The Credit Economy and the Economic Rationality of Science

Kevin J.S. Zollman*

Forthcoming in Journal of Philosophy
Draft: June 18, 2017

Abstract

Theories of scientific rationality typically pertain to belief. This paper argues that we should expand our focus to include motivations as well as belief. An economic model is used to evaluate whether science is best served by scientists motivated only by truth, only by credit, or by both truth and credit. In many, but not all, situations scientists motivated by both truth and credit should be judged as the most rational scientists.

Understanding rational scientific behavior has long been a central interest in philosophy of science. Almost all the varied approaches to this problem share an assumption of passive individualism. There is a single scientist. The scientist is entertaining several theories which are all live alternatives. Evidence arrives that might relate to some of those theories, either by confirming, disconfirming, falsifying, or otherwise testing one or more of them.

Philosophy of science has focused most of its attention on determining how individual scientists should respond to evidence. Less attention has been paid to the more active part of science: where scientists choose what evidence to seek out. Almost none is given to the other aspects of a scientist's life: hiring, grant writing, publication, and so on.

*My thanks for helpful comments from anonymous reviewer, Liam Kofi Bright, Remco Heesen, Simon Huttegger, Philip Kitcher, Brian Skyrms, Katie Steele, Wei Wang and audiences at Columbia University, Australian National University, CLMPS in Helsinki, University of California Irvine, SEP in Hamilton, Tsinghua University, and Vanderbilt.

In part, this focus springs from the distinction between *epistemic rationality* and *instrumental rationality*.[1] Epistemic rationality pertains to the rationality of beliefs given one's evidence and one's prior beliefs, while instrumental rationality focuses on whether an action is proper given one's beliefs and one's ends. Stated crudely, epistemic rationality is exclusively about beliefs; instrumental rationality includes one's motivations and actions as well.

In philosophy of science the focus on epistemic rationality, at the expense of instrumental concerns, is a mistake. Several scholars have argued that many central issues in philosophy of science depend on understanding the rationality of science through both an epistemic and instrumental lens.[2] I illustrate this by focusing on one central part of the scientific enterprise: scientists desire to receive credit for their work. I call this the "credit economy" in science. In this paper, I show that our aims for science are usually, but not always, furthered by the credit economy.

I. The ends of scientists

What is the purpose of theories of scientific rationality? It is not merely an empty accolade. Scholars often relate the rationality of science to other "big" questions in the philosophy of science. If scientists are rational when they discard one theory in favor of another, science is progressing through individually rational decisions.[3] Furthermore, if science is progressing, it becomes possible (although not necessary) that science incrementally advancing toward the truth.

When analyzing the question of the rationality of science, philosophers usually analyze one question: what are rational constraints on how a scientist changes her mind in light of the available evidence? As an illustra-

[1]Thomas Kelly, "Epistemic Rationality As Instrumental Rationality: A Critique" *Philosophy and Phenomenological Research* LXVI, 3 (2003): 612–40.

[2]For example: David Hull, *Science as a Process* (Chicago: University of Chicago Press, 1988); Philip Kitcher, "The Division of Cognitive Labor" *The Journal of Philosophy* LXXVII, 1 (1990): 5–22; *The Advancement of Science* (New York: Oxford University Press, 1993); Miriam Solomon, *Social Empiricism* (Cambridge: MIT Press, 2001); Conor Mayo-Wilson, Kevin J.S. Zollman, and David Danks "The Independence Thesis : When Individual and Social Epistemology Diverge," *Philosophy of Science* LXXVIII, 4 (2011): 653–77.

[3]Larry Laudan, *Beyond Positivism and Relativism: Theory, Method, and Evidence*, (Boulder: Westview Press, 1996).

tive example, the Bayesian approach to this problem involves assigning prior probabilities to each available hypothesis and likelihood judgments to each potential piece of evidence conditional on each theory. A rational scientist, according to the Bayesian, then modifies her beliefs after receiving evidence by updating her probability distribution using Bayes' rule.

A Popperian would reject notions of prior probability. Instead, the scientist has a set of theories that are unfalsified and a particular conjecture chosen from among the live alternatives. If the evidence falsifies a theory, it is then discarded and a new conjecture is chosen. Alternatives to these two abound, but all define the problem of rationality in the same terms. They all impose constraints on how one changes one's beliefs or conjectures in light of evidence.

Indeed, scientists are often asked to reform their opinions because of new evidence. But, a scientist confronts many decisions that fall outside this narrow class. During a normal day, a scientist might ask herself: What experiments should I conduct? Should I continue collecting evidence or should I stop and try to publish? To which journal should I send my paper? How should I present my findings? Should I apply for this grant opportunity? Can I afford to hire another postdoc? And so on. Undoubtedly, a scientist's beliefs about various theories will bear on these decisions, but her beliefs alone will not determine the answers to these questions. Issues of instrumental rationality run throughout them; a scientist must combine her beliefs with her goals to arrive at a plan of action.

A philosopher might regard these decisions as outside the purview of a theory of "the rationality of science." Scientific rationality, one might say, is only epistemic rationality; instrumental rationality plays no part. One is free to define one's terms in whatever way one sees fit, but to follow this path would be to diminish the importance of our theories of scientific rationality.

First and foremost, this view would leave out many questions that are critical to the progress of science. A scientist's answer to the instrumental questions influences scientific progress in various ways. If a scientist publishes erroneous results, she might mislead. If she keeps her achievements secret, others might pursue dead ends. She might hire another postdoc who would make a new discovery. Should philosophy of science distance itself from these instrumental questions, we would open a chasm between our theory of scientific rationality and scientific progress.

Furthermore, treating scientific rationality as equivalent to epistemic rationality has the unpleasant effect of making actual science appear irrational.

Kitcher argues that conflating of scientific and epistemic rationality – a view he calls "Legend" – makes philosophical positions about the progress of science needlessly susceptible to criticism.[4] Decades of sociological study have demonstrated that scientists are often motivated by considerations beyond purely epistemic ones. If we insist on Legend, we will find ourselves committed to the irrationality of most existent and historical science.

The reality that scientists' are motivated by non-epistemic concerns has inspired substantial research. Scholars have investigated how scientist's career motivations might influence several large scale decisions in science. These include how to work with others,[5] the willingness to engage in sloppy or fraudulent science,[6] the choice scientific problems,[7] the desire of scientists to

[4]Kitcher, *The Advancement of Science*, *op. cit.*

[5]Siddhartha Banerjee, Ashish Goel, and Anilesh K. Krishnaswamy, "Re-incentivizing Discovery: Mechanisms for Partial-progress Sharing in Research," *Proceedings of the Fifteenth ACM Conference on Economics and Computation* (2014): 149–66; Thomas Boyer, "Is a Bird in the Hand Worth Two in the Bush? Or, Whether Scientists Should Publish Intermediate Results" *Synthese* CXCI, 1 (2014): 17–35; Justin Bruner and Cailin O'Connor "Power, Bargaining, and Collaboration" in Thomas Boyer, Conor Mayo-Wilson, and Michael Weisberg, eds., *Scientific Collaboration and Collective Knowledge* (Oxford: Oxford University Press, forthcoming); Remco Heesen, "Communism and the Incentive to Share in Science" *Philosophy of Science* (forthcoming); Kevin J.S. Zollman, "Learning to Collaborate," in Thomas Boyer, Conor Mayo-Wilson, and Michael Weisberg, eds., *Scientific Collaboration and Collective Knowledge* (Oxford: Oxford University Press, forthcoming).

[6]Liam Kofi Bright, "On Fraud", *Philosophical Studies* CLXXIV 2 (2017), 291–310; Remco Heesen, "Expediting the Flow of Knowledge Versus Rushing into Print," *Manuscript*; Paul E. Smaldino, and Richard McElreath "The Natural Selection of Bad Science," *Royal Society Open Science* III: 160384.

[7]Carl T. Bergstrom, Jacob G. Foster, and Yangbo Song, "Why Scientists Chase Big Problems: Individual Strategy and Social Optimality" *Manuscript*; Partha Dasgupta, "The Economics of Parallel Research"in Frank Hahn, ed., *The Economics of Missing Markets, Information, and Games* (Oxford: Clarendon Press, 1990), pp. 129–62; Alvin Goldman *Liasons: Philosophy Meets the Cognitive Sciences,* (Cambridge: MIT Press, 1992), pp. 225–54; John Kleinberg and Sigal Oren "Mechanisms for (Mis)allocating Scientific Credit," *Proceedings of the 43rd annual ACM symposium on theory of computing* (2011), 529—-538; Kitcher, "The Division of Cognitive Labor," *op. cit.*; *The Advancement of Science*, *op. cit.*; Erich Kummerfeld and Kevin J.S. Zollman "Conservatism and the Scientific State of Nature" *British Journal for the Philosophy of Science* LXVI, 4 (2017): 1057–76; Michael Strevens, "The Role of the Priority Rule in Science" *Journal of Philosophy* C, 2 (2006): 55–79.

replicate the work of others,[8] amount of effort to dedicate to a publication,[9] and what methodological choices to make.[10] Collectively these studies have shown that real scientists, motivated by things other than truth, will sometimes (but not always) behave in ways that step outside of the narrow path described by theories of epistemic rationality.

