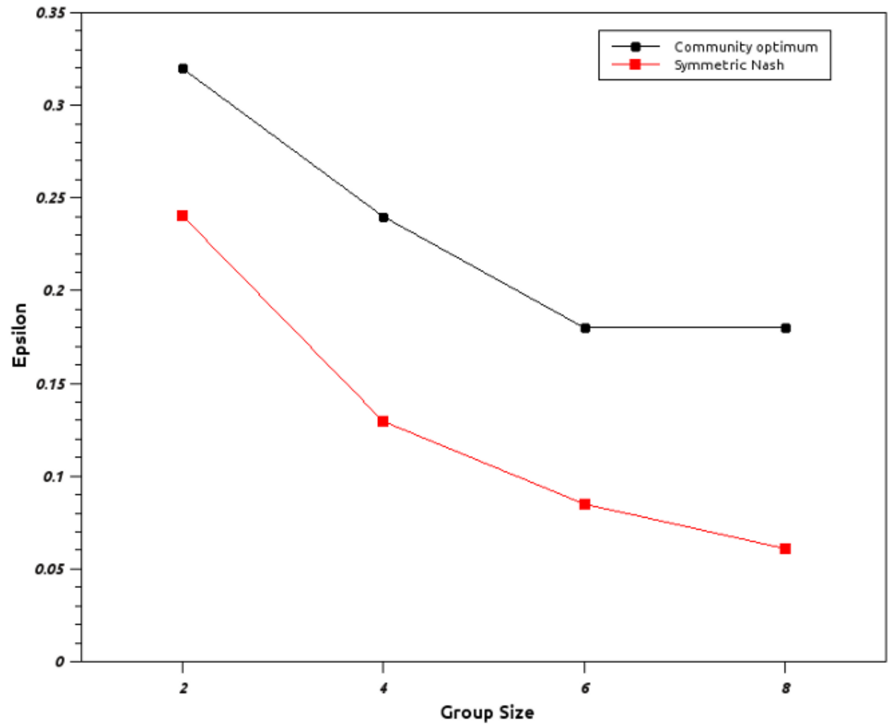

Figure 2: The location of the community optimum ϵ and symmetric Nash equilibrium ϵ for different group sizes. The separation is indicative of a free-rider problem.

linear regressions fit the data very well,[11] and this leads us to believe that the actual best response curve for each group size is a function close to the appropriate linear regression. Figure 1 also contains the line $x = y$, where the hero agent is playing the same ϵ value as the group. The intersections of this line with the regressions for the various group sizes represent the symmetric Nash equilibria for each group size.

The symmetric Nash equilibria, as determined by the intersections in Figure 1, are plotted in red squares in Figure 2. The optimal community ϵ's are also plotted. As is found in social dilemmas, like the Prisoner's dilemma, individual optimization leads to collective behavior that is worse than what could be secured by enforcing a collective agreement to do otherwise. Therein lies the free-rider problem: every agent's best response to the community optimum is to experiment less (see Figure 1), in the same way that it is advantageous to defect against a cooperator in the Prisoner's Dilemma. But if every agent experiments less (defects) then every agent is worse off. In

[11] Mean squared errors for fully-connected network, for group sizes $2, 4, 6, 8 : 2.42e^{-04}, 2.22e^{-04}, 2.59e^{-04}, 5.80e^{-05}$.

terms of scientific practice, this means that scientists trying to maximize their individual epistemic contributions would be expected to choose scientific methodologies conservatively. That is, we would expect epistemically self-interested scientists to adopt the methodology that appears to be most successful at the moment more often than is beneficial for the scientific community at large. This behavior is not the result of funding practices, peer review, graduate student training, or social power, but endogenously arises as a result of the underlying problem the scientists face.