What should we make of this? A tempting solution – advocated by many – is that science does not progress in the way we thought. In contrast, some authors have argued that epistemically irrational scientists might best contribute to progressive science. David Hull states the position most emphatically, "Although objective knowledge through bias and commitment sounds as paradoxical as bombs for peace, I agree that the existence and ultimate rationality of science can be explained in terms of bias, jealousy, and irrationality."[11] Similar, but softer, positions have been articulated by a range of philosophers.[12]

Whatever one's opinion about Hull's claim, it is at least possible that epistemic irrationality could contribute to scientific progress. In the economic domain this was the core idea behind Adam Smith's invisible hand. Individual profit maximization, Smith argued, might lead to a society that is beneficial to its members. Scholars have either explicitly or implicitly appealed to this idea by arguing that science might progress through epistemically irrational actions of scientists.[13]

[8]Justin Bruner, "Policing Epistemic Communities" *Episteme* X, 4 (2013): 403–16; Felipe Romero, "Novelty vs Replicability : Virtues and Vices in the Reward System of Science," *Philosophy of Science* (forthcoming).

[9]Max Albert, "Product Quality in Scientific Competition," *Manuscript*; Liam Kofi Bright, "Decision Theoretic Model of the Productivity Gap," *Erkenntnis* LXXXIII, 2 (2017): 421–42.

[10]Hull, *Science as a Process*, *op. cit.*; Max Albert, "Methodology and Scientific Competition," *Episteme* VIII, 2 (2011): 165–83.

[11]Hull, *Science as a Process*, *op. cit.*, p. 32

[12]Karl Popper, "The Rationality of Scientific Revolutions,"in Rom Harre, ed., *Problems of Scientific Revolution: Progress and Obstacles to Progress* (Oxford: Clarendon Press, 1975); Kitcher, "The Division of Cognitive Labor," *op. cit.*; Kitcher, *The Advancement of Science*, *op. cit.*; Miriam Solomon, "Scientific Rationality and Human Reasoning," *Philosophy of Science* LIX, 3 (1992): 439–55; Solomon, *Social Empiricism*, *op. cit.*; Mayo-Wilson, Zollman, and Danks "The Independence Thesis : When Individual and Social Epistemology Diverge," *op. cit.*; Kevin J.S. Zollman, "Computer Simulation and Emergent Reliability in Science," *Journal of Artifical Societies and Social Simulation* XIV, 4 (2011): 15.

[13]Hull, *Science as a Process*, *op. cit.*; David Hull, "What's Wrong with Invisible-Hand

I do not mean to suggest that these views should be accepted uncritically. But the possibility that individual actions aimed at one goal might create a society that achieves a different end illustrates an important possibility. Treating scientific rationality as equivalent to individual epistemic rationality, and analyzing it as such, runs the risk that the rationality of science is divorced from scientific progress. Demonstrating that individual scientists are epistemically rational (or irrational) does nothing to establish that the community of scientists will effectively uncover truth.[14] Without a tight connection between these two, a theory of scientific rationality might become an exercise in hollow accolades.

One might avoid conflating scientific and epistemic rationality but argue that an approach for the instrumental rationality of science is already available. We have a well developed theory of instrumental rationality: expected utility maximization. For all its flaws, this theory – or a closely related one – appears to capture many of our intuitions about instrumentally rational behavior. A scientist is rational, one might say, if her response to evidence is epistemically rational and if she maximizes expected utility according to her own goals.

This solution is too weak, however. Expected utility theory is, by design, neutral with respect to an agent's goals. In many contexts this flexibility is a virtue. The theory's value-neutrality allows its incorporation into liberal political theories that allow citizens to pursue a life of their own design. But this feature is less appealing when we think about *scientific* rationality. A scientist who aims to minimize the amount of truth she produces is equally instrumentally rational as one who aims to maximize truth. Both can be captured as maximizing a utility function. While we might want to regard

Explanations?" *Philosophy of Science* LXIV (1997): 117–26; Thomas C. Leonard "Reflection on Rules in Science: An Invisible-Hand Perspective," *Journal of Economic MethodologyIX*, 2 (2002): 141–68; Michael Polanyi, "The Republic of Science: Its Political and Economic Theory" *Minerva* I (1962): 54–74; Gordon Tullock *The Organization of Inquiry* (Duke University Press, 1966); Kitcher, "The Division of Cognitive Labor," *op. cit.*; Kitcher, *The Advancement of Science*, *op. cit.*

[14]With Conor Mayo-Wilson and David Danks, I have called this claim the Independence Thesis: that social rationality is independent of individual rationality. With the context of a particular model of rational behavior we prove various parts of the independence thesis. See Mayo-Wilson, Zollman, and Danks "The Independence Thesis: When Individual and Social Epistemology Diverge," *op. cit.*; Conor Mayo-Wilson, Kevin J.S. Zollman, and David Danks "Wisdom of the Crowds vs. Groupthink: Learning in Groups and in Isolation," *International Journal of Game Theory,* XCII (2013): 695–723.

both scientists as equally *instrumentally* rational, we should not characterize them as equally *scientifically* rational. We should exclude some ends.

If we are to develop a theory of instrumental scientific rationality, we must then pick out those aims for science that contribute to a rational scientific enterprise. Truth seems like a laudable aim. Perhaps it is the only one. Adam Smith falsely claimed that great scientists were not motivated by the accolades of others. This single-minded focus on the truth, Smith argued, prevents them from developing "factions and cabals."[15] W.E.B. Du Bois states this position most clearly when he enjoins scientists to have "but one simple aim: the discovery of truth."[16] Call this position the "only-truth" position.

Real scientists deviate from the norm of pure truth-seeking. Most well known is scientists' desire to receive due credit for their work. A scientist doesn't merely want to advance knowledge, she wants everyone to know she pushed us forward. She wants to be known for her discoveries.

Boettke and O'Donnell advocate for a position antipodal to Smith and DuBois.[17] They argue that scientists should be motivated *solely* by credit and should ignore truth. "We argue that the only social responsibility of economists is to maximize their career advancement within the scientific community of economists… Rather than 'good' scientists, good rules of scientific engagement for ongoing contestation of ideas through open, critical discourse are required."[18] I will call this position the "only-credit" view.

Scientists need not have a single motive, they could have many goals. I will here consider a "hybrid" view which argues that the most rational scientists are motivated by truth and credit – they desire both. Where possible,

---

[15]Adam Smith, *The Theory of Moral Sentiments* Eleventh edition (London: Cadel & Davies, 1812): 213–15.

[16]W.E.B. Du Bois, "The Study of the Negro Problems," *Annal of the American Academy of Political and Social Science* XI: 1–23 at p. 16. It is important to note that the context of Du Bois' defense of this aim is different than our focus here, and so the arguments that follow are not aimed at refuting Du Bois' reasons even if they put some pressure on his position. For a more detailed discussion of Du Bois' philosophy of science, see Bright, "On Fraud" *op. cit.*; Liam Kofi Bright "Du Bois' Democratic Defence of the Value Free Ideal" *Synthese* (forthcoming).

[17]Peter J. Boettke and Kyle W. O'Donnell "The Social Responsibility of Economists" in George DeMartino and Deirdre McCloskey, eds., *The Oxford Handbook of Professional Economics Ethics* (New York, Oxford University Press, 2016), 116–36.

[18]ibid., abstract available online-only at `http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199766635.001.0001/oxfordhb-9780199766635-e-007`

they will try to maximize both the number of truths and the total amount of credit they receive.

The presence of these different positions open a question: what aims should we regard as rational aims for scientists? I will show two respects in which the previous positions are flawed. First, there is no unequivocal answer to this question – the best motivation depends on the social situation in which a scientist finds herself. As a result, those scholars who suggest that there is a single best aim for scientists, always and everywhere, are incorrect. Second, and contrary to existing discussions, I will show that the closest we can come to a general correct motivation is the hybrid view, where scientists are motivated by both credit and truth.

II. Forms of credit

Many existing discussions of the credit-motive in science conflate two different aspects of credit. We should consider the motives of individual scientists separately from the social norms of the community for how accolades are distributed. While these two features together determine the behavior of scientists, they are conceptually distinct and could be modified independently of one another.[19]

This paper focuses on two primary motivations of scientists: they desire (among other things) truth and credit. Scientists, we presume, are interested in better understanding some aspect of the world. In addition, they want to be recognized by their peers for exceptional work. Honest scientists desire acclaim for their own work; the more devious are happy to take credit for the discoveries of others. Regardless of their nobility, all else equal, scientists are happier when they receive more credit.