One might wonder the degree to which Nash equilibria represent appropriate predictions. Indeed, the use of the Nash equilibrium as a predictive tool is often criticized. Critics presume that the maximization calculations performed by the modeler represent a cognitive process used by real individuals when making decision. These critics believe that game theoretic models require that each individual carefully predicts what the other actors will do and calculates what she should do in response. Given the complexity of this problem, it is unreasonable to suppose that scientists at various stages in their careers are estimating exactly what level of experimentation is appropriate.

There is an alternative interpretation of the Nash equilibrium concept that avoids these criticisms. Consider a population of scientists who are not playing a Nash equilibrium in this game – in particular, presume that some have higher ϵ's than others. Those with lower ϵ's will, on average, make a larger total epistemic contribution because they leave the experimentation to others and increase their individual scientific productivity (at the cost of the community's productivity). Others might imitate them, or the more successful scientists might have more students who they train to have the same general attitude to scientific problems. Over time, the group of scientists that is out of equilibrium might come to move in the direction of the equilibrium.

This interpretation of what can be learned from finding Nash equilibria of a game is often called "evolutionary game theory," where evolution is taken to mean any change over time and not only biological evolution via natural selection. Although we have not offered an explicit evolutionary model, we support this interpretation of our work and we believe it avoids concerns that attribute too sophisticated Machiavellian planning to individual scientists.

4 Structure of communication

Zollman (2007, 2010) considered a situation where individuals play a strategy that closely resem-

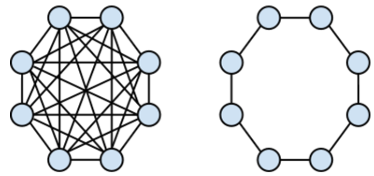

Figure 3: The complete and ring networks with 8 agents.

bled the Exploiter strategy (or alternatively a situation where $\epsilon = 0$).[12] He found that because the individual scientists did not experiment sufficiently on their own, the community did better when communication was restricted. By restricting communication, diversity in theory choice was preserved and the community as a whole did better. In this section we investigate the effect that limiting communication might have on our results.

Figure 3 shows two different types of networks of communication that are explored in this paper. In these graphs, nodes represent agents and lines represent symmetric channels of communication, such that connected agents share all their evidence with each other. The left hand graph represents a *fully-connected* network of agents which we have analyzed in the preceding section. In fully-connected networks, every agent has access to the evidence of the entire community before making each choice. Combined with the assumption that all agents share the same utility function, this results in the entire community agreeing each round on which lever is the greedy option.

On the right is a *ring* network which we will analyze in this section. In a ring network, every agent communicates with only two agents, her neighbors, and her neighbors don't communicate with one another (assuming more than 3 agents). As such, an agent doesn't have access to some of the evidence that her neighbors have. No two agents get to see the same set of evidence, and so there may be no consensus on which lever is the greedy option, meaning that two ϵ-greedy agents could both act greedily but pull different levers.

Again, we estimated the optimum ϵ for ring-connected populations of agents by running simulations with ring-connected communities playing a variety of different ϵ's. Each data point was the result of 2 million simulated iterations of the community playing a 50 round bandit problem with a given ϵ. The data points for the optimal ϵ of fully-connected networks in Figure 2 were estimated

[12]The agents Zollman considered used Bayesian reasoning combined with an exploiting response rule. Because the agents had a prior about the value of different arms, they might ignore early bad results because their prior was not sufficiently altered by the data. In this respect the agents in his model were slightly different from pure exploiting agents.

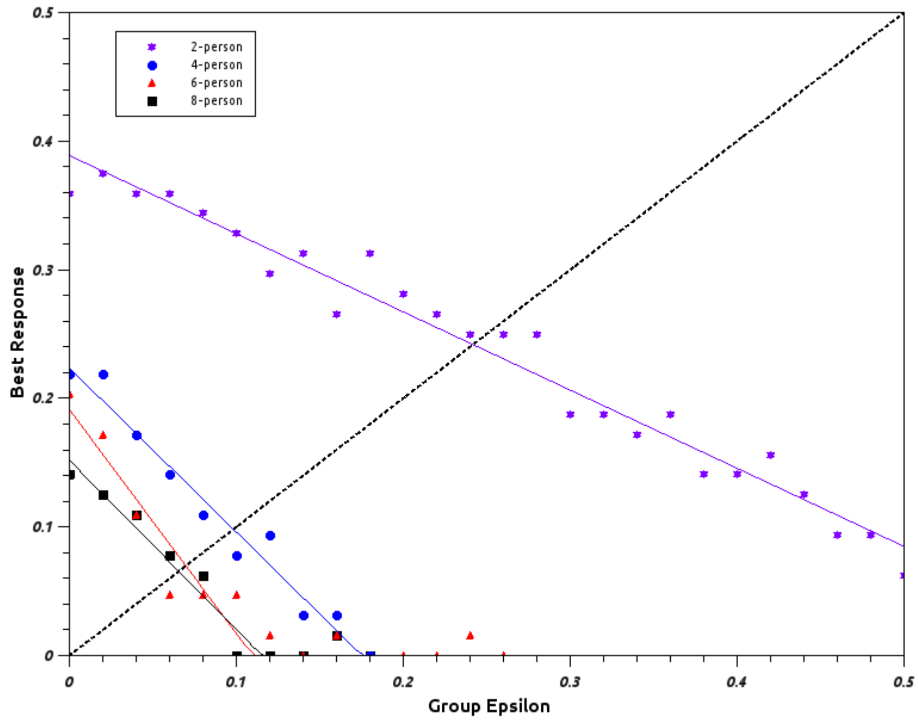

Figure 4: Regressions on simulation data with ring.

in the same way.

Figure 4 shows the linear regressions for each group size. The linear regressions fit the simulation data well,[13] so again there is reason to believe that the actual best response functions for these heroes are close to the regressions. Some of these regressions are notably different from those of the fully-connected network. In particular, the regressions for the larger group sizes show significant deviation from those of the fully-connected network, with much lower intercepts but also much shallower slopes. One might notice that the line from the 6 person ring intersects with the line for the 8 person ring close to the origin. We believe this is a result of the noise introduced by there being very few group ϵ's where the hero should adopt any $\epsilon > 0$ and not representative of a significant result.

Figure 5 compares the location of the community optimum and the symmetric Nash equilibrium. Again we find the same pattern as we do in a society that communicates widely. The community would do better if they could enforce a higher rate of exploration than would be obtained naturally

[13]Mean squared errors for the ring network, for group sizes $2, 4, 6, 8 : 3.04e^{-04}, 2.29e^{-04}, 3.14e^{-04}, 1.46e^{-04}$.