This motivation tells us little about how scientists will act, however, because it does not yet tell us how the community confers credit. The sociologist Robert Merton is duly famous for his investigation into this aspect of the credit system.[20]

The priority rule requires that the lion's share of credit goes to the person who discovered a phenomena first. A second-place discoverer gets less (or no) credit even if the discovery was made independently. *The Matthew*

[19]Michael Strevens, "Economic Approaches to Understanding Scientific Norms," *Episteme* VIII, 2 (2012): 184–200.

[20]Robert Merton, *The Sociology of Science: Theoretical and Empiricial Investigations* (Chicago: University of Chicago Press, 1973).

effect can be a countervailing influence.[21] When discoveries are made nearly simultaneously, or by teams of scientists working together, most of the credit is conferred on the more famous of the discoverers.

These two norms for the allocation of credit are independent of the motivation of the scientists. We could easily imagine alternative schemes for assigning credit that do not follow either of these rules.[22] There are constraints. For example, it is often difficult to observe the effort of scientists directly, we can only observe outcomes.[23] As a result, social norms that reward scientific effort directly are infeasible.

Different social norms for credit conferral may cause different scientific outcomes.[24] We must, therefore, take care to separate those features of scientific behavior that follow from the credit-motive alone from those that depend on the priority rule specifically.

In the following investigation, I take pains to divide scientist's motivations for credit from these social norms. Each section makes minimal assumptions about a credit system needed to secure a result. In this respect, our investigation is focused most squarely on the credit-motivation and only utilizes the norms of credit conferral where necessary. The conclusions will be very general and apply to many conceivable credit systems beyond those that incorporate the priority rule or Matthew effect.

Furthermore, my focus will be exclusively normative. Many interesting and complicated questions about the history of these norms remain unanswered. Several scholars have offered divergent theories of the emergence of the priority rule.[25] But, this is not my purpose here. I wish to uncover the

[21] ibid.; Michael Strevens, "The Role of the Matthew Effect in Science," *Studies in History and Philosophy of Science* XXXVII, 2 (2006): 159–70.

[22] Strevens, "The Role of the Priority Rule in Science" *op. cit.*

[23] Partha Dasgupta and Paul A. David, "Information Disclosure and the Economics of Science and Technology," in George R. Feiwel, ed. *Arrow and the Ascent of Modern Economic Theory* (New York, New York University Press, 1987), pp. 519–42.

[24] Kleinberg and Oren "Mechanisms for (Mis)allocating Scientific Credit" *op. cit.*

[25] Albert, "Product Quality in Scientific Competition" *op. cit.*; Paul A. David "Common Agency Contracting and the Emergence of 'Open Science'," *The American Economic Review* LXXXVIII, 2 (1998): 15–21; "Understanding the emergence of 'open science' institutions: Functionalist economics in historical context," *Industrial and Corporate Change XIII*, 4 (2004): 571–89; Strevens, "The Role of the Priority Rule in Science" *op. cit.*; "The Role of the Matthew Effect in Science, " *op. cit.*; Jesús P. Zamora Bonilla, "Scientific Inference And The Pursuit Of Fame: A Contractarian Approach," *Philosophy of Science* LXIX, 2 (2002): 300–23; "Cooperation, Competition, and the Contractarian View," *Ethics & Politics* XV, 2 (2013): 14–24.

effects, and by extension the normative standing, of the credit-motivation. I will leave the genealogy of these desires and norms for another time.

When considering scientists' motivations, there are many perspectives one might take. This paper will take a social and consequentialist focus. I will determine the normative standing of the truth, credit, and hybrid motives by looking at how those motives affect a community of scientists that adopt them.

As noted in the introduction, scientists make many decisions throughout their scientific careers. We are unable to analyze how the credit system affects all of them. Instead, I will focus on two central decisions – the choice of how much of one's labor to allocate to science and the choice of which project, among many, to pursue.

III. Scientific production

Our first decision: how much time should a scientist dedicate to her craft? The knowledge generated by scientific activity has all the hallmarks of a public good. One cannot easily prevent others from acquiring knowledge without paying – it is intrinsically non-excludable.[26] One person consuming knowledge does not prevent another from consuming it – knowledge is non-rivalrous.[27]

Markets are notoriously bad at incentivizing the creation of public goods, because it is difficult to recover the cost of production.[28] Dasgupta, David, and Stephan argue that one important consequence of the priority rule is to solve this public goods problem.[29] They point out that the priority rule con-

[26] By "intrinsically non-excludable" I mean that without government intervention it is non-excludable. Patents and copyrights are one method of converting non-excludable knowledge into intellectual property which is excludable.

[27] This observation goes back at least to Thomas Jefferson who said, "He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me" Thomas Jefferson "Letter to Isaac McPherson" reprinted in Philip B. Kurland and Ralph Lerner *The Founders Constitution vol. 3* (Indianpolis: Liberty Fund, 2001).

[28] This represents the central argument for the government to fund scientific research in Richard Nelson, "The Simple Economics of Basic Scientific Research," *Journal of Political Economy* LXVII, 3 (June 1959): 297–306.

[29] Partha Dasgupta and Paul A. David, "Toward a New Economics of Science," *Research Policy* XXIII, 5 (1993): 487–521; Paula E. Stephan, "The Economics of Science," *Journal of Economic Literature* XXXIV, 3 (1996): 1199–235.

verts the non-rivalrous and non-excludable good, knowledge, into a rivalrous and excludable good, priority. A scientist can prevent others from claiming priority and if she enjoys priority for a discovery, no one else can. In this way, they argue, the priority rule helps to mitigate the public good problem inherent in knowledge production.

In this section, I will argue that the credit-motive is more central to this issue than the priority rule norm. The priority rule is not unique among ways of apportioning credit. Any system that tracks (on average) the effort of the scientists will achieve this end. So, while the priority rule represents one way of solving this problem other credit norms would as well.

III.1. Professor Crusoe. Although our ultimate goal is to understand a scientific community, it will be helpful to start by considering a society with a lone scientist, Professor Crusoe. Crusoe has a simple choice, he must decide how much effort to dedicate to scientific activity and how much to leave to other things. Time is limited. He values scientific pursuits because he desires truth. Crusoe also enjoys acquiring food and engaging in leisure activities.

To model this, Crusoe will choose a percentage of his total time, $x \in [0, 1]$ to dedicate to science, leaving the remaining effort for other things. Displaying an academic's biases, I will henceforth call all non-science activities "leisure."[30] Crusoe has a focal scientific project where he will dedicate his scientific activity. This project may succeed at revealing some important fact about the world, or it might fail. Crusoe can affect the probability his project succeeds with more effort, but he cannot guarantee success.

Crusoe has a utility function over his available choices. Assume the function is additively separable, it takes the form $u(x) = s(x) + l(x)$. $s(x)$ represents joy from scientific discovery and $l(x)$ represents the joy of leisure activities.

These already impose constraints on Crusoe's preferences. First, his preferences must be amenable to representation by a utility function. Given the uncertainty of science, this means that Crusoe's preferences over uncertain prospects obey one of the standard decision theories which allow such a representation. Furthermore, we constrain his preferences to those that can be broken into two different parts which combine via addition. While this

[30]Leisure denotes anything that Crusoe does that benefits him alone without benefiting other scientists. This might even include scientific work on topics that only interests Crusoe, work that is kept secret, and work done for industry.

strikes me as a reasonable assumption, at least for illustrative purposes, it is not completely above reproach.

Crusoe's labor should increase the probability that his scientific project succeeds, and Crusoe wants his project to succeed. This entails that $s(x)$ is increasing with x. Crusoe also enjoys leisure, which entails that $l(x)$ is *decreasing* in x because as he dedicates time to science, he has less time for leisure.

Finally, both leisure and science feature decreasing marginal gains. In the case of leisure, this might because as he spends more time on leisure activities the less joy he takes in them – he gets bored. With science, he may exhaust the "low hanging fruit" – easy ways to further improve the probability that his scientific project succeeds. Each next bit of effort improves his scientific prospects ever less. This assumption about scientific effort was first made by Pierce:

> We thus see that when an investigation is commenced, after the initial expenses are once paid, at little cost we improve our knowledge and improvement then is especially valuable; but as the investigation goes on, additions to our knowledge cost more and more and at the same time are of less and less worth... All the sciences exhibit the same phenomenon; and so does the course of life. At first, we learn very easily, and the interest of experience is very great; but it becomes harder and harder and less and less worthwhile, until we are glad to sleep in death.[31]

In order to provide a concrete illustration, let $s(x) = \sqrt{x}$ and $l(x) = a\sqrt{1-x}$. The variable a represents the value of leisure relative to science. When $0 < a < 1$, we will say that science is more valuable than leisure. Conversely, when $a > 1$ leisure is more valuable. (These functions are used for illustration. I generalize the results in the appendix.)