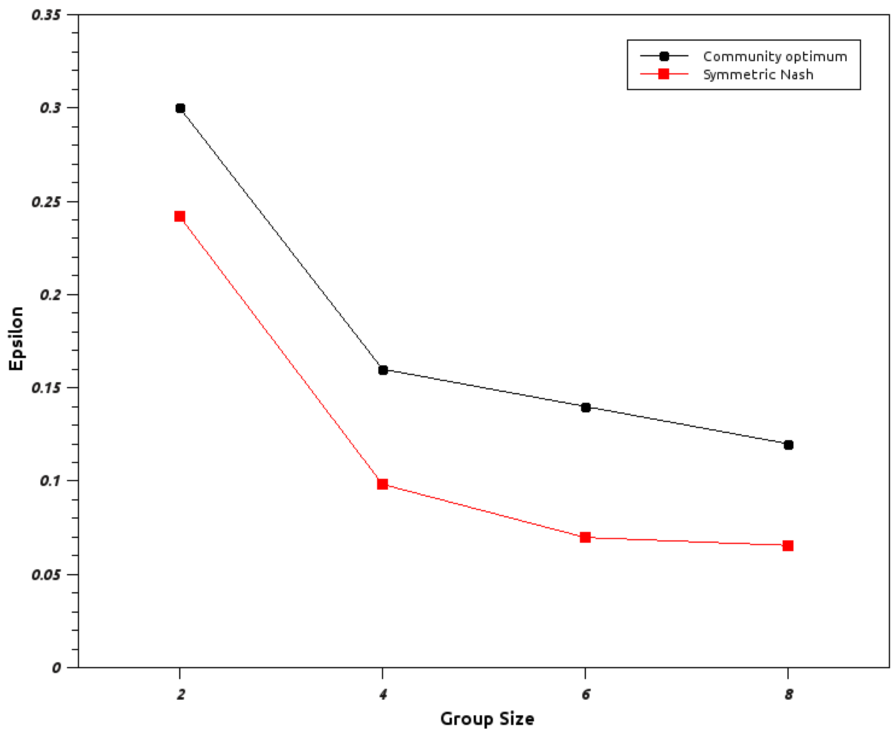

Figure 5: The location of the community optimum ϵ and symmetric Nash equilibrium ϵ for different group sizes arranged in the ring. Again we find a free-rider problem.

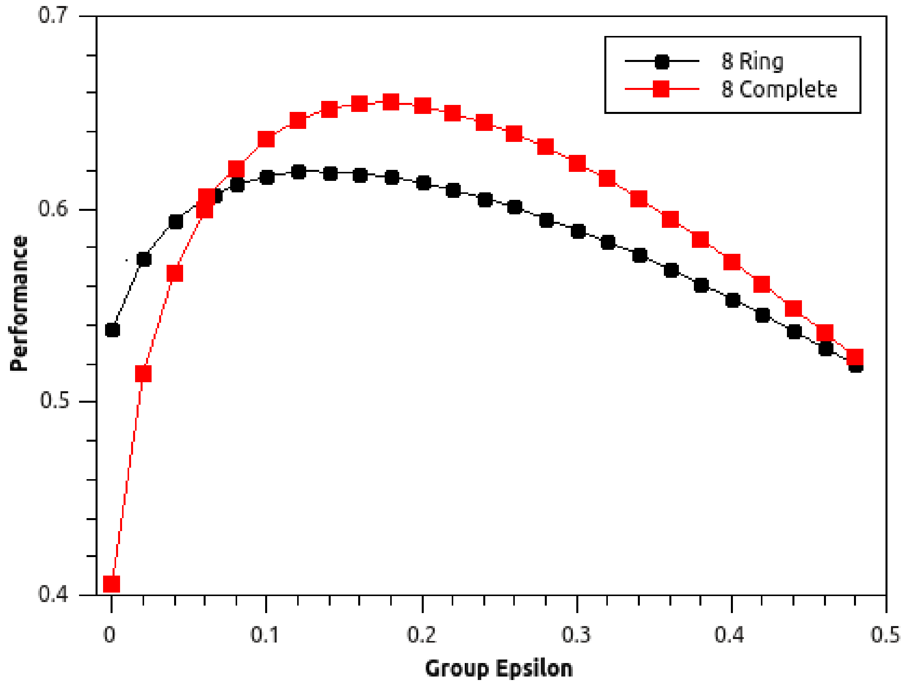

Figure 6: The expected payoff from universal adoption of various ϵ for communities arranged in the ring and complete graph.

by free choice of the individual scientists.

Recall that Zollman (2007, 2010) found that, for non-exploring agents, the ring outperformed the complete graph. Does this same result obtain when agents can choose their rate of exploration? Figure 6 illustrates the performance of 8 agents connected in a ring compared to the performance of the same number of fully-connected agents. In these simulations, the entire group plays the same ϵ, as determined by the x-axis, with the resulting expected payoff displayed by the y-axis. As we expected, the optimal community ϵ for a ring is less than for the fully-connected network because a ring community maintains a level of diversity longer through restricted information flow. With lower ϵ's experimentation is maintained by the structure of the community, and they therefore perform better. On the other hand, for groups that experiment more, we can see from Figure 6 that they do better if they are fully-connected than if they are in a ring. This represents an important caveat to Zollman's (2007; 2010) work. Communication only is harmful when exploration is low.[14]

[14]To be clear, we are not claiming to have demonstrated that Zollman's results are artifacts of an idealization in his model. Instead, his model looked at one extreme, and his results are robust to small deviations from that extreme. However, they are not robust to *all* deviations. This suggests that the degree to which his results are true depend on

If one compares the the performance of the two groups when each group is at its Nash equilibrium, one finds that both the ring and complete network do equally well. We do not expect that this is a general phenomena, but rather an accident of our particular assumptions. We leave exploring this issue for future work.