Crusoe will maximize his utility function. This occurs when

$$x = \frac{1}{1+a^2}$$

[31] Charles Sanders Peirce "Note on the Theory of the Economy of Research" in *Report of the Superintendent of the United States Coast Survey Showing the Progress of the Work for the Fiscal Year Ending with June, 1876* (Washington, DC: Government Printing Office, 1879), pp. 197–201 at 198.

This equation captures the predictable consequence of our assumptions: as leisure becomes more valuable, Crusoe dedicates less effort to science.

Crusoe deserves no criticism for this. Crusoe maximizes scientific productivity for himself. No one else is affected; why should he do more? On the island, credit has no use. All this changes when the population of the island grows.

III.2. A scientific community. Now suppose a community of individuals, each of whom must allocate time to scientific activity. The scientists in this community embody Du Bois' ideal: they care nothing for credit. When it comes to science, all they desire is to produce truth. Like Professor Crusoe, they care about other things in their lives outside of science, however.

Let us begin by supposing a simple society with two scientists. One allocates x amount of time to science while the other chooses y. Both desire truth, so they gain from one another's contribution to science. But they do not enjoy the other's leisure – this is private.

The first scientist's utility is given by this equation:[32]

$$u_1(x, y) = \sqrt{x + y} + a\sqrt{1 - x}$$

The utility function for scientist 2 is the identical except with x and y swapped.[33] Effort in science is additive: both scientists are equally good at science and they don't have synergistic effects.

Like before, our scientists must choose how much time to allocate to science. Unlike with Crusoe, each scientist's choice is influenced by the behavior

[32]The astute reader will note that the first term, $\sqrt{x+y}$ is not a probability because it can take values above 1. We have therefore violated description of the model from above. This is not a problem, because one can modify a utility function by multiplying by a positive constant without any loss of generality. So, while I do not express it here for ease of calculations, one can assume that this utility function is multiplied by $\frac{1}{\sqrt{2}}$ in order to keep the first term below 1.

[33]In our operationalization of the only-truth conjecture we have assumed that scientists care about truth in itself. This assumption makes our scientists maximally ideal – they don't care who produces the truth, only that truth is produced. Kitcher calls these scientists "altruistic," but I will not use this term here. In particular, my scientists are not altruistic toward one another – neither scientist takes pleasure in the leisure of the other. I believe this utility function represents what most proponents of the only-truth position have in mind. Kitcher's target was slightly different, and as a result he made different assumptions. The contrasts with Kitcher will be discussed in more detail in section 4. See Kitcher, *The Advancement of Science*, *op. cit.* at p. 311, footnote 7.

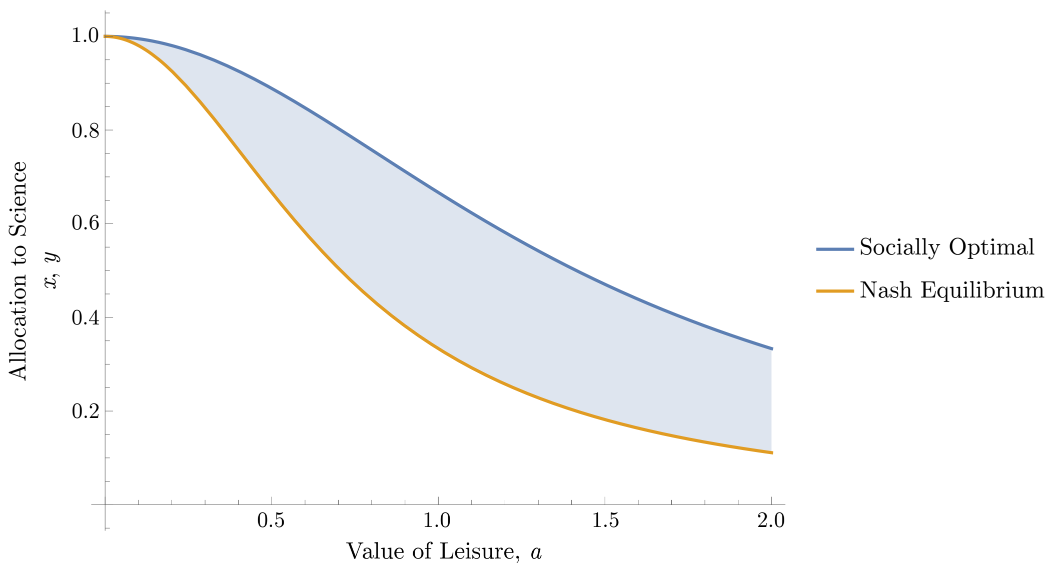

Figure 1: An illustration of the gap between the socially optimal amount of scientific production and the Nash equilibrium relative to the value of leisure, a.

of the other scientist. For example, the first scientist maximizes her utility by setting:

$$x = \frac{1 - ya^2}{1 + a^2}$$

(The equation for scientist 2 is the same with x and y reversed.) If scientist 1 allocates a significant amount of time to science, then scientist 2 will be happy for 1's effort and allocate more time to her own leisure.

To determine how rational scientists will behave, we find an allocation that is simultaneously optimal for both scientists – the Nash equilibrium of our scientific game. Consider the allocation where both scientist dedicate the same amount of effort. This is known as the symmetric Nash equilibrium and occurs when:

$$x = y = \frac{1}{1 + 2a^2}$$

For Crusoe, we were in no position to criticize his decision. After all, he was alone on the island – he owed nothing to anyone else. But now, there are two people. Even though they are both doing some science, they both might be unhappy with their situation.

In fact, they are. *Both* scientists would prefer that they instead choose a higher level of effort in science. Figure 1 illustrates the gap between the individually rational allocation (the Nash equilibrium) and the socially optimal allocation.

This problem arises because in this model a public good – science – is produced by paying a private cost – the loss of leisure. Both scientists would prefer that they both dedicate more to science. But the situation where they do more is unstable because each would have an incentive to cheat. Given the choice, each scientist will allocate a little less to science. When both cheat, they are both made worse off.[34] Situations of this form are familiar from social dilemmas like the Prisoner's dilemma. The truth-only position is susceptible to the public goods problem. If scientists only find value in truths produced by science, then science will be under-produced.

The notion of social optimality I use considers only on the two scientists in this model. If our island were inhabited with other non-scientist citizens who also desired to know the truth, but could not produce it, the situation would be even worse. The non-scientists would desire more science but would not gain from the scientists' leisure.

One might criticize our simple model for a number of reasons: there are only two scientist, they are both identical in their ability to do science, and they both care about leisure to the same degree. The general conclusion I reached does not critically depend on these assumptions. In the appendix this result is presented with a high degree of generality. The underlying public goods problem remains with more scientists, with scientists who differ in terms of ability and their desire for leisure, with different scientific production functions, and with many leisure functions.

In these models we have taken the scientists preferences as fixed and exogenous to our models. This need not the be case however. It is well understood that our current process of education, hiring, and tenure serves to drive out those academics who have too strong a preference for leisure. While deserving of study in its own right, this fact does not change the central result of this section. Even if our scientists care little for leisure, they would face a social dilemma. Only if our scientist were completely bored by all leisure activities would the dilemma vanish.

These models treat the population of scientists as stable and unchanging.

[34] Formally speaking, in this context the socially optimal state Pareto dominates the Nash equilibrium.

Of course, new people join the scientific community while others depart. If the new scientists had exactly the same abilities and preferences for leisure as those who depart, the model would not change at all (scientists with identical ability and preferences are interchangeable). If the younger generation had structurally different preferences, the location of the equilibrium would shift over time, but the social dilemma would remain. So long as the population settled near the Nash equilibrium, the scientists would all regret their situation.

But, should we expect the Nash equilibrium to predict how a population will behave? The Nash equilibrium concept, as a solution to games, is often criticized. It is said that the use of the Nash equilibrium presumes a significant amount of knowledge on the part of the various participants. While the most common way of motivating Nash equilibria involves substantial knowledge and foresight, research into evolutionary game theory has shown that lower rationality learning often – but not always – results in Nash equilibrium play.[35]

If individuals are maximizing relative to their beliefs but are systematically wrong, one might find a different outcome. Perhaps scientists have incorrect beliefs about the choices of others or perhaps they have false beliefs about the effect of their scientific labor.[36] Our modeling assumptions may not be robust to alterations of this form, but space prevents a full exploration of this here. Moving beyond the question of whether we should expect people to play the Nash equilibrium, relying on individuals' ignorance to solve a problem does not represent effective social design.

III.3. Adding credit. Our scientific community confronts a public goods problem. Dasgupta, David, and Stephan argue that the priority rule solves this difficulty. While I will show they are correct, their argument is overly narrow. The motivation for credit, I argue, provides a solution to the public goods problem, and this solution would persist under a large number of different credit allocation norms.

To add credit to our model scientific community, we must create a function to represent how much scientific credit, on average, is distributed to a

[35]For a systematic presentation see William H. Sandholm *Population Games and Evolutionary Dynamics* (Cambridge: MIT Press, 2010).

[36]See Ryan Muldoon and Michael Weisberg, "Robustness and Idealization in Models of Cognitive Labor," *Synthese* CLXXXIII, 2 (2010): 161–74.

scientist given her and her colleagues' effort. Recall, when scientists 1 and 2 dedicate x and y to science respectively, the probability of success is proportional to $\sqrt{x+y}$. We might assume a credit system works like this: if the project fails, no credit is given. If the project succeeds each scientists receives credit relative to their contribution to the success of the project. This might take many forms, for illustration assume that credit is divided in proportion to their *marginal* contribution – the increase in the probability of success attributable to each scientist's effort.

Scientist 1's utility function will now be $u_1(x, y) = \sqrt{x+y} + a\sqrt{1-x} + c_1(x, y)$. The credit part of scientist 1's utility function is:

$$c_1(x, y) = c\sqrt{x+y}(\sqrt{x+y} - \sqrt{y})$$

Where c represents how much scientist 1 enjoys credit. (Scientist 2 has the same function with x and y transposed.)[37]

Each scientist is assigned credit based on how much she unilaterally adds to the success of the project (that is the increase in success probability from her contribution, holding fixed the contribution of the other). This is not the only way to assign credit. An alternative credit function might divide the total probability of success between the two scientists in proportion to their effort. In the context of this problem, this does not alter the conclusions. However, in the subsequent section these decisions will become important.

We assume that the scientists care about credit intrinsically – that is their desire for credit is not founded in other things that credit produces. Scientists might desire credit because it affords them greater opportunities to travel, more income, or some other tangible asset. This would complicate the model, because credit could be transformed into leisure at a later date.[38]

Each scientist is maximizing the sum of their desire for truth, their desire for leisure, and their desire for credit. Intuitively, the desire for credit should increase scientists' contributions to science since credit represents a second benefit from scientific activity. By engaging in science the scientists gain knowledge – which they continue to desire – and they also gain a chance at credit.

The expression for the Nash equilibrium for our truth and credit seeking

[37]The same caveat about multiplicative constants from footnote 32 should be applied here.

[38]Some exploration of this idea has been done by David M. Levy, "The Market for Fame and Fortune," *History of Political Economy* XX, 4 (1988), 615–25.

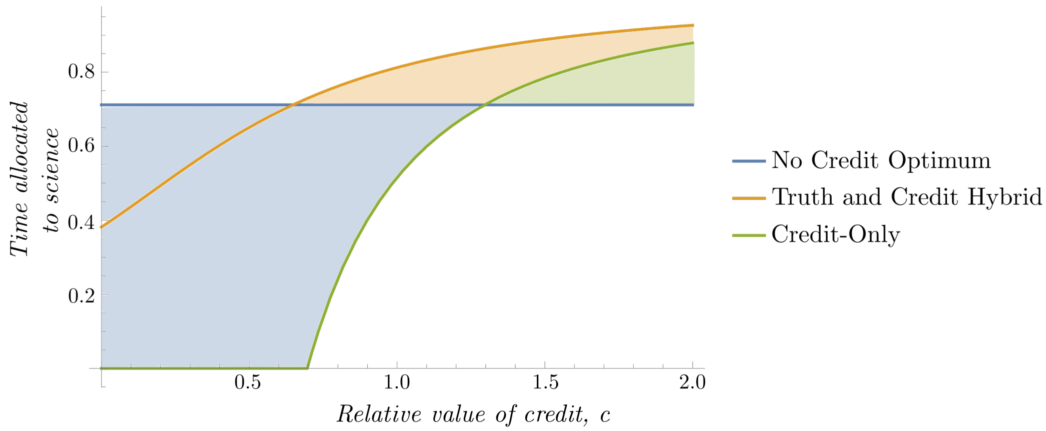

Figure 2: An illustration of the Nash equilibria for a two-scientist economy relative to the value of credit, c.

scientists is more complicated and not particularly illustrative.[39] Figure 2 illustrates the underlying result. As credit becomes more valuable, scientific productivity increases. Once the scientists desire truth to a sufficient degree, they produce science at a rate equal to (or greater than) the rate desired by the truth-only motivated scientists.[40]

Figure 2 also illustrates how scientists would allocate their effort if they cared solely for credit. Unsurprisingly, if scientists did not care about truth, they would spend less time on science than if they cared about both truth and credit.[41]

This provides a strong argument against the only-truth position. Science and scientists are benefited when scientists desire both credit and truth. Scientific productivity increases as the desire for credit (c) grows. The more scientists seek credit, the better.[42]

---

[39] All relevant equations will be provided in a supplementary Mathematica file.

[40] These results are not unique to this particular way of allocating credit. The appendix illustrates that many credit allocation norms function in the same way.

[41] Whether or not the truth-only or credit-only scientists allocate more to science will depend on the value of leisure, a, and credit, c. For some values the credit motivated scientist will allocate more to science, for other values the truth-motivated ones will allocate more.

[42] In passing, Congleton notes that status-seeking behavior can help to serve the public good when those seeking status – in this case scientists – produce something that is beneficial to others. Although the underlying approach is quite different, this is a similar observation to the one made here. See Roger Congleton, "Efficient status seeking:

This model provides a slightly weaker argument against the only-credit position. Unless the scientists desire for credit is very strong, credit will provide an insufficient motivation to produce science. The truth-motivation provides a helpful, but not necessary, supplement to the credit-motivation.

All of these statements are true within the model just presented, but real science is more complicated. Scientists are usually not restricted to a single project. Will these same conclusions hold true if the model becomes more realistic?

IV. Choosing projects

Perice was the first to recognize the importance of allocating one's labor between several potential research projects.[43] He also saw that this question has a distinctively economic dimension because it requires the comparison of various costs and benefits.

In many scientific domains, one can approach a problem in may ways: there may be multiple ways to synthesize a molecule, to discover the structure of a hidden part of nature, or to develop a theory to explain some phenomena. Furthermore, because science has a high degree of uncertainty, one will not know ahead of time which approach will ultimately succeed. In such situations it maybe optimal to divide labor between both projects to maximize the probability that at least one of the projects succeeds.[44]

Polyani asserts that the truth-motive of scientists achieves this optimal division of cognitive labor.[45] He therefore thinks that truth-motivated scientists are in need of no external guidance about what projects to pursue. Regarding the credit-motive, Kitcher and Strevens argue that the priority rule achieves a near-optimal allocation of scientists between different research programs.[46]

Externalities, and the evolution of status games," *Journal of Economic Behavior and Organization* XI, 2 (1989), 175–90 at p. 185.

[43] Peirce "Note on the Theory of the Economy of Research" *op. cit.*

[44] Dasgupta "The Economics of Parallel Research" *op. cit.*; Kitcher, "The Division of Cognitive Labor," *op. cit.*

[45] Polanyi "The Republic of Science" *op. cit.*

[46] This is in contrast to an argument from Dasgupta and Maskin. They suggest that the priority rule leads to waste because it encourages two scientists to correlate their research programs successes or failures. This possibility is not represented in Kitcher's and Strevens' models, and it is not addressed in the model presented in this paper. See Partha Dasgupta and Eric Maskin, "The Simple Economics of Research Portfolios," *The Economic Journal*

If Kitcher's and Streven's claims are true, it would undermine the only-truth position further. Not only would the desire for credit serve to motivate scientists to do more science, it would also motivate them to distribute their labor within science in a beneficial way. This section will show, however, that Kitcher and Strevens cannot be marshaled against the only-truthers in this way – the truth-motive optimally distributes labor without supplement. So as a result, credit can *at best* have no effect. Some systems of credit, like the one used in the last section, are instead counterproductive.

IV.1. Professor Crusoe, redux. Let us return to Professor Crusoe. For the moment suppose he is no longer contemplating how much effort to invest in science – he has made this decision already. He now must chose what fraction of his scientific work to dedicate to one of two projects, s and t. Let x now represent the proportion of his effort dedicated to project s.

I will again assume separable additivity, which here is a much more significant assumption. Many scientific problems with multiple approaches might feature a synergy whereby contribution to one project helps another.[47] We will set these aside and assume Crusoe maximizes the following utility function $u(x) = s(x) + t(x)$. s is increasing and t is decreasing in x and both feature decreasing marginal returns.