5 Discussion

While our simulations used specific parameters, we believe that our results are robust across a large range of such parameters. What is crucial, however, is that the general structure of the learning problem faced by scientists is appropriately modeled as a bandit problem. Should a community of scientists face a different learning problem – and some certainly do – our results would not apply. As illustrated by the examples of peptic ulcer disease and concept formation, many scientists do face bandit-like problem, and therefore our results illustrate an important problem with the scientific state of nature.

Our model makes a number of idealizing assumptions which should be made explicit. First, we assume that scientists are all equally good at generating results utilizing the different theoretical approaches. Undoubtedly, in reality some scientists might be better or worse at implementing particular theories, and this might lead them to try out the theory with which they are more adept. Our results would be tempered (or possibly totally eliminated) by this diversity. However, we do not believe that the success of science should rest on this diversity arising endogenously because it would be easy for external factors to destroy it. Instead, we argue, this sort of diversity should be actively encouraged in science through some external institutional involvement.

Second, we assume that all scientists attribute the same intrinsic value to a particular contribution. Contra Kuhn, they do not disagree about the standards of the scientific enterprise. Given the apparent large scale agreement among scientists regarding what does or does not represent a significant contribution to our body of knowledge, we do not believe this to be an unreasonable assumption. Kuhn and others have made much out of apparent disagreements over epistemological standards during revolutionary periods in science. Although scientists do occasionally disagree over

the degree to which scientists are willing to explore risky alternatives, and this may vary from field to field and over time. Indeed, Zollman's central conclusion – that transient diversity is necessary for effective sciences – is replicated by this study which illustrates another method by which that diversity can be achieved.

the contribution of a given work, we believe those disagreements are relatively small compared to what remains shared and that they occur relatively rarely.

Even if scientists disagree about fundamental parts of science in way supposed by Kuhn, our model would still apply if scientists found themselves facing the same incentives. If, for instance, the scientists were all seeking credit from the same central body (a scientific society or funding agency) there incentives would track what that body found interesting. In such a case the scientists would be appropriately modeled in the way we have done here.[15]

We have shown that under these idealized circumstances scientists left to their own devices will tend to experiment less than would otherwise be desired. There is a free-rider effect, where scientists have a positive incentive to leave experimentation to others. This shows that the scientific state of nature valorized by Bush, Polyani, and others may not be as rosy as they suggest.

We should caution against a few misreadings of this conclusion. First, we do not claim that science as it is now practiced is in need of *further* institutional control. Scientists do not now operate in the scientific state of nature. There are grants and prizes (like the Nobel prize) some of which explicitly encourage innovative or novel research. These may function to counteract the free-rider problem we identify here.[16]

We do believe that further research should address what institutional controls may already exist or might be further created to help to counter this apparent problem. Our second set of simulations shows that one solution proposed in the literature, restricting the scientists' communication, will not succeed at combating this problem. The literature on free-rider problems is filled with a number of solutions, but all of them turn, essentially, on changing the underlying choice problem by (a) providing a mechanism for reward or punishment or (b) by introducing some mechanism for cooperative individuals to associate disproportionately with one another. Because one cannot observe another's ϵ directly, this latter solution is difficult, while the former solution would represent a move away from the scientific state of nature.

Another possible counteracting force is that scientists might be altruistic. As mentioned in Section 2, by modeling scientists as ϵ-greedy agents we are treating them as epistemically selfish

[15] We thank an anonymous reviewer for making this point.