This section will use the same illustrative model from section III. There are now two scientific approaches to a single focal problem. Crusoe would like to solve the problem but does not know which approach will ultimately succeed. The scientist's work on an approach increases the probability of success. Crusoe wants to solve the problem, but does not care how that end is achieved.

Following the model of the last section, suppose $s(x) = \sqrt{x}$ and $t(x) = b\sqrt{1-x}$. Parallel to a in the context of leisure, b can be said to represent the relative quality of project t.[48] In this model, Crusoe can divide his labor

XCVII, 387 (1987): 581–95; Kitcher, "The Division of Cognitive Labor," *op. cit.*; Kitcher, *The Advancement of Science*, *op. cit.*; Strevens, "The Role of the Priority Rule in Science" *op. cit.*

[47] Strevens, "The Role of the Priority Rule in Science" *op. cit.*

[48] Again we have the same slight abuse of mathematics noted in footnote 32. For many values of b, the total probability of success will be above 1, and thus not really a probability. As before, the reader is asked to assume that the utility function is multiplied by a constant to reduce the highest value to less than 1. We are making a further assumption that the projects are mutually exclusive – the probability that at least one of them succeeds is

between two projects without penalty. In some real contexts, this division of attention might be difficult. We are setting these concerns to the side for the purposes of philosophical exploration.

The mathematics for Crusoe is identical to the previous model. We have just reinterpreted leisure as a second scientific project. As before, we cannot criticize our lone scientist. Crusoe allocates his labor optimally because he is the only person who stands to benefit. The parallels disappear when we turn to the community of scientists.

IV.2. Another scientific community. Now there is a second scientist who also must allocate her labor between s and t. The utility for scientist 1 changes because scientist 1 benefits from scientist 2's work on both projects s and t.

$$u_1(x, y) = \sqrt{x + y} + b\sqrt{2 - x - y}$$

This minor change also affects the symmetric Nash equilibrium, which is now,

$$x = \frac{1}{1 + b^2}$$

More importantly, the optimal allocation is identical to the Nash equilibrium. That is, the scientists are individually choosing to allocate their labor in the best way possible. In the appendix we show that this is general result that does not depend on the particular functions $s(x)$ and $t(x)$ chosen here for illustration.

As was the case with the choice between a single project and leisure, this qualitative result would remain even if the population of scientists changed over time. As new scientists entered or left the population, the location of the equilibrium would change, but it would also coincide with the socially optimal state.

How does this result square with the claims of Kitcher and Strevens?[49] Let us start with the latter. There is not significant disagreement with

given by the sum of the probability that each succeeds on its own. Our utility function is also a close approximation when they are not mutually exclusive but the probability that either project succeeds is very small.

[49]One superficial difference is that Kitcher and Strevens restrict scientists to allocate all of their labor to one project rather than allowing them, as I do, to split their efforts. While this will make a different to what counts as the optimal distribution in a given situation, it seems unlikely to alter the relative quality of the only-truth, only-credit, and hybrid position.

Strevens. Strevens' project is to compare different systems of credit allocation in order to determine which, among many, credit systems would be superior. Strevens does not explicitly consider the motivation for truth. So our model does not contradict his, since he does address the truth-motive directly.

Kitcher on the other hand, takes himself to be addressing a closely related issue. In his earlier paper, he considers three ways for a truth-motivated scientist to choose between two scientific projects.[50] The first and second strategy involve the scientist choosing the "intrinsically" better project.[51] In our illustrative model a scientist would allocate all her effort to project s (i.e. choose $x = 1$) if $b < 1$ and allocate all her effort to project t (i.e. choose $x = 0$) if $b > 1$.

Why should scientists choose such a bad approximation of the rational strategy? One possibility, Kitcher argues, is that scientists may be ignorant of the allocations of others. This might lead to the simplistic choice strategies he describes. This possibility has been explored in some detail by Muldoon and Weisberg.[52] Even in the case of partial ignorance, Muldoon and Wiesberg's results question the degree to which credit-motivated scientists would be superior to truth-motivated ones.

Kitcher's third method from his earlier paper is the method we describe here – a scientist allocates her time where it would maximize the total probability of success.[53] Later, in *The Advancement of Science*, Kitcher recognizes that truth-motivated scientists will lead to an optimal distribution, but he argues that no contemporary scientist is motivated solely by truth.[54] Since our current question is a normative one – should scientists be so motivated – this objection is not relevant here.

IV.3. Adding credit. Even though credit could not improve the situation, we should investigate what would happen if scientists seek credit. This motivation cannot be productive – the scientists are already performing optimally without it. So, at best, the credit-motive has no effect. Otherwise the desire

[50] Kitcher, "The Division of Cognitive Labor," *op. cit.* at pp. 13–14.

[51] For the interested reader the first strategy is described at the bottom of page 13 and the second strategy is that numbered "(1)" on the top of page 14.

[52] Muldoon and Weisberg "Robustness and idealization in models of cognitive labor," *op. cit.*

[53] Kitcher, "The Division of Cognitive Labor," *op. cit.*

[54] Kitcher, *The Advancement of Science*, *op. cit.*, p. 344, footnote 26.

for credit will disturb a well functioning scientific community.

Assume a credit system identical to the one in section III. Now there are two scientific projects, s and t. If project s succeeds, then the scientists are rewarded proportional to their marginal contribution to project s. Similarly for t; if this project succeeds, each scientist receives some credit.[55]

Formally the credit function for project s is:

$$c_{s,1}(x) = c\left(\sqrt{x+y}\left(\sqrt{x+y} - \sqrt{y}\right)\right)$$

And the credit function for project t is:

$$c_{t,1}(x) = c\left(b\sqrt{2-x-y}\left(b\sqrt{2-x-y} - b\sqrt{1-y}\right)\right)$$

We can again solve for the Nash equilibrium where scientist care both about truth and credit. A general statement of the solution for all values of the variables is unwieldy. But we can graph the solution for particular values to get a sense of the effects of this credit system.

I will focus on the case where $c = 1$, where the scientists care as much for credit as they do for truth. Figure 3 illustrates the problem with this credit rule, it encourages an overallocation to the superior project. Strevens found the same effect in his model.[56]

The credit-only view fares worse. Scientists motivated solely by our credit function – who care nothing for truth – will almost always allocate all of their labor to one project. The only exception occurs when both projects have identical probability of being successful, when $b = 1$. (This result is proven in the appendix.)

This is a feature of the particular credit function we chose. Our credit function only gives credit for effort if the project is successful. If the project fails, scientists receive no credit for their effort. This choice is reasonable as it tracks how actual credit works, effort that produces no scientific results rarely receives acclaim.

[55]Following Kitcher and Strevens, I assume that there is a single goal that either project could achieve and that the credit conferred is the same if project s succeeds as it is if t succeeds. Kleinberg and Oren consider a model where the amount of credit given depends on the project and where scientists are solely motivated by credit. They show that scientists will allocate their labor optimally for some ways of conferring credit, but not for others. See Kleinberg and Oren "Mechanisms for (Mis) allocating Scientific Credit" *op. cit.*

[56]Strevens, "The Role of the Priority Rule in Science" *op. cit.*

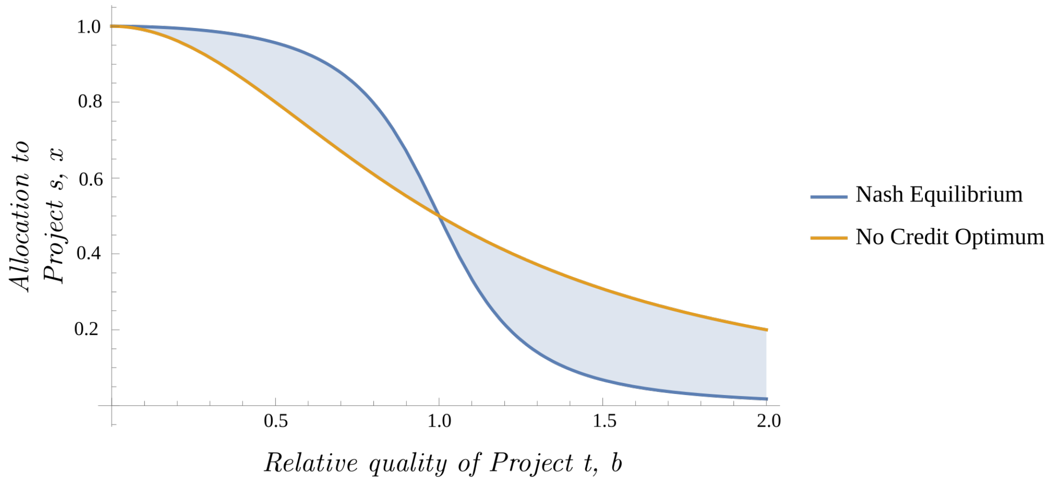

Figure 3: An illustration of the Nash equilibria for a two-scientist, two-project economy varying the relative quality of the two projects.