[16] Kitcher (1993) provides a justification for non-epistemic motivations of scientists which is similar to the one we discuss here. Kitcher focuses on the choice between two potential solutions to a single problem where the different solutions benefit from the number of scientists who pursue that solution. Our choice situation is rather different. While not a reproduction of his argument, we believe that our's is complementary to his.

– they care only about maximizing *their* contributions to the corpus of knowledge. If scientists inhabited the other extreme – they cared only about maximizing the corpus of knowledge without any interest in who made the contribution – our free-rider problem could not arise. Free-riding does not arise in situations where all agents receive identical payoffs, which would occur if the scientists were maximally altruistic. (There are other problems, known as "coordination problems" that can arise, but these are beyond the scope of this paper.) What is most likely, however, is that scientists occupy a middle ground, they care both about maximizing knowledge and about how much they themselves have contributed. In such a situation there is room for the free-rider problem to arise, although its severity would depend on many particulars.

We have modeled our scientists as "risk neutral" individuals, in the economic sense of that word. Our scientists neither enjoy nor fear taking risk for its own sake – they consider only the expected payoff of an arm and not the risk associated with it. It seems clear that, in this model, were scientists risk averse the problem we present would be further exasperated as the riskier arm is also the superior one. But, were scientists risk seeking – where they to enjoy taking risks to such an extent that they would seek out theories that are worse but riskier – then our results might no longer hold. We do not believe this is a likely model of scientists' preferences, but further research should be conducted to determine the robustness of our conclusions to this assumption.

Even if it turns out that contemporary controls are not sufficient to counteract the free-rider problem we identify here, this research most certainly does not justify just any form of institutional control. While we argue that the scientific state of nature is sub-optimal, it could undoubtedly be made worse by institutional settings that actively discourage novel and risky research – indeed many scientists currently argue this is the case with government controlled funding.

What our research adds to this debate is not a conclusion about the effectiveness of current institutional controls, instead it shows that one cannot immediately dismiss institutional control as prima facia bad for science. The scientific state of nature is not a utopia, even if everyone is acting in good faith. Instead, we must engage in the complicated discussion of the effect of this or that institutional control for the effectiveness of science as a whole.

References

Bala, V. and S. Goyal (1998). Learning from Neighbours. *Review of Economic Studies 65*, 565–621.

Berry, D. A. and B. Fristedt (1985). *Bandit Problems: Sequential Allocation of Experiments.* London: Chapman and Hall.

Bolton, P. and C. Harris (1999, March). Strategic Experimentation. *Econometrica 67*(2), 349–374.

Gumbel, A. (2007, October). Mario Capecchi: The man who changed our world. *The Independent*.

Harford, T. (2011, May). Positive Black Swans How to fund research so that it generates insanely great ideas, not pretty good ones. *Slate*.

Huttegger, S. (2011). Carnapian Inductive Logic for Two-Armed Bandits. *Manuscript*.

Kallen, H. M. (1946). The Meanings of " Unity " Among the Sciences , Once More. *Philosophy and Phenomenological Research 6*(4), 493–496.

Kitcher, P. (1993). *The Advancement of Science.* New York: Oxford University Press.

Mayo-Wilson, C., K. J. S. Zollman, and D. Danks (2011). The Independence Thesis : When Individual and Social Epistemology Diverge. *Philosophy of Science 78*(4), 653–677.

Polanyi, M. (1962). The Republic of Science : Its Political and Economic Theory. *Minerva 1*, 54–74.

Sutton, R. and A. Barto (1998). *Reinforcement learning.* Cambridge: MIT Press.

Zollman, K. J. S. (2007, December). The Communication Structure of Epistemic Communities. *Philosophy of Science 74*(5), 574–587.

Zollman, K. J. S. (2009, June). Optimal Publishing Strategies. *Episteme 6*(2), 185–199.

Zollman, K. J. S. (2010). The Epistemic Benefit of Transient Diversity. *Erkenntnis 72*(1), 17–35.