If we had not made this assumption – if scientists were rewarded for their marginal contribution to each project regardless of the project's ultimate success – we would have a different result. With this credit norm, scientists desire for credit won't affect the final allocation. In this case, the truth-only, credit-only, and hybrid views all result in the same outcome.

In contrast to section III, this model gives a weak defense of the truth-only position. When the public goods problem is set aside, optimal scientific communities can be comprised by scientists motivated solely by truth. Those who are motivated (partially or entirely) by credit might opt to over-allocate their labor to safer alternatives. Whether this occurs will depend on the social norms governing the distribution of credit.

V. Two projects with leisure

The preceding two models pull in two directions with respect to the truth-motivation of scientists. Focusing on the public goods problem, the credit-motivation appears productive. When we look at the issue of allocation of effort to multiple project we see that the credit-motivation would have at best no effect and at worst lead to a systematic misallocation of scientific work. Real scientists face both choices. At home on a Saturday afternoon, one must choose whether to work and one's topic for the day. Can we make

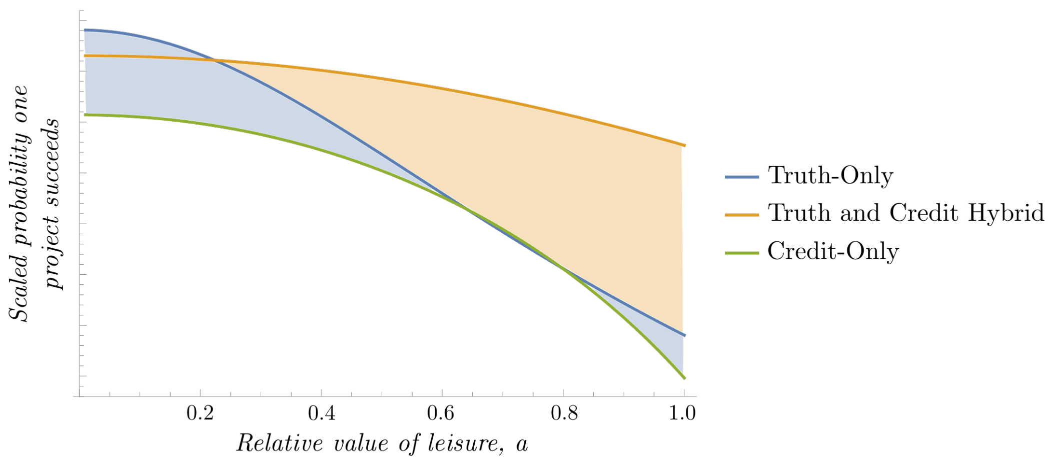

Figure 4: A comparison of the scientific output of three scientific groups. In all cases, a community of two scientists is choosing between two scientific projects and leisure. The y-axis represents the total scientific output, interpreted as a rescaled probability. In this diagram $b = 0.5$ and $c = 1$ (for the hybrid and credit-only curves).

any general pronouncements about the value of credit in this setting?

Utilizing the illustrative models from the previous two sections, we can. Scientists will make a three way choice between leisure, project s, and project t. One result is pictured in figure 4. The y-axis of this figure represents a rescaled probability that one of the projects succeeds.

When a is low – when the value of leisure is much less than the value of science – the credit-motivation has negative effects. The credit-motivation increases the total amount of labor allocated to science, but this motive also systematically misallocates labor between the two projects. When leisure is not particularly enjoyable the negative effects of the latter swamp the former. In such cases, the truth-only position wins out.

As a grows, however, the relative costs and benefits reverse. Now as leisure becomes more valuable, scientific productivity is benefited by a hybrid credit- and truth-motive. While credit still leads to a misallocation of labor between the two projects, this harm is swamped by the gain from increased effort. In these cases, the credit-only position continues to do poorly. Scientists motivated solely by credit fail to produce enough to overcome the radical misallocation of effort between two scientific projects.

These results are for the credit system where only the successful project receives credit. Had we used a system of credit allocation where even the unsuccessful project was awarded credit, then the hybrid view would outperform the truth-only position no matter what is the value of leisure. Similarly the hybrid position would outperform the credit-only position, but only because the desire for truth gives an extra incentive to do more science.

VI. Conclusion

While we have not provided an in-all-cases argument against the truth-only position, defenders of this view find themselves on difficult terrain. Relative to these models, the truth-only view results in a well functioning scientific society only when scientists have little to distract them from science.

While perhaps independently wealthy scientists of a past era might have eschewed other activities, today's scientists face complex labor trade-offs with family, teaching, private consulting, and true leisure. The complexity of modern life, the requirements of teaching and university administration, and the demands of work-life balance for professionalized scientists have created a situation where non-scientific endeavors occupy more of our time. In such a case, the credit-motive appears to have a productive influence on science.

Within the confines of these models, we find strong arguments against the credit-only view advocated by Boettke and O'Donnell.[57] In these models, communities of scientists motivated by credit were always inferior to those also motivated by truth. The hybrid view wins out over the credit-only view in these settings.

Of course these models are limited in two critical ways. First, they idealizing assumptions about the structure of the scientific community and about individual scientists. Real scientists and scientific communities are more complicated than our models, and it is always possible that a critical causal factor has been left out. In each section, I have argued why I do not think the models have critical lacunae, but this remains a possibility.

Beyond the idealization of the models, this paper only addresses two types of decisions that scientists might face: how much time to dedicate two one of two scientific projects or to non-scientific endeavors. There are a wide array of other decisions confronted by scientists. As noted in section I, many scholars have investigated how the credit-motive shapes a number of scientific

[57]Boettke and O'Donnell "The Social Responsibility of Economists," *op. cit.*

decisions. Less is known about how scientists would behave under the truth-only or hybrid motivations (see Bright's work for a notable exception).[58]

Significant work remains. Scientists make many choices over their careers. Different sciences confront inquirers with varying trade-offs. The credit norms vary across scientific cultures. But, if we ever hope to develop a robust sense of scientific rationality this multifaceted project must be undertaken.

VII. Appendix: Formal Results

In this appendix, I generalize some of the result from earlier sections to more general scientific production, leisure, and credit functions. In addition, a few claims made in the text are proven.

VII.1. Science and leisure.

VII.1.1. Without credit. To begin, Prof. Crusoe chooses between science and leisure. Crusoe's utility function is given by $u(x) = s(x) + l(x)$, where s is interpreted as science and l as leisure. Science and leisure are both enjoyed by Crusoe, so $s(x)$ is increasing in x and $l(x)$ is decreasing in x. For mathematical convenience assume both functions are twice differentiable. Finally, both science and leisure feature decreasing marginal utility for him, so $s''(x) < 0$ and $l''(x) < 0$.

Crusoe will choose x in order to maximize his utility function. The maximum of the utility function occurs where $s'(x) \leq -l'(x)$ and if $x > 0$, $s'(x) = -l'(x)$. That is, he chooses x so that the rate of gains from scientific production equal the rate of loss from being unable to do other things.

Now consider a community of $n \geq 2$ scientists. Each scientist i chooses an amount of effort to dedicated to science x_i. This produces a vector of labor choices, denoted by $\mathbf{x} = \langle x_1, x_2, \ldots, x_n \rangle$. Now each scientist's utility function has the form:

$$u_i(\mathbf{x}) = s(X) + l_i(x_i)$$

I will use X to represent the total weighted contributions to science, $X = \sum_i w_i x_i$. The weight terms w_i are used to represent differential ability levels for different scientists. Higher w_i corresponds to i's labor being more efficiently converted into truths. I will normalize $w_i \geq 1$.

[58] Bright, "On Fraud" *op. cit.*

Since s represents the truths gained from science, the s function is the same for every scientist. There are decreasing marginal gains from engaging in more science and increasing marginal losses from losing additional time for other things. But, scientists might value leisure differently, so we do not assume identical l_i functions.

What would a social planner do? We assume the social planner can force each individual scientist's hand. The social planner desires to maximize the total utility in the society.[59] The social planner wants to choose an allocation, $\mathbf{x}^\circ$ for each scientist that maximizes the total aggregate scientific output minus the total cost from lost time doing other things. Mathematically, the social planner wants to maximize

$$\sum_i s(X) + l_i(x_i)$$

This occurs when for every i,

$$\frac{\partial}{\partial x_i} \sum_j s(X) + l_j(x_j) \leq 0 \tag{1}$$

In words, the social planner will allocate an amount of effort to each scientist such that the marginal costs of doing more science imposed on that scientist are equal to the marginal total social good generated by that scientist for all members of the community.

The Nash equilibrium is a set of allocations, $\mathbf{x}^*$ which satisfies the following set of constraints,

$$\frac{d}{dx_i} s(w_i x_i + \sum_{j \neq i} w_j x_j^*) + l_i(x_i) \leq 0 \tag{2}$$

Here each individual chooses an x_i where the marginal gains *to her* for participating in more science equal the marginal loses from losing out on other things.

Under these, relatively weak assumptions, our community faces a social dilemma. The total amount allocated to science will be less in the Nash equilibrium than would be regarded as optimal by the social planner.

[59] This is not to say there is, or should be, any such person. It merely represents a heuristic tool for finding the *socially optimal state*. To engage in these calculations we must assume that we have some method for the interpersonal comparison of utility so that the sum of utilities is meaningful.

Proposition. *Whenever the socially optimal state involves some scientific effort, total scientific labor in the socially optimal state is higher than in the Nash equilibrium. That is, if $X^\circ > 0$, then $X^\circ > X^*$*

Proof. There are two relevant cases.

Case 1: For all i such that $x_i^* > 0$, $x_i^\circ > x_i^*$ – that is, every agent who contributes to science in the Nash equilibrium contributes more in the socially optimal state. In this case the theorem is obviously true because for all j, $x_j^\circ \geq 0$.

Case 2: There exists an i such that $x_i^* > 0$ and $x_i^* \geq x_i^\circ$. Because l_i' is decreasing, this entails that $l_i'(x_i^\circ) \geq l_i'(x_i^*)$. Letting $X_{-i} = \sum_{j \neq i} w_j x_j$, equation 2 is equivalent to

$$w_i s'(w_i x_i + X_{-i}^*) \leq -l_i'(x_i^*)$$

Since $x_i^* > 0$, then $w_i s'(X^*) = -l_i'(x_i^*)$.

Similarly, equation 1 is equivalent to

$$n w_i s'(w_i x_i + X_{-i}^\circ) \leq -l_i'(x_i^\circ)$$

The preceding two equations entail that $w_i s'(X^*) \leq w_i n s'(X^\circ)$, which also entails $s'(X^*) > s'(X^\circ)$. Because s' is decreasing this entails that $X^\circ > X^*$. □

A usual characteristic of social dilemmas is that the socially optimal solution is regarded as superior to the Nash equilibrium by every member of the society. In more technical terminology, the socially optimal solution is Pareto superior to the Nash equilibrium.

Because of the asymmetries present in this model, the social optimum need not Pareto dominate the Nash equilibrium. If we assume all scientists are identical in their abilities and desire for leisure, then the socially optimal allocation is better for every scientist than the symmetric Nash equilibrium allocation. (These are merely sufficient, but not necessary conditions for the socially optimal allocation to be Pareto superior to the Nash equilibrium.)

Proposition. *If $l_i = l_j$ and $w_i = w_j$ for all i and j and $X^\circ > 0$ then assuming that x^* is a symmetric Nash equilibrium, for all i, $u_i(x^\circ) > u_i(x^*)$*

Proof. Because $w_i = w_j$ we can assume without loss of generality that $w_i = 1$. The symmetry conditions imply that the the socially optimal allocation must involve the same effort level for every scientists, $x_i^\circ = x_j^\circ$. By the definition of the socially optimal state, it is the case that:

$$\sum_i s_i(X^\circ) + l_i(x_i^\circ) \geq \sum_i s_i(X^*) + l_i(x_i^*)$$

Because of the symmetries in the system, this is equivalent to (for arbitrary i):

$$s_i(X^\circ) + l_i(x_i^\circ) \geq s_i(X^*) + l_i(x_i^*)$$

By the previous proposition, $\mathbf{x}^\circ \neq \mathbf{x}^*$. Because of the assumptions about the structure of s_i and l_i, this entails the inequality is strict. □

VII.1.2. With credit. Assume the same heterogeneous community of individuals as before, but now in addition to the direct benefits from the generation of truth, the individuals benefit by receiving credit. I will again assume an additively separable utility function:

$$u_i(\mathbf{x}) = s(X) + l_i(x_i) + c_i(x_i, x_{-i})$$

$c_i(x_i, x_{-i})$ represents the expected credit for individual i receives from allocating effort x_i when the other's choose effort according to the vector x_{-i}. Credit is always preferred by the scientists ($c_i \geq 0$), and the expected utility of credit is increasing in the first argument. We make no assumptions about how credit is sensitive to other arguments – how scientist i is effected by an increase or decrease in scientist j's effort. The relationship might plausibly go either way. On the one hand, other scientists working on an approach might increase its probability of success, thus increasing the likelihood that i gets credit. On the other hand, other's effort might crowd out i and reduce her expected credit. Finally, c_i features decreasing marginal gains in x_i and is twice differentiable.

In a credit economy the Nash equilibrium allocation $\mathbf{x}^C$ satisfies these constraints:

$$w_i s'(w_i x_i^C + X_{-i}^C) + l_i'(x_i^C) + c_i'(x_i^C, x_{-i}^C) \leq 0$$

Proposition. *Assuming that there is some scientific production under credit, $X^C > 0$, the total amount of science is higher in the Nash equilibrium under the credit system than without credit. That is $X^C > X^*$*

Proof. A similar proof strategy works here as above. Suppose an agent i such that $x_i^* > x_i^C$ and $x_i^* > 0$. Because l' is decreasing, this entails that $l'(x_i^C) \geq l'(x_i^*)$. Since $x_i^* > 0$, then $w_i s'(X^*) = -l_i(x_i^*)$ and $w_i s'(X^C) \leq -l_i(x_i^C) - c_i(x_i^C, x_{-i}^C)$. Because $c_i' > 0$, this entails that $w_i s'(X^*) > w_i s'(X^C)$ which entails that $X^C > X^*$. □

VII.2. Two projects. Now we consider a problem where scientists are choosing between two scientific projects.

VII.2.1. Without credit. There is a community of scientists each of whom must allocate effort to one of two projects s and t. I will here use an even more general utility function than before. Instead of assuming that the total scientific productivity is a function of a weighted sum of effort, we assume it is any function whatsoever of the vector of effort.

Each individual scientist acts to maximize the following utility function,[60]

$$u_i(\mathbf{x}) = s(x_i, x_{-i}) + t(1 - x_i, 1 - x_{-i})$$

The socially optimal state occurs where (when $1 > x_i > 0$):

$$\frac{\partial}{\partial x_i} ns(x_j, x_{-j}) = -\frac{\partial}{\partial x_i} nt(1 - x_j, 1 - x_{-j})$$

If we now turn our focus to individual allocations, the conditions on Nash equilibria, $\mathbf{x}^*$ are (again when $1 > x_i > 0$)

$$\frac{d}{dx_i} s_i(x_i, x_{-i}^*) = -\frac{d}{dx_i} t_i(1 - x_i, 1 - x_{-i}^*)$$

which is equivalent to the socially optimal state for all functions.

Because the Nash equilibrium is always socially optimal in this setting, the addition of credit cannot – by definition – improve the situation. Some credit functions may have no effect on the Nash allocation, and as a result the credit-motive does not harm the community. But, other credit functions will lead to a misallocation.

[60] I will abuse notation slightly and write $1 - x_{-i}$ to represent the vector made by subtracting each element in x_{-i} from 1.

VII.2.2. Credit only. Suppose now two scientists who are only motivated by credit and credit is allocated according to the conditional marginal credit rule discussed above. That is, a scientist receives credit for her work on a project only if that project succeeds. If the project succeeds, then the amount of credit she receives is proportional to her marginal contribution.

Formally, scientist 1's utility function is:

$$\begin{aligned} u_1(x,y) =& \sqrt{x+y}\left(\sqrt{x+y}-\sqrt{y}\right) \\ &+ b\sqrt{2-x-y}\left(b\sqrt{2-x-y}-b\sqrt{1-y}\right) \end{aligned}$$

Scientist 2's utility function is the same, with x and y reversed.

If there is an interior solution, the derivative of these functions will be zero. For scientist 1's utility function, this occurs when

$$1-\frac{y}{2\sqrt{y(x+y)}}-b^2\left(1-\frac{1-y}{2\sqrt{(1-y)(2-x-y)}}\right)=0$$

This is equivalent to:

$$2(1-b^2)=\sqrt{\frac{y}{x+y}}-b^2\sqrt{\frac{1-y}{2-x-y}}$$

For scientist 2, the constraint is:

$$2(1-b^2)=\sqrt{\frac{x}{x+y}}-b^2\sqrt{\frac{1-x}{2-x-y}}$$

these are only jointly satisfiable when $b=1$. Therefore, when $b\neq 1$, there will be no interior solution. That means, at least one scientist is dedicating all of her time to a single project